

# **Introduction to Statistical Time Series**

**Second Edition**

**WAYNE A. FULLER**

Iowa State University



**A Wiley-Interscience Publication**

**JOHN WILEY & SONS, INC.**

**New York • Chichester • Brisbane • Toronto • Singapore**

This Page intentionally left blank

# **Introduction to Statistical Time Series**

**WILEY SERIES IN PROBABILITY AND STATISTICS**

Established by **WALTER A. SHEWHART and SAMUEL S. WILKS**

Editors: *Vic Barnett, Ralph A. Bradley, Nicholas I. Fisher, J. Stuart Hunter, J. B. Kadane, David G. Kendall, David W. Scott, Adrian F. M. Smith, Jozef L. Teugels, Geoffrey S. Watson*

A complete list of the titles in this series appears at the end of this volume

# **Introduction to Statistical Time Series**

**Second Edition**

**WAYNE A. FULLER**

Iowa State University



**A Wiley-Interscience Publication**

**JOHN WILEY & SONS, INC.**

**New York • Chichester • Brisbane • Toronto • Singapore**

#### A NOTE TO THE READER

This book has been electronically reproduced from digital information stored at John Wiley & Sons, Inc. We are pleased that the use of this new technology will enable us to keep works of enduring scholarly value in print as long as there is a reasonable demand for them. The content of this book is identical to previous printings.

This text is printed on acid-free paper.

Copyright © 1996 by John Wiley & Sons, Inc.

All rights reserved. Published simultaneously in Canada.

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, scanning or otherwise, except as permitted under Sections 107 or 108 of the 1976 United States Copyright Act, without either the prior written permission of the Publisher, or authorization through payment of the appropriate per-copy fee to the Copyright Clearance Center, 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470. Requests to the Publisher for permission should be addressed to the Permissions Department, John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, (201) 748-6011, fax (201) 748-6008, E-Mail PERMREQ@WILEY.COM.

To order books or for customer service please, call 1(800)-CALL-WILEY (225-5945).

#### *Library of Congress Cataloging in Publication Data:*

Fuller, Wayne A.

Introduction to statistical time series / Wayne A. Fuller. -- 2nd ed.

p. cm. --- (Wiley series in probability and statistics)

"A Wiley-Interscience publication."

Includes bibliographical references and index.

ISBN 0-471-55239-9 (cloth : alk. paper)

1. Time-series analysis. 2. Regression analysis. I. Title

II. Series.

QA280.F84 1996

519.5'5--dc20

95-14875

**To Evelyn**

This Page intentionally left blank

# Contents

<b>Preface to the First Edition</b>	<b>xi</b>
<b>Preface to the Second Edition</b>	<b>xiii</b>
<b>List of Principal Results</b>	<b>xv</b>
<b>List of Examples</b>	<b>xxi</b>
<b>1. Introduction</b>	<b>1</b>
1.1 Probability Spaces	1
1.2 Time Series	3
1.3 Examples of Stochastic Processes	4
1.4 Properties of the Autocovariance and Autocorrelation Functions	7
1.5 Complex Valued Time Series	12
1.6 Periodic Functions and Periodic Time Series	13
1.7 Vector Valued Time Series	15
References	17
Exercises	17
<b>2. Moving Average and Autoregressive Processes</b>	<b>21</b>
2.1 Moving Average Processes	21
2.2 Absolutely Summable Sequences and Infinite Moving Averages	26
2.3 An Introduction to Autoregressive Time Series	39
2.4 Difference Equations	41
2.5 The Second Order Autoregressive Time Series	54
2.6 Alternative Representations of Autoregressive and Moving Average Processes	58
2.7 Autoregressive Moving Average Time Series	70
2.8 Vector Processes	75
2.9 Prediction	79
2.10 The Wold Decomposition	94

2.11	Long Memory Processes	98
	References	101
	Exercises	101
<b>3.</b>	<b>Introduction to Fourier Analysis</b>	<b>112</b>
3.1	Systems of Orthogonal Functions—Fourier Coefficients	112
3.2	Complex Representation of Trigonometric Series	130
3.3	Fourier Transform—Functions Defined on the Real Line	132
3.4	Fourier Transform of a Convolution	136
	References	139
	Exercises	139
<b>4.</b>	<b>Spectral Theory and Filtering</b>	<b>143</b>
4.1	The Spectrum	143
4.2	Circulants—Diagonalization of the Covariance Matrix of Stationary Process	149
4.3	The Spectral Density of Moving Average and Autoregressive Time Series	155
4.4	Vector Processes	169
4.5	Measurement Error—Signal Detection	181
4.6	State Space Models and Kalman Filtering	187
	References	205
	Exercises	205
<b>5.</b>	<b>Some Large Sample Theory</b>	<b>214</b>
5.1	Order in Probability	214
5.2	Convergence in Distribution	227
5.3	Central Limit Theorems	233
5.4	Approximating a Sequence of Expectations	240
5.5	Estimation for Nonlinear Models	250
5.5.1	Estimators that Minimize an Objective Function	250
5.5.2	One-Step Estimation	268
5.6	Instrumental Variables	273
5.7	Estimated Generalized Least Squares	279
5.8	Sequences of Roots of Polynomials	290
	References	299
	Exercises	299

CONTENTS	ix
<b>6. Estimation of the Mean and Autocorrelations</b>	<b>308</b>
6.1 Estimation of the Mean	308
6.2 Estimators of the Autocovariance and Autocorrelation Functions	313
6.3 Central Limit Theorems for Stationary Time Series	320
6.4 Estimation of the Cross Covariances	339
References	348
Exercises	348
<b>7. The Periodogram, Estimated Spectrum</b>	<b>355</b>
7.1 The Periodogram	355
7.2 Smoothing, Estimating the Spectrum	366
7.3 Other Estimators of the Spectrum	380
7.4 Multivariate Spectral Estimates	385
References	400
Exercises	400
<b>8. Parameter Estimation</b>	<b>404</b>
8.1 First Order Autoregressive Time Series	404
8.2 Higher Order Autoregressive Time Series	407
8.2.1 Least Squares Estimation for Univariate Processes	407
8.2.2 Alternative Estimators for Autoregressive Time Series	413
8.2.3 Multivariate Autoregressive Time Series	419
8.3 Moving Average Time Series	421
8.4 Autoregressive Moving Average Time Series	429
8.5 Prediction with Estimated Parameters	443
8.6 Nonlinear Processes	451
8.7 Missing and Outlier Observations	458
8.8 Long Memory Processes	466
References	471
Exercises	471
<b>9. Regression, Trend, and Seasonality</b>	<b>475</b>
9.1 Global Least Squares	476
9.2 Grafted Polynomials	480
9.3 Estimation Based on Least Squares Residuals	484
9.3.1 Estimated Autocorrelations	484
9.3.2 Estimated Variance Functions	488

9.4	Moving Averages—Linear Filtering	497
9.4.1	Moving Averages for the Mean	497
9.4.2	Moving Averages of Integrated Time Series	502
9.4.3	Seasonal Adjustment	504
9.4.4	Differences	507
9.5	Structural Models	509
9.6	Some Effects of Moving Average Operators	513
9.7	Regression with Time Series Errors	518
9.8	Regression Equations with Lagged Dependent Variables and Time Series Errors	530
	References	538
	Exercises	538
10.	Unit Root and Explosive Time Series	546
10.1	Unit Root Autoregressive Time Series	546
10.1.1	The Autoregressive Process with a Unit Root	546
10.1.2	Random Walk with Drift	565
10.1.3	Alternative Estimators	568
10.1.4	Prediction for Unit Root Autoregressions	582
10.2	Explosive Autoregressive Time Series	583
10.3	Multivariate Autoregressive Processes with Unit Roots	596
10.3.1	Multivariate Random Walk	596
10.3.2	Vector Process with a Single Unit Root	599
10.3.3	Vector Process with Several Unit Roots	617
10.4	Testing for a Unit Root in a Moving Average Model	629
	References	638
	Exercises	638
10.A	Percentiles for Unit Root Distributions	641
10.B	Data Used in Examples	653
	Bibliography	664
	Index	689

# Preface to the First Edition

This textbook was developed from a course in time series given at Iowa State University. The classes were composed primarily of graduate students in economics and statistics. Prerequisites for the course were an introductory graduate course in the theory of statistics and a course in linear regression analysis. Since the students entering the course had varied backgrounds, chapters containing elementary results in Fourier analysis and large sample statistics, as well as a section on difference equations, were included in the presentation.

The theorem-proof format was followed because it offered a convenient method of organizing the material. No attempt was made to present the most general results available. Instead, the objective was to give results with practical content whose proofs were generally consistent with the prerequisites. Since many of the statistics students had completed advanced courses, a few theorems were presented at a level of mathematical sophistication beyond the prerequisites. Homework requiring application of the statistical methods was an integral part of the course.

By emphasizing the relationship of the techniques to regression analysis and using data sets of moderate size, most of the homework problems can be worked with any of a number of statistical packages. One such package is SAS (Statistical Analysis System, available through the Institute of Statistics, North Carolina State University). SAS contains a segment for periodogram computations that is particularly suited to this text. The system also contains a segment for regression with time series errors compatible with the presentation in Chapter 9. Another package is available from International Mathematical and Statistical Library, Inc.; this package has a chapter on time series programs.

There is some flexibility in the order in which the material can be covered. For example, the major portions of Chapters 1, 2, 5, 6, 8, and 9 can be treated in that order with little difficulty. Portions of the later chapters deal with spectral matters, but these are not central to the development of those chapters. The discussion of multivariate time series is positioned in separate sections so that it may be introduced at any point.

I thank A. R. Gallant for the proofs of several theorems and for the repair of others; J. J. Goebel for a careful reading of the manuscript that led to numerous substantive improvements and the removal of uncounted mistakes; and D. A.

Dickey, M. Hidiroglou, R. J. Klemm, and G. H. K. Wang for computing examples and for proofreading. G. E. Battese, R. L. Carter, K. R. Crouse, J. D. Cryer, D. P. Hasza, J. D. Jobson, B. Macpherson, J. Mellon, D. A. Pierce and K. N. Wolter also read portions of the manuscript. I also thank my colleagues, R. Groeneveld, D. Isaacson, and O. Kempthorne, for useful comments and discussions. I am indebted to a seminar conducted by Marc Nerlove at Stanford University for the organization of some of the material on Fourier analysis and spectral theory. A portion of the research was supported by joint statistical agreements with the U.S. Bureau of the Census.

I thank Margaret Nichols for the repeated typings required to bring the manuscript to final form and Avonelle Jacobson for transforming much of the original illegible draft into typescript.

WAYNE A. FULLER

*Ames, Iowa  
February 1976*

## Preface to the Second Edition

Considerable development in statistical time series has occurred since the first edition was published in 1976. Notable areas of activity include nonstationary models, nonlinear estimation, multivariate models, state space representations and empirical model identification. The second edition attempts to incorporate new results and to respond to recent emphases while retaining the basic format of the first edition.

With the exception of new sections on the Wold decomposition, partial autocorrelation, long memory processes, and the Kalman filter, Chapters one through four are essentially unchanged from the first edition. Chapter 5 has been enlarged, with additional material on central limit theorems for martingale differences, an expanded treatment of nonlinear estimation, a section on estimated generalized least squares, and a section on the roots of polynomials. Chapter 6 and Chapter 8 have been revised using the asymptotic theory of Chapter 5. Also, the discussion of estimation methods has been modified to reflect advances in computing. Chapter 9 has been revised and the material on the estimation of regression equations has been expanded.

The material on nonstationary autoregressive models is now in a separate chapter, Chapter 10. New tests for unit roots in univariate processes and in vector processes have been added.

As with the first edition, the material is arranged in sections so that there is considerable flexibility to the order in which topics can be covered.

I thank David Dickey and Heon Jin Park for constructing the tables of Chapter 10. I thank Anthony An, Rohit Deo, David Hasza, N. K. Nagaraj, Sastry Pantula, Heon Jin Park, Savas Papadopoulos, Sahadeb Sarkar, Dongwan Shin, and George H. K. Wang for many useful suggestions. I am particularly indebted to Sastry Pantula who assisted with the material of Chapters 5, 8, 9, and 10 and made substantial contributions to other parts of the manuscript, including proofs of several results. Sahadeb Sarkar contributed to the material on nonlinear estimation of Chapter 5, Todd Sanger contributed to the discussion of estimated generalized least squares, Yasuo Amemiya contributed to the section on roots of polynomials, Rohit Deo contributed to the material on long memory processes, Sastry Pantula, Sahadeb Sarkar and Dongwan Shin contributed to the material on the limiting

distribution of estimators for autoregressive moving averages, and Heon Jin Park contributed to the sections on unit root autoregressive processes. I thank Abdoulaye Adam, Jay Breidt, Rohit Deo, Kevin Dodd, Savas Papadopoulos, and Anindya Roy for computing examples. I thank SAS Institute, Cary, NC, for providing computing support to Heon Jin Park for the construction of tables for unit root tests. The research for the second edition was partly supported by joint statistical agreements with the U.S. Bureau of the Census.

I thank Judy Shafer for the extensive word processing required during preparation of the second edition.

WAYNE A. FULLER

*Ames, Iowa  
November 1995*

# List of Principal Results

<i>Theorem</i>	<i>Topic</i>
1.4.1	Covariance function is positive semidefinite, 7
1.4.2	Covariance function is even, 8
1.4.3	Correlation function on real line is a characteristic function, 9
1.4.4	Correlation function on integers is a characteristic function, 9
1.5.1	Covariance function of complex series is positive semidefinite, 13
2.2.1	Weighted average of random variables, where weights are absolutely summable, defines a random variable, 31
2.2.2	Covariance of two infinite sums of random variables, 33
2.2.3	Convergence in mean square of sum of random variables, 35
2.4.1	Order of a polynomial is reduced by differencing, 46
2.4.2	Jordan canonical form of a matrix, 51
2.6.1	Representation of autoregressive process as an infinite moving average, 59
2.6.2	Representation of invertible moving average as an infinite autoregression, 65
2.6.3	Moving average representation of a time series based on covariance function, 66
2.6.4	Canonical representation of moving average time series, 68
2.7.1	Representation of autoregressive moving average as an infinite moving average, 72
2.7.2	Representation of autoregressive moving average as an infinite autoregression, 74
2.8.1	Representation of vector autoregression as an infinite moving average, 77
2.8.2	Representation of vector moving average as an infinite autoregression, 78
2.9.1	Minimum mean square error predictor, 80
2.9.2	Durbin–Levinson algorithm for constructing predictors, 82

- 2.9.3 Predictors as a function of previous prediction errors, 86  
2.9.4 Limit of prediction error is a moving average, 89  
2.10.1 Limit of one period prediction error, 94  
2.10.2 Wold decomposition, 96  
3.1.1 Sine and cosine functions form an orthogonal basis for  $N$  dimensional vectors, 112  
3.1.2 Sine and cosine functions are orthogonal on  $[-\pi, \pi]$ , 116  
3.1.3 Bessel's inequality, 118  
3.1.4 Fourier coefficients are zero if and only if function is zero, 119  
3.1.5 If Fourier coefficients are zero, integral of function is zero, 119  
3.1.6 Integral of function defined in terms of Fourier coefficients, 120  
3.1.7 Pointwise representation of a function by Fourier series, 123  
3.1.8 Absolute convergence of Fourier series for a class of functions, 125  
3.1.9 The correlation function defines a continuous spectral density, 127  
3.1.10 The Fourier series of a continuous function is Cesàro summable, 129  
3.3.1 Fourier integral theorem, 133  
3.4.1 Fourier transform of a convolution, 137  
4.2.1 Approximate diagonalization of covariance matrix with orthogonal sine–cosine functions, 154  
4.3.1 Spectral density of an infinite moving average of a time series, 156  
4.3.2 Moving average representation of a time series based on covariance function, 162  
4.3.3 Moving average representation of a time series based on continuous spectral density, 163  
4.3.4 Autoregressive representation of a time series based on continuous spectral density, 165  
4.3.5 Spectral density of moving average with square summable coefficients, 167  
4.4.1 Spectral density of vector process, 179  
4.5.1 Linear filter for time series observed subject to measurement error, 183  
5.1.1 Chebyshev's inequality, 219  
5.1.2 Common probability limits, 221  
5.1.3 Convergence in  $r$ th mean implies convergence in probability, 221  
5.1.4 Probability limit of a continuous function, 222  
5.1.5 The algebra of  $O_p$ , 223  
5.1.6 The algebra of  $o_p$ , 224  
5.1.7 Taylor expansion about a random point, 226  
5.2.1 Convergence in law when difference converges in probability, 228

- 5.2.2 Helly–Bray Theorem, 230  
5.2.3 Joint convergence of distribution functions and characteristic functions, 230  
5.2.4 Convergence in distribution of continuous functions, 230  
5.2.5 Joint convergence in law of independent sequences, 230  
5.2.6 Joint convergence in law when one element converges to a constant, 232  
5.3.1 Lindeberg central limit theorem, 233  
5.3.2 Liapounov central limit theorem, 233  
5.3.3 Central limit theorem for vectors, 234  
5.3.4 Central limit theorem for martingale differences, 235  
5.3.5 Functional central limit theorem, 236  
5.3.6 Convergence of functions of partial sums, 237  
5.3.7 Multivariate functional central limit theorem, 238  
5.3.8 Almost sure convergence of martingales, 239  
5.4.1 Moments of products of sample means, 242  
5.4.2 Approximate expectation of functions of means, 243  
5.4.3 Approximate expectation of functions of vector means, 244  
5.4.4 Bounded integrals of functions of random variables, 247  
5.5.1 Limiting properties of nonlinear least squares estimator, 256  
5.5.2 Limiting distribution of estimator defined by an objective function, 260  
5.5.3 Consistency of nonlinear estimator with different rates of convergence, 262  
5.5.4 Limiting properties of one-step Gauss–Newton estimator, 269  
5.6.1 Limiting properties of instrumental variable estimators, 275  
5.7.1 Convergence of estimated generalized least squares estimator to generalized least squares estimator, 280  
5.7.2 Central limit theorem for estimated generalized least squares, 284  
5.7.3 Estimated generalized least squares with finite number of covariance parameters, 286  
5.7.4 Estimated generalized least squares based on simple least squares residuals, 289  
5.8.1 Convergence of roots of a sequence of polynomials, 295  
5.8.2 Differentials of roots of determinantal equation, 298  
6.1.1 Convergence in mean square of sample mean, 309  
6.1.2 Variance of sample mean as function of the spectral density, 310  
6.1.3 Limiting efficiency of sample mean, 312  
6.2.1 Covariances of sample autocovariances, 314  
6.2.2 Covariances of sample autocovariances, mean estimated, 316

- 6.2.3 Covariances of sample correlations, 317
- 6.3.1 Central limit theorem for  $m$ -dependent sequences, 321
- 6.3.2 Convergence in probability of mean of an infinite moving average, 325
- 6.3.3 Central limit theorem for mean of infinite moving average, 326
- 6.3.4 Central limit theorem for linear function of infinite moving average, 329
- 6.3.5 Consistency of sample autocovariances, 331
- 6.3.6 Central limit theorem for autocovariances, 333
- 6.4.1 Covariances of sample autocovariances of vector time series, 342
- 6.4.2 Central limit theorem for cross covariances one series independent  $(0, \sigma^2)$ , 345
- 7.1.1 Expected value of periodogram, 359
- 7.1.2 Limiting distribution of periodogram ordinates, 360
- 7.2.1 Covariances of periodogram ordinates, 369
- 7.2.2 Limiting behavior of weighted averages of periodogram ordinates, 372
- 7.3.1 Limiting behavior of estimated spectral density based on weighted autocovariances, 382
- 7.4.1 Diagonalization of covariance matrix of bivariate process, 387
- 7.4.2 Distribution of sample Fourier coefficients for bivariate process, 389
- 7.4.3 Distribution of smoothed bivariate periodogram, 390
- 8.2.1 Limiting distribution of regression estimators of parameters of  $p$ th order autoregressive process, 408
- 8.2.2 Equivalence of alternative estimators of parameters of autoregressive process, 418
- 8.2.3 Limiting distribution of estimators of parameter of  $p$ th order vector autoregressive process, 420
- 8.3.1 Limiting distribution of nonlinear estimator of parameter of first order moving average, 424
- 8.4.1 Limiting distribution of nonlinear estimator of vector of parameters of autoregressive moving average, 432
- 8.4.2 Equivalence of alternative estimators for autoregressive moving average, 434
- 8.5.1 Order of error in prediction due to estimated parameters, 444
- 8.5.2 Expectation of prediction error, 445
- 8.5.3 Order  $n^{-1}$  approximation to variance of prediction error, 446
- 8.6.1 Polynomial autoregression, 452
- 8.8.1 Maximum likelihood estimators for long memory processes, 470
- 9.1.1 Limiting distribution of simple least squares estimated parameters of regression model with time series errors, 478

- 9.1.2 Spectral representation of covariance matrix of simple least squares estimator and of generalized least squares estimator, 479  
9.1.3 Asymptotic relative efficiency of simple least squares and generalized least squares, 480  
9.3.1 Properties of autocovariances computed from least squares residuals, 485  
9.4.1 Centered moving average estimator of polynomial trend, 501  
9.4.2 Trend moving average removes autoregressive unit root, 502  
9.4.3 Effect of a moving average for polynomial trend removal, 504  
9.7.1 Asymptotic equivalence of generalized least squares and estimated generalized least squares for model with autoregressive errors, 521  
9.7.2 Limiting distribution of maximum likelihood estimator of regression model with autoregressive errors, 526  
9.8.1 Limiting distribution of least squares estimator of model with lagged dependent variables, 530  
9.8.2 Limiting distribution of instrumental variable estimator of model with lagged dependent variables, 532  
9.8.3 Limiting properties of autocovariances computed from instrumental variable residuals, 534  
10.1.1 Limiting distribution of least squares estimator of first order autoregressive process with unit root, 550  
10.1.2 Limiting distribution of least squares estimator of  $p$ th order autoregressive process with a unit root, 556  
10.1.3 Limit distribution of least squares estimator of first order unit root process with mean estimated, 561  
10.1.4 Limiting distribution of least squares estimator of  $p$ th order process with a unit root and mean estimated, 563  
10.1.5 Limiting distribution of least squares estimator of unit root autoregression with drift, 566  
10.1.6 Limiting distribution of least squares estimator of unit root autoregression with time in fitted model, 567  
10.1.7 Limiting distribution of symmetric estimator of first order unit root autoregressive process, 570  
10.1.8 Limiting distribution of symmetric estimators adjusted for mean and time trend, 571  
10.1.9 Limiting distribution of symmetric test statistics for  $p$ th order autoregressive process with a unit root, 573  
10.1.10 Limiting distribution of maximum likelihood estimator of unit root, 573  
10.1.11 Order of error in prediction of unit root process with estimated parameters, 582

- 10.2.1 Limiting distribution of explosive autoregressive estimator, 585
- 10.2.2 Limiting distribution of vector of least squares estimators for  $p$ th order autoregression with an explosive root, 589
- 10.3.1 Limiting distribution of least squares estimator of vector of estimators of equation containing the lagged dependent variable with unit coefficient, 600
- 10.3.2 Limiting distribution of least squares estimator when one of the explanatory variables is a unit root process, 603
- 10.3.3 Limiting distribution of least squares estimator of coefficients of vector process with unit roots, 610
- 10.3.4 Limiting distribution of maximum likelihood estimator for multivariate process with a single root, 613
- 10.3.5 Limiting distribution of maximum likelihood estimator for multivariate process with  $g$  unit roots, 619

# List of Examples

<i>Number</i>	<i>Topic</i>
1.3.1	Finite index set, 4
1.3.2	White noise, 5
1.3.3	A nonstationary time series, 5
1.3.4	Continuous time series, 6
2.1.1	First order moving average time series, 23
2.1.2	Second order moving average time series, 24
2.3.1	First order autoregressive time series, 40
2.5.1	Correlogram of time series, 57
2.9.1	Prediction for unemployment rate, 90
2.9.2	Prediction for autoregressive moving average, 91
4.1.1	Spectral distribution function composed of jumps, 147
4.1.2	Spectral distribution function of series with white noise component, 148
4.5.1	Filter for time series observed subject to measurement error, 184
4.6.1	Kalman Filter for Des Moines River data, 192
4.6.2	Kalman filter, Des Moines River, mean unknown, 193
4.6.3	Kalman filter, autoregressive unit root, 195
4.6.4	Kalman filter for missing observations, 198
4.6.5	Predictions constructed with the Kalman filter, 202
5.1.1	Taylor expansion about a random point, 227
5.4.1	Approximation to expectation, 248
5.5.1	Transformation of model with different rates of convergence, 265
5.5.2	Failure of convergence of second derivative of nonlinear model, 267
5.5.3	Gauss–Newton estimation, 272
5.6.1	Instrumental variable estimation, 277
6.3.1	Sample autocorrelations and means of unemployment rate, 336

- 6.4.1 Autocorrelations and cross correlations of Boone–Saylorville data, 345
- 7.1.1 Periodogram for wheat yield data, 363
- 7.2.1 Periodogram of autoregression, 375
- 7.2.2 Periodogram of unemployment rate, 379
- 7.4.1 Cross spectrum computations, 394
- 8.2.1 Autoregressive fit to unemployment rate, 412
- 8.3.1 Estimation of first order moving average, 427
- 8.4.1 Autoregressive moving average fit to artificial data, 439
- 8.4.2 Autoregressive moving average fit to unemployment rate, 440
- 8.5.1 Prediction with estimated parameters, 449
- 8.6.1 Nonlinear models for lynx data, 455
- 8.7.1 Missing observations, 459
- 8.7.2 Outlier observations, 462
- 9.2.1 Grafted quadratic fit to U.S. wheat yields, 482
- 9.3.1 Variance as a function of the mean, 490
- 9.3.2 Stochastic volatility model, 495
- 9.5.1 Structural model for wheat yields, 510
- 9.7.1 Regression with autoregressive errors, spirit consumption, 522
- 9.7.2 Nonlinear estimation of trend in wheat yields with autoregressive error, 527
- 9.7.3 Nonlinear estimation for spirit consumption model, 528
- 9.8.1 Regression with lagged dependent variable and autoregressive errors, 535
- 10.1.1 Estimation and testing for a unit root, ordinary least squares, 564
- 10.1.2 Testing for a unit root, symmetric and likelihood procedures, 577
- 10.1.3 Estimation for process with autoregressive root in  $(-1, 1]$ , 581
- 10.1.4 Prediction for a process with a unit root, 583
- 10.2.1 Estimation with an explosive root, 593
- 10.2.2 Prediction for an explosive process, 596
- 10.3.1 Estimation of regression with autoregressive explanatory variable, 606
- 10.3.2 Estimation and testing for vector process with unit roots, 624
- 10.4.1 Test for a moving average unit root, 635

# **Introduction to Statistical Time Series**

This Page intentionally left blank

## CHAPTER 1

# Introduction

The analysis of time series applies to many fields. In economics the recorded history of the economy is often in the form of time series. Economic behavior is quantified in such time series as the consumer price index, unemployment, gross national product, population, and production. The natural sciences also furnish many examples of time series. The water level in a lake, the air temperature, the yields of corn in Iowa, and the size of natural populations are all collected over time. Growth models that arise naturally in biology and in the social sciences represent an entire field in themselves.

The mathematical theory of time series has been an area of considerable activity in recent years. Applications in the physical sciences such as the development of designs for airplanes and rockets, the improvement of radar and other electronic devices, and the investigation of certain production and chemical processes have resulted in considerable interest in time series analysis. This recent work should not disguise the fact that the analysis of time series is one of the oldest activities of scientific man.

A successful application of statistical methods to the real world requires a melding of statistical theory and knowledge of the material under study. We shall confine ourselves to the statistical treatment of moderately well-behaved time series, but we shall illustrate some techniques with real data.

### 1.1. PROBABILITY SPACES

When investigating outcomes of a game, an experiment, or some natural phenomenon, it is useful to have a representation for all possible outcomes. The individual outcomes, denoted by  $\omega$ , are called *elementary events*. The set of all possible elementary events is called the *sure event* and is denoted by  $\Omega$ . An example is the tossing of a die, where we could take  $\Omega = \{\text{one spot shows, two spots show, \dots, six spots show}\}$  or, more simply,  $\Omega = \{1, 2, 3, 4, 5, 6\}$ .

Let  $A$  be a subset of  $\Omega$ , and let  $\mathcal{A}$  be a collection of such subsets. If we observe the outcome  $\omega$  and  $\omega$  is in  $A$ , we say that  $A$  has occurred. Intuitively, it is possible

to specify  $P(A)$ , the probability that (or expected long-run frequency with which)  $A$  will occur. It is reasonable to require that the function  $P(A)$  satisfy:

**AXIOM 1.**  $P(A) \geq 0$  for every  $A$  in  $\mathcal{A}$ .

**AXIOM 2.**  $P(\Omega) = 1$ .

**AXIOM 3.** If  $A_1, A_2, \dots$  is a countable sequence from  $\mathcal{A}$  and  $A_i \cap A_j$  is the null set for all  $i \neq j$ , then  $P(\bigcup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$ .

Using our die tossing example, if the die is fair we would take  $P(A) = \frac{1}{6}$  [the number of elementary events  $\omega$  in  $A$ ]. Thus  $P(\{1, 3, 5\}) = \frac{1}{6} \times 3 = \frac{1}{2}$ . It may be verified that Axioms 1 to 3 are satisfied for  $\mathcal{A}$  equal to  $\mathcal{P}(\Omega)$ , the collection of all possible subsets of  $\Omega$ .

Unfortunately, for technical mathematical reasons, it is not always possible to define  $P(A)$  for all  $A$  in  $\mathcal{P}(\Omega)$  and also to satisfy Axiom 3. To eliminate this difficulty, the class of subsets  $\mathcal{A}$  of  $\Omega$  on which  $P$  is defined is restricted. The collection  $\mathcal{A}$  is required to satisfy:

1. If  $A$  is in  $\mathcal{A}$ , then the complement  $A^c$  is also in  $\mathcal{A}$ .
2. If  $A_1, A_2, \dots$  is a countable sequence from  $\mathcal{A}$ , then  $\bigcup_{i=1}^{\infty} A_i$  is in  $\mathcal{A}$ .
3. The null set is in  $\mathcal{A}$ .

A nonempty collection  $\mathcal{A}$  of subsets of  $\Omega$  that satisfies conditions 1 to 3 is said to be a *sigma-algebra* or *sigma-field*.

We are now in a position to give a formal definition of a probability space. A *probability space*, represented by  $(\Omega, \mathcal{A}, P)$ , is the sure event  $\Omega$  together with a sigma-algebra  $\mathcal{A}$  of subsets of  $\Omega$  and a function  $P(A)$  defined on  $\mathcal{A}$  that satisfies Axioms 1 to 3.

For our purposes it is unnecessary to consider the subject of probability spaces in detail. In simple situations such as tossing a die,  $\Omega$  is easy to enumerate, and  $P$  satisfies Axioms 1 to 3 for  $\mathcal{A}$  equal to the set of all subsets of  $\Omega$ .

Although it is conceptually possible to enumerate all possible outcomes of an experiment, it may be a practical impossibility to do so, and for most purposes it is unnecessary to do so. It is usually enough to record the outcome by some function that assumes values on the real line. That is, we assign to each outcome  $\omega$  a real number  $X(\omega)$ , and if  $\omega$  is observed, we record  $X(\omega)$ . In our die tossing example we could take  $X(\omega) = 1$  if the player wins and  $-1$  if the house wins.

Formally, a *random variable*  $X$  is a real valued function defined on  $\Omega$  such that the set  $\{\omega : X(\omega) \leq x\}$  is a member of  $\mathcal{A}$  for every real number  $x$ . The function  $F_x(x) = P(\{\omega : X(\omega) \leq x\})$  is called the *distribution function* of the random variable  $X$ .

The reader who wishes to explore further the subjects of probability spaces, random variables, and distribution functions for stochastic processes may read Tucker (1967, pp. 1–33). The preceding brief introduction will suffice for our purposes.

## 1.2. TIME SERIES

Let  $(\Omega, \mathcal{A}, P)$  be a probability space, and let  $T$  be an index set. A real valued *time series* (or *stochastic process*) is a real valued function  $X(t, \omega)$  defined on  $T \times \Omega$  such that for each fixed  $t$ ,  $X(t, \omega)$  is a random variable on  $(\Omega, \mathcal{A}, P)$ . The function  $X(t, \omega)$  is often written  $X_t(\omega)$  or  $X_t$ , and a time series can be considered as a collection  $\{X_t : t \in T\}$  of random variables.

For fixed  $\omega$ ,  $X(t, \omega)$  is a real valued function of  $t$ . This function of  $t$  is called a *realization* or a *sample function*. If we look at a plot of some recorded time series such as the gross national product, it is important to realize that conceptually we are looking at a plot of  $X(t, \omega)$  with  $\omega$  fixed. The collection of all possible realizations is called the *ensemble* of functions or the ensemble of realizations.

If the index set contains exactly one element, the stochastic process is a single random variable and we have defined the distribution function of the process. For stochastic processes with more than one random variable we need to consider the *joint distribution function*. The joint distribution function of a finite set of random variables  $\{X_{t_1}, X_{t_2}, \dots, X_{t_n}\}$  from the collection  $\{X_t : t \in T\}$  is defined by

$$\begin{aligned} F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_{t_1}, x_{t_2}, \dots, x_{t_n}) \\ = P\{\omega : X(t_1, \omega) \leq x_{t_1}, \dots, X(t_n, \omega) \leq x_{t_n}\}. \end{aligned} \quad (1.2.1)$$

A time series is called *strictly stationary* if

$$\begin{aligned} F_{X_{t_1}, X_{t_2}, \dots, X_{t_n}}(x_{t_1}, x_{t_2}, \dots, x_{t_n}) \\ = F_{X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h}}(x_{t_1}, x_{t_2}, \dots, x_{t_n}), \end{aligned}$$

where the equality must hold for all possible (nonempty finite distinct) sets of indices  $t_1, t_2, \dots, t_n$  and  $t_1 + h, t_2 + h, \dots, t_n + h$  in the index set and all  $(x_{t_1}, x_{t_2}, \dots, x_{t_n})$  in the range of the random variable  $X_t$ . Note that the indices  $t_1, t_2, \dots, t_n$  are not necessarily consecutive. If a time series is strictly stationary, we see that the distribution function of the random variable is the same at every point in the index set. Furthermore, the joint distribution depends only on the distance between the elements in the index set, and not on their actual values. Naturally this does not mean a particular realization will appear the same as another realization.

If  $\{X_t : t \in T\}$  is a strictly stationary time series with  $E\{|X_t|\} < \infty$ , then the expected value of  $X_t$  is a constant for all  $t$ , since the distribution function is the same for all  $t$ . Likewise, if  $E\{X_t^2\} < \infty$ , then the variance of  $X_t$  is a constant for all  $t$ .

A time series is defined completely in a probabilistic sense if one knows the cumulative distribution (1.2.1) for any finite set of random variables  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$ . However, in most applications, the form of the distribution function is not known. A great deal can be accomplished, however, by dealing

only with the first two moments of the time series. In line with this approach we define a time series to be *weakly stationary* if:

1. The expected value of  $X_t$  is a constant for all  $t$ .
2. The covariance matrix of  $(X_{t_1}, X_{t_2}, \dots, X_{t_n})$  is the same as the covariance matrix of  $(X_{t_1+h}, X_{t_2+h}, \dots, X_{t_n+h})$  for all nonempty finite sets of indices  $(t_1, t_2, \dots, t_n)$ , and all  $h$  such that  $t_1, t_2, \dots, t_n, t_1 + h, t_2 + h, \dots, t_n + h$  are contained in the index set.

As before  $t_1, t_2, \dots, t_n$  are not necessarily consecutive members of the index set. Also, since the expected value of  $X_t$  is a constant, it may conveniently be taken as 0. The covariance matrix, by definition, is a function only of the distance between observations. That is, the covariance of  $X_{t+h}$  and  $X_t$  depends only on the distance,  $h$ , and we may write

$$\text{Cov}\{X_t, X_{t+h}\} = E\{X_t X_{t+h}\} = \gamma(h),$$

where  $E\{X_t\}$  has been taken to be zero. The function  $\gamma(h)$  is called the *autocovariance* of  $X_t$ . When there is no danger of confusion, we shall abbreviate the expression to *covariance*.

The terms *stationary in the wide sense*, *covariance stationary*, *second order stationary*, and *stationary* are also used to describe a weakly stationary time series. It follows from the definitions that a strictly stationary process with the first two moments finite is also weakly stationary. However, a strictly stationary time series may not possess finite moments and hence may not be covariance stationary.

Many time series as they occur in practice are not stationary. For example, the economies of many countries are developing or growing. Therefore, the typical economic indicators will be showing a "trend" through time. This trend may be in either the mean, the variance, or both. Such nonstationary time series are sometimes called *evolutionary*. A good portion of the practical analysis of time series is connected with the transformation of an evolving time series into a stationary time series. In later sections we shall consider several of the procedures used in this connection. Many of these techniques will be familiar to the reader because they are closely related to least squares and regression.

### 1.3. EXAMPLES OF STOCHASTIC PROCESSES

**Example 1.3.1.** Let the index set be  $T = \{1, 2\}$ , and let the space of outcomes be the possible outcomes associated with tossing two dice, one at "time"  $t = 1$  and one at time  $t = 2$ . Then

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

Define

$$X(t, \omega) = t + [\text{value on die } t]^2.$$

Therefore, for a particular  $\omega$ , say  $\omega_3 = (1, 3)$ , the realization or sample function would be  $(1 + 1, 2 + 9) = (2, 11)$ . In this case, both  $\Omega$  and  $T$  are finite, and we are able to determine that there are 36 possible realizations.  $\blacktriangle\blacktriangle$

**Example 1.3.2.** One of the most important time series is a sequence of uncorrelated random variables, say  $\{e_t : t \in (0, \pm 1, \pm 2, \dots)\}$ , each with zero mean and finite variance,  $\sigma^2 > 0$ . We say that  $e_t$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables.

This time series is sometimes called *white noise*. Note that the index set is the set of all integers. The set  $\Omega$  is determined by the range of the random variables. Let us assume that the  $e_t$  are normally distributed and therefore have range  $(-\infty, \infty)$ . Then  $\omega \in \Omega$  is a real valued infinite-dimensional vector with an element associated with each integer. The covariance function of  $\{e_t\}$  is given by

$$\gamma_e(h) = \begin{cases} \sigma^2, & h = 0, \\ 0 & \text{otherwise.} \end{cases}$$

Because of the importance of this time series, we shall reserve the symbol  $e_t$  for a sequence of uncorrelated random variables with zero mean and positive finite variance. On occasion we shall further assume that the variables are independent and perhaps of a specified distribution. These additional assumptions will be stated when used.

In a similar manner the most commonly used index set will be the set of all integers. If the index set is not stated, the reader may assume that it is the set of all integers.  $\blacktriangle\blacktriangle$

**Example 1.3.3.** Consider the time series

$$X_t = \hat{\beta}_1 + \hat{\beta}_2 t, \quad t \in [0, 1],$$

where  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2)'$  is distributed as a bivariate normal random variable with mean  $\beta = (\beta_1, \beta_2)'$  and covariance matrix

$$\begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{12} & \sigma_{22} \end{pmatrix}.$$

Any realization of this process yields a function continuous on the interval  $[0, 1]$ , and the process therefore is called *continuous*. Such a process might represent the outcome of an experiment to estimate the linear response to an input variable measured on the interval  $[0, 1]$ . The  $\hat{\beta}$ 's are then the estimated regression coefficients. The set  $\Omega$  is the two-dimensional space of real numbers. Each

experiment conducted to obtain a set of estimates of  $\hat{\beta}_1$  and  $\hat{\beta}_2$  constitutes a realization of the process.

The mean function of  $X_t$  is given by

$$E\{X_t\} = E\{\hat{\beta}_1 + \hat{\beta}_2 t\} = \beta_1 + \beta_2 t,$$

and the covariance function by

$$\begin{aligned} & E\{(X_t - E\{X_t\})(X_{t+h} - E\{X_{t+h}\})\} \\ &= E\{[(\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2)t][(\hat{\beta}_1 - \beta_1) + (\hat{\beta}_2 - \beta_2)(t+h)]\} \\ &= \sigma_{11} + \sigma_{12}(t+h) + \sigma_{12}t + \sigma_{22}t(t+h), \quad t, t+h \in [0, 1]. \end{aligned}$$

It is clear that this process is not stationary. ▲▲

**Example 1.3.4.** For the reader familiar only with the study of multivariate statistics, the idea of a random variable that is both continuous and random in time may require a moment of reflection. On the other hand, such processes occur in nature. For example, the water level of a lake or river can be plotted as a continuous function of time. Furthermore, such a plot might appear so smooth as to support the idea of a derivative or instantaneous rate of change in level.

Consider such a process,  $\{X_t : t \in (-\infty, \infty)\}$ , and let the covariance of the process be given by

$$\gamma(h) = E\{(X_t - \mu)(X_{t+h} - \mu)\} = Ae^{-\alpha h^2}, \quad \alpha > 0.$$

Thus, for example, if time is measured in hours and the river level at 7:00 A.M. is reported daily, then the covariance between daily reports is  $Ae^{-576\alpha}$ . Likewise, the covariance between the change from 7:00 A.M. to 8:00 A.M. and the change from 8:00 A.M. to 9:00 A.M. is given by

$$E\{(X_{t+1} - X_t)(X_{t+2} - X_{t+1})\} = -A[1 - 2e^{-\alpha} + e^{-4\alpha}].$$

Note that the variance of the change from  $t$  to  $t+h$  is

$$\text{Var}\{X_{t+h} - X_t\} = 2A[1 - e^{-\alpha h^2}],$$

and

$$\lim_{h \rightarrow 0} \text{Var}\{X_{t+h} - X_t\} = 0.$$

A process with this property is called *mean square continuous*.

We might define the rate of change per unit of time by

$$R_t(h) = \frac{X_{t+h} - X_t}{h}.$$

For a fixed  $h > 0$ ,  $R_i(h)$  is a well-defined random variable and

$$\text{Var}\{R_i(h)\} = \frac{2A[1 - e^{-\alpha h^2}]}{h^2}.$$

Furthermore, by L'Hospital's rule,

$$\begin{aligned}\lim_{h \rightarrow 0} \text{Var}\{R_i(h)\} &= \lim_{h \rightarrow 0} \frac{2A[2\alpha h e^{-\alpha h^2}]}{2h} \\ &= 2A\alpha.\end{aligned}$$

Stochastic processes for which this limit exists are called *mean square differentiable*. ▲▲

#### 1.4. PROPERTIES OF THE AUTOCOVARIANCE AND AUTOCORRELATION FUNCTIONS

To compare the basic properties of time series, it is often useful to have a function that is not influenced by the units of measurement. To this end, we define the *autocorrelation function* of a stationary time series by

$$\rho(h) = \frac{\gamma(h)}{\gamma(0)}.$$

Thus the autocorrelation function is the autocovariance function normalized to be one at  $h = 0$ . As with the autocovariance function, we shall abbreviate the expression to correlation function when no confusion will result.

The autocovariance and autocorrelation functions of stationary time series possess several distinguishing characteristics. A function  $f(x)$  defined for  $x \in \chi$  is said to be *positive semidefinite* if it satisfies

$$\sum_{j=1}^n \sum_{k=1}^n a_j a_k f(t_k - t_j) \geq 0 \quad (1.4.1)$$

for any set of real numbers  $(a_1, a_2, \dots, a_n)$  and for any  $(t_1, t_2, \dots, t_n)$  such that  $t_i - t_j$  is in  $\chi$  for all  $(i, j)$ . Some authors reserve the term positive semidefinite for a function that is zero for at least one nonzero vector  $(a_1, a_2, \dots, a_n)$  and use *nonnegative definite* to describe functions satisfying (1.4.1). We shall use both terms for functions satisfying (1.4.1).

**Theorem 1.4.1.** The covariance function of the stationary time series  $\{T_t : t \in T\}$  is nonnegative definite.

**Proof.** Without loss of generality, let  $E\{X_i\} = 0$ . Let  $(t_1, t_2, \dots, t_n) \in T$ , let  $(a_1, a_2, \dots, a_n)$  be any set of real numbers, and let  $\gamma(t_k - t_j)$  be the covariance

between  $X_{t_j}$  and  $X_{t_k}$ . We know that the variance of a random variable, when it is defined, is nonnegative. Therefore,

$$\begin{aligned} 0 &\leq \text{Var}\left\{\sum_{j=1}^n a_j X_{t_j}\right\} = E\left\{\sum_{j=1}^n \sum_{k=1}^n a_j a_k X_{t_j} X_{t_k}\right\} \\ &= \sum_{j=1}^n \sum_{k=1}^n a_j a_k \gamma(t_k - t_j). \end{aligned}$$
▲

If we set  $n = 2$  in (1.4.1), we have

$$0 \leq a_1^2 \gamma(0) + a_2^2 \gamma(0) + 2a_1 a_2 \gamma(t_1 - t_2),$$

which implies

$$0.5(a_1^2 + a_2^2) \geq -a_1 a_2 [\gamma(0)]^{-1} \gamma(t_1 - t_2).$$

For  $t_1 - t_2 = h$  we set  $-a_1 = a_2 = 1$  and then set  $a_1 = a_2 = 1$  to obtain the well-known property of correlations:

$$|\rho(h)| \leq 1. \quad (1.4.2)$$

The concepts of *even* and *odd* functions (about zero) will prove useful in our study. For example,  $\cos t$  is an *even function*. The use of this description is apparent when one notes the symmetry of  $\cos t$  on the interval  $[-\pi, \pi]$ . Similarly,  $\sin t$  is an *odd function*. In general, an even function,  $f(t)$ , defined on a domain  $T$ , is a function that satisfies  $f(t) = f(-t)$  for all  $t$  and  $-t$  in  $T$ . An odd function  $g(t)$  is a function satisfying  $g(t) = -g(-t)$  for all  $t$  and  $-t$  in  $T$ . Many simple properties of even and odd functions follow immediately. For example, if  $f(t)$  is an even function and  $g(t)$  is an odd function, where both are integrable on the interval  $[-A, A]$ , then, for  $0 \leq b \leq A$ ,

$$\int_{-b}^b g(t) dt = 0,$$

$$\int_{-b}^b f(t) dt = 2 \int_0^b f(t) dt.$$

As an exercise, the reader may verify that the product of two even functions is even, the product of two odd functions is even, the product of an odd and an even function is odd, the sum of even functions is even, and the sum of odd functions is odd.

**Theorem 1.4.2.** The covariance function of a real valued stationary time series is an even function of  $h$ . That is,  $\gamma(h) = \gamma(-h)$ .

**Proof.** We assume, without loss of generality, that  $E\{X_t\} = 0$ . By stationarity,

$$E\{X_t X_{t+h}\} = \gamma(h)$$

for all  $t$  and  $t+h$  contained in the index set. Therefore, if we set  $t_0 = t_1 - h$ ,

$$\gamma(h) = E\{X_{t_0} X_{t_0+h}\} = E\{X_{t_1-h} X_{t_1}\} = \gamma(-h). \quad \blacktriangle$$

Given this theorem, we shall often evaluate  $\gamma(h)$  for real valued time series for nonnegative  $h$  only. Should we fail to so specify, the reader may always safely substitute  $|h|$  for  $h$  in a covariance function.

In the study of statistical distribution functions the characteristic function of a distribution function is defined by

$$\varphi(h) = \int e^{ixh} dG(x),$$

where the integral is a Lebesgue-Stieltjes integral,  $G(x)$  is the distribution function, and  $e^{ixh}$  is the complex exponential defined by  $e^{ixh} = \cos xh + i \sin xh$ . It is readily established that the function  $\varphi(h)$  satisfies:

1.  $\varphi(0) = 1$ .
2.  $|\varphi(h)| \leq 1$  for all  $h \in (-\infty, \infty)$ .
3.  $\varphi(h)$  is uniformly continuous on  $(-\infty, \infty)$ .

See, for example, Tucker (1967, p. 42 ff.) or Gnedenko (1967, p. 266 ff.).

It can be shown that a continuous function  $\rho(h)$  with  $\rho(0) = 1$  is a characteristic function if and only if it is nonnegative definite. For example, see Gnedenko (1967, pp. 290, 387).

**Theorem 1.4.3.** The real valued function  $\rho(h)$  is the correlation function of a mean square continuous stationary real valued time series with index set  $T = (-\infty, \infty)$  if and only if it is representable in the form

$$\rho(h) = \int_{-\infty}^{\infty} e^{ixh} dG(x),$$

where  $G(x)$  is a symmetric distribution function, and the integral is a Lebesgue-Stieltjes integral.

For the index set  $T = \{0, \pm 1, \pm 2, \dots\}$ , the corresponding theorem is:

**Theorem 1.4.4.** The real valued function  $\rho(h)$  is the correlation function of a real valued stationary time series with index set  $T = \{0, \pm 1, \pm 2, \dots\}$  if and only if

it is representable in the form

$$\rho(h) = \int_{-\pi}^{\pi} e^{i x h} dG(x),$$

where  $G(x)$  is a symmetric distribution function.

This theorem will be discussed in Chapter 3.

Associated with the autocorrelation function of a time series is the partial autocorrelation function. Before discussing ideas of partial correlation for time series, we recall the partial correlation coefficient of multivariate analysis. Let  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_p)$  be a  $p$ -variate random variable with nonsingular covariance matrix. Then the partial correlation between  $Y_i$  and  $Y_j$  after  $Y_k$  is

$$\rho_{ij\cdot k} = [(1 - \rho_{13}^2)(1 - \rho_{23}^2)]^{-1/2} (\rho_{12} - \rho_{13}\rho_{23}), \quad (1.4.3)$$

where  $\rho_{ij} = (\sigma_{ii}\sigma_{jj})^{-1/2}\sigma_{ij}$  and  $\sigma_{ij}$  is the covariance between  $Y_i$  and  $Y_j$ . An alternative definition of  $\rho_{ij\cdot k}^2$  in terms of regression coefficients is

$$\rho_{ij\cdot k}^2 = \beta_{12\cdot k} \beta_{21\cdot k}, \quad (1.4.4)$$

where  $\beta_{ij\cdot k}$  is the partial regression coefficient for  $Y_i$  in the regression of  $Y_j$  on  $Y_i$  and  $Y_k$ . For example,

$$\begin{pmatrix} \beta_{12\cdot 3} \\ \beta_{13\cdot 2} \end{pmatrix} = \begin{pmatrix} \sigma_{22} & \sigma_{23} \\ \sigma_{32} & \sigma_{33} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{12} \\ \sigma_{13} \end{pmatrix},$$

$$\begin{pmatrix} \beta_{21\cdot 3} \\ \beta_{23\cdot 1} \end{pmatrix} = \begin{pmatrix} \sigma_{11} & \sigma_{13} \\ \sigma_{31} & \sigma_{33} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{21} \\ \sigma_{23} \end{pmatrix}.$$

Recall that  $\beta_{ij\cdot k}$  is the simple population regression coefficient for the regression of

$$Y_i - \beta_{ik} Y_k \quad \text{on} \quad Y_j - \beta_{jk} Y_k,$$

where  $\beta_{ik} = \sigma_{kk}^{-1}\sigma_{ik}$ . Therefore, the partial correlation between  $Y_i$  and  $Y_j$  after  $Y_k$  is the simple correlation between  $Y_i - \beta_{ik} Y_k$  and  $Y_j - \beta_{jk} Y_k$ . Also note that the sign of  $\beta_{ij\cdot k}$  is always equal to the sign of  $\beta_{ji\cdot k}$ . The multiple correlation associated with the regression of  $Y_i$  on  $Y_j$  and  $Y_k$  is denoted by  $R_{i(j,k)}^2$  and is defined by the equation

$$1 - R_{i(j,k)}^2 = (1 - \rho_{ij}^2)(1 - \rho_{ik\cdot j}^2) = (1 - \rho_{ik}^2)(1 - \rho_{ij\cdot k}^2). \quad (1.4.5)$$

Using (1.4.4), the extension of the definition to higher order partial correlations is straightforward. The squared partial correlation between  $Y_i$  and  $Y_l$  after adjusting for  $Y_k$  and  $Y_t$  is

$$\rho_{ij\cdot kl}^2 = \beta_{ij\cdot kl} \beta_{jl\cdot kl}, \quad (1.4.6)$$

where

$$\begin{pmatrix} \beta_{ij \cdot kl} \\ \beta_{ik \cdot jl} \\ \beta_{il \cdot jk} \end{pmatrix} = \begin{pmatrix} \sigma_{jj} & \sigma_{jk} & \sigma_{jl} \\ \sigma_{kj} & \sigma_{kk} & \sigma_{kl} \\ \sigma_{lj} & \sigma_{lk} & \sigma_{ll} \end{pmatrix}^{-1} \begin{pmatrix} \sigma_{ij} \\ \sigma_{ik} \\ \sigma_{il} \end{pmatrix}. \quad (1.4.7)$$

The partial correlation  $\rho_{ij \cdot kl}$  can also be interpreted as the simple correlation between  $Y_i - \beta_{ik \cdot i} Y_k - \beta_{il \cdot i} Y_l$  and  $Y_j - \beta_{jk \cdot i} Y_k - \beta_{jl \cdot i} Y_l$ . The multiple correlation associated with the regression of  $Y_i$  on the set  $(Y_j, Y_k, Y_l)$  satisfies

$$\begin{aligned} 1 - R_{i(j,kl)}^2 &= (1 - \rho_{ij}^2)(1 - \rho_{ik \cdot j}^2)(1 - \rho_{il \cdot jk}^2) \\ &= (1 - \rho_{ik}^2)(1 - \rho_{il \cdot k}^2)(1 - \rho_{ij \cdot kl}^2). \end{aligned} \quad (1.4.8)$$

The squared partial correlation between  $Y_1$  and  $Y_2$  after adjusting for  $(Y_3, Y_4, \dots, Y_p)$  is

$$\rho_{12 \cdot 3,4,\dots,p}^2 = \beta_{12 \cdot 3,4,\dots,p} \beta_{21 \cdot 3,4,\dots,p},$$

where  $\beta_{12 \cdot 3,4,\dots,p}$  is the population regression coefficient for  $Y_2$  in the regression of  $Y_1$  on  $Y_2, Y_3, \dots, Y_p$ , and  $\beta_{21 \cdot 3,4,\dots,p}$  is the population regression coefficient for  $Y_1$  in the regression of  $Y_2$  on  $Y_1, Y_3, Y_4, \dots, Y_p$ . The multiple correlation between  $Y_1$  and the set  $(Y_2, Y_3, \dots, Y_p)$  satisfies the equation

$$1 - R_{1(2,3,\dots,p)}^2 = (1 - \rho_{12}^2)(1 - \rho_{13 \cdot 2}^2)(1 - \rho_{14 \cdot 2,3}^2) \cdots (1 - \rho_{1p \cdot 2,3,\dots,p-1}^2). \quad (1.4.9)$$

For a covariance stationary time series  $Y_t$  with  $R_{1(2,3,\dots,r-1)}^2 < 1$ , the *partial autocorrelation function* is denoted by  $\phi(h)$  and, for  $0 < h \leq r$ , is the partial correlation between  $Y_t$  and  $Y_{t+h}$  after adjusting for  $Y_{t+1}, Y_{t+2}, \dots, Y_{t+h-1}$ . It is understood that  $\phi(0) = 1$  and  $\phi(1) = \rho(1)$ . The partial autocorrelation function is defined in an analogous way for  $h < 0$ . Because  $\rho(h)$  is symmetric,  $\phi(h)$  is symmetric. Let  $\theta_{ih}$  be the population regression coefficient for  $Y_{t-i}$ ,  $1 \leq i \leq h$ , in the regression of  $Y_t$  of  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-h}$ . The population regression equation is

$$Y_t = \theta_{0h} + \theta_{1h} Y_{t-1} + \theta_{2h} Y_{t-2} + \cdots + \theta_{hh} Y_{t-h} + a_{ht}, \quad (1.4.10)$$

where

$$\begin{pmatrix} \theta_{1h} \\ \theta_{2h} \\ \vdots \\ \theta_{hh} \end{pmatrix} = \begin{pmatrix} 1 & \rho(1) & \rho(2) & \cdots & \rho(h-1) \\ \rho(1) & 1 & \rho(1) & \cdots & \rho(h-2) \\ \vdots & \vdots & \vdots & & \vdots \\ \rho(h-1) & \rho(h-2) & \rho(h-3) & \cdots & 1 \end{pmatrix}^{-1} \begin{pmatrix} \rho(1) \\ \rho(2) \\ \vdots \\ \rho(h) \end{pmatrix}, \quad (1.4.11)$$

the matrix of correlations is positive definite by the assumption that  $R_{1(2,3,\dots,h-1)}^2 < 1$ , and  $a_{hi}$  is the population regression residual. From (1.4.11),

$$\phi(h) = \theta_{hh} \quad (1.4.12)$$

because the coefficient for  $Y_{t-h}$  in the regression of  $Y_t$  on  $Y_{t-1}, Y_{t-2}, \dots, Y_{t-h}$  is equal to the coefficient for  $Y_t$  in the regression of  $Y_{t-h}$  on  $Y_{t-h+1}, Y_{t-h+2}, \dots, Y_t$ . The equality follows from the symmetry of the autocorrelation function  $\rho(h)$ . From (1.4.9) we have

$$E\{a_{hi}^2\} = \gamma_r(0) \prod_{i=1}^h [1 - \phi^2(i)]. \quad (1.4.13)$$

## 1.5. COMPLEX VALUED TIME SERIES

Occasionally it is advantageous, from a theoretical point of view, to consider complex valued time series. Letting  $X_t$  and  $Y_t$  be two real valued time series, we define the complex valued time series  $Z_t$  by

$$Z_t = X_t + iY_t. \quad (1.5.1)$$

The expected value of  $Z_t$  is given by

$$E\{Z_t\} = E\{X_t\} + iE\{Y_t\}, \quad (1.5.2)$$

and we note that

$$E^*\{Z_t\} = E\{Z_t^*\} = E\{X_t\} - iE\{Y_t\}, \quad (1.5.3)$$

where the symbol “\*” is used to denote the complex conjugate.

The covariance of  $Z_t$  and  $Z_{t+h}$  is defined as

$$\begin{aligned} \text{Cov}\{Z_t, Z_{t+h}\} &= E\{(Z_t - E\{Z_t\})(Z_{t+h}^* - E\{Z_{t+h}^*\})\} \\ &= \text{Cov}\{X_t, X_{t+h}\} + i\text{Cov}\{Y_t, X_{t+h}\} \\ &\quad - i\text{Cov}\{X_t, Y_{t+h}\} + \text{Cov}\{Y_t, Y_{t+h}\}. \end{aligned} \quad (1.5.4)$$

Note that the variance of a complex valued process is always real and nonnegative, since it is the sum of the variances of two real valued random variables.

The definitions of stationarity for complex time series are completely analogous to those for real time series. Thus, a complex time series  $Z_t$  is weakly stationary if the expected value of  $Z_t$  is a constant for all  $t$  and the covariance matrix of  $(Z_{t_1}, Z_{t_2}, \dots, Z_{t_n})$  is the same as the covariance matrix of  $(Z_{t_1+h}, Z_{t_2+h}, \dots, Z_{t_n+h})$ , where all indices are contained in the index set.

From (1.5.4), the autocovariance of a stationary complex time series  $Z_t$  with zero mean is given by

$$\begin{aligned}\gamma_Z(h) &= [E\{X_t X_{t+h}\} + E\{Y_t Y_{t+h}\}] + i[E\{Y_t X_{t+h}\} - E\{X_t Y_{t+h}\}] \\ &\stackrel{\text{(say)}}{=} g_1(h) + ig_2(h).\end{aligned}\quad (1.5.5)$$

We see that  $g_1(h)$  is a symmetric or even function of  $h$ , and  $g_2(h)$  is an odd function of  $h$ . By the definition of the complex conjugate, we have

$$\gamma_Z^*(h) = \gamma_Z(-h). \quad (1.5.6)$$

Therefore, the autocovariance function of a complex time series is *skew symmetric*, where (1.5.6) is the definition of a skew symmetric function.

A complex valued function  $\gamma(\cdot)$ , defined on the integers, is positive semidefinite if

$$\sum_{j=1}^n \sum_{k=1}^n v_j v_k^* \gamma(t_k - t_j) \geq 0 \quad (1.5.7)$$

for any set of  $n$  complex numbers  $(v_1, v_2, \dots, v_n)$  and any integers  $(t_1, t_2, \dots, t_n)$ . Thus, as in the real valued case, we have:

**Theorem 1.5.1.** The covariance function of a stationary complex valued time series is positive semidefinite.

It follows from Theorem 1.5.1 that the correlation inequality holds for complex random variables; that is,

$$\rho_Z(h)\rho_Z^*(h) = \frac{\gamma_Z(h)\gamma_Z^*(h)}{[\gamma_Z(0)]^2} \leq 1.$$

In the sequel, if we use the simple term "time series," the reader may assume that we are speaking of a real valued time series. All complex valued time series will be identified as such.

## 1.6. PERIODIC FUNCTIONS AND PERIODIC TIME SERIES

Periodic functions play an important role in the analysis of empirical time series. We define a function  $f(t)$  with domain  $T$  to be *periodic* if there exists an  $H > 0$  such that for all  $t$ ,  $t + H \in T$ ,

$$f(t + H) = f(t),$$

where  $H$  is the *period* of the function. That is, the function  $f(t)$  takes on all of its

possible values in an interval of length  $H$ . For any positive integer,  $k$ ,  $kH$  is also a period of the function. While examples of perfectly periodic functions are rare, there are situations where observed time series may be decomposed into the sum of two time series, one of which is periodic or nearly periodic. Even casual observation of many economic time series will disclose seasonal variation wherein peaks and troughs occur at approximately the same month each year. Seasonal variation is apparent in many natural time series, such as the water level in a lake, daily temperature, wind speeds and velocity, and the levels of the tides. Many of these time series also display regular daily variation that has the appearance of rough "cycles."

The trigonometric functions have traditionally been used to approximate periodic behavior. A function obeying the sine-cosine type periodicity is completely specified by three parameters. Thus, we write

$$f(t) = A \sin(\lambda t + \varphi), \quad (1.6.1)$$

where the *amplitude* is the absolute value of  $A$ ,  $\lambda$  is the *frequency*, and  $\varphi$  is the *phase angle*. The frequency is the number of times the function repeats itself in a period of length  $2\pi$ . The phase angle is a "shift parameter" in that it determines the points  $(-\lambda^{-1}\varphi + \text{integer multiples of } \pi)$  where the function is zero. A parametrization that is more useful in the estimation of such a function can be constructed by using the trigonometric identity

$$\sin(\lambda t + \varphi) = \sin \lambda t \cos \varphi + \cos \lambda t \sin \varphi.$$

Thus (1.6.1) becomes

$$f(t) = B_1 \cos \lambda t + B_2 \sin \lambda t, \quad (1.6.2)$$

where  $B_1 = A \sin \varphi$  and  $B_2 = A \cos \varphi$ .

Let us consider a simple type of time series whose realizations will display perfect periodicity. Define  $\{X_t : t \in (0, \pm 1, \pm 2, \dots)\}$  by

$$X_t = e_1 \cos \lambda t + e_2 \sin \lambda t, \quad (1.6.3)$$

where  $e_1$  and  $e_2$  are independent drawings from a normal  $(0, 1)$  population. Note that the realization is completely determined by the two random variables  $e_1$  and  $e_2$ . The amplitude and phase angle vary from realization to realization, but the period is the same for every realization.

The stochastic properties of this time series are easily derived:

$$\begin{aligned} E\{X_t\} &= E\{e_1 \cos \lambda t + e_2 \sin \lambda t\} \\ &= 0; \\ \gamma(h) &= E\{(e_1 \cos \lambda t + e_2 \sin \lambda t)[e_1 \cos \lambda(t+h) + e_2 \sin \lambda(t+h)]\} \\ &= E\{e_1^2 \cos \lambda t \cos \lambda(t+h) + e_1 e_2 \cos \lambda t \sin \lambda(t+h) + e_1 e_2 \sin \lambda t \cos \lambda(t+h) + e_2^2 \sin \lambda t \sin \lambda(t+h)\} \end{aligned} \quad (1.6.4)$$

$$\begin{aligned}
 & + e_1 e_2 \cos \lambda t \sin \lambda(t+h) + e_2 e_1 \sin \lambda t \cos \lambda(t+h) \\
 & = \cos \lambda t \cos \lambda(t+h) + \sin \lambda t \sin \lambda(t+h) \\
 & = \cos \lambda h.
 \end{aligned}$$

We see that the process is stationary. This example also serves to emphasize the fact that the covariance function is obtained by averaging the product  $X_t X_{t+h}$  over realizations.

It is clear that any time series defined by the finite sum

$$X_t = \sum_{j=1}^M e_{1j} \cos \lambda_j t + \sum_{j=1}^M e_{2j} \sin \lambda_j t, \quad (1.6.5)$$

where  $e_{1j}$  and  $e_{2j}$  are independent drawings from a normal  $(0, \sigma_j^2)$  population and  $\lambda_j \neq \lambda_k$  for  $k \neq j$ , will display periodic behavior. The representation (1.6.5) is a useful approximation for portions of some empirical time series, and we shall see that ideas associated with this representation are important for the theoretical study of time series as well.

## 1.7. VECTOR VALUED TIME SERIES

Most of the concepts associated with univariate time series have immediate generalizations to vector valued time series. We now introduce representations for multivariate processes.

The  $k$ -dimensional time series  $\{\mathbf{X}_t : t = 0, \pm 1, \pm 2, \dots\}$  is defined by

$$\mathbf{X}_t = [X_{1t}, X_{2t}, \dots, X_{kt}]', \quad (1.7.1)$$

where  $\{X_{it} : t = 0, \pm 1, \pm 2, \dots\}$ ,  $i = 1, 2, \dots, k$ , are scalar time series. The expected value of  $\mathbf{X}'$  is

$$E\{\mathbf{X}'\} = [E\{X_{1t}\}, E\{X_{2t}\}, \dots, E\{X_{kt}\}].$$

Assuming the mean is zero, we define the covariance matrix of  $\mathbf{X}_t$  and  $\mathbf{X}_{t+h}$  by

$$E\{\mathbf{X}_t \mathbf{X}'_{t+h}\} = \begin{bmatrix} E\{X_{1t} X_{1,t+h}\} & E\{X_{1t} X_{2,t+h}\} & \cdots & E\{X_{1t} X_{k,t+h}\} \\ E\{X_{2t} X_{1,t+h}\} & E\{X_{2t} X_{2,t+h}\} & \cdots & E\{X_{2t} X_{k,t+h}\} \\ \vdots & \vdots & \vdots & \vdots \\ E\{X_{kt} X_{1,t+h}\} & E\{X_{kt} X_{2,t+h}\} & \cdots & E\{X_{kt} X_{k,t+h}\} \end{bmatrix}. \quad (1.7.2)$$

As with scalar time series, we define  $\mathbf{X}_t$  to be covariance stationary if:

1. The expected value of  $\mathbf{X}_t$  is a constant function of time.
2. The covariance matrix of  $\mathbf{X}_t$  and  $\mathbf{X}_{t+h}$  is the same as the covariance matrix of  $\mathbf{X}_j$  and  $\mathbf{X}_{j+h}$  for all  $t, t+h, j, j+h$  in the index set.

If a vector time series is stationary, then every component scalar time series is stationary. However, a vector of scalar stationary time series is not necessarily a vector stationary time series.

The second stationarity condition means that we can express the covariance matrix as a function of  $h$  only, and, assuming  $E\{\mathbf{X}_t\} = 0$ , we write

$$\Gamma(h) = E\{\mathbf{X}_t \mathbf{X}'_{t+h}\}$$

for stationary time series. Note that the diagonal elements of this matrix are the autocovariances of the  $X_{jt}$ . The off-diagonal elements are the *cross covariances* of  $X_{it}$  and  $X_{jt}$ . The element  $E\{X_{it} X_{j,t+h}\}$  is not necessarily equal to  $E\{X_{i,t+h} X_{jt}\}$ , and hence  $\Gamma(h)$  is not necessarily equal to  $\Gamma(-h)$ . For example, let

$$\begin{aligned} X_{1t} &= e_t, \\ X_{2t} &= e_t + \beta e_{t-1}. \end{aligned} \tag{1.7.3}$$

Then

$$\begin{aligned} \gamma_{12}(1) &= E\{X_{1t} X_{2,t+1}\} = \beta \sigma^2, \\ \gamma_{12}(-1) &= E\{X_{1,t+1} X_{2t}\} = 0. \end{aligned}$$

However,

$$\begin{aligned} \gamma_{21}(1) &= E\{X_{2t} X_{1,t+1}\} = 0 = \gamma_{12}(-1), \\ \gamma_{21}(-1) &= E\{X_{2t} X_{1,t-1}\} = \beta \sigma^2 = \gamma_{12}(1), \end{aligned}$$

and it is clear that

$$\Gamma(h) = \Gamma'(-h). \tag{1.7.4}$$

It is easy to verify that (1.7.4) holds for all vector stationary time series, and we state the result as a lemma.

**Lemma 1.7.1.** The autocovariance matrix of a vector stationary time series satisfies  $\Gamma(h) = \Gamma'(-h)$ .

The nonnegative definite property of the scalar autocovariance function is maintained essentially unchanged for vector processes.

**Lemma 1.7.2.** The covariance function of a vector stationary time series  $\{\mathbf{X}_t : t \in T\}$  is a nonnegative definite function in that

$$\sum_{j=1}^n \sum_{i=1}^n \mathbf{a}'_j \boldsymbol{\Gamma}(t_i - t_j) \mathbf{a}_i \geq 0$$

for any set of real vectors  $\{\mathbf{a}_1, \mathbf{a}_2, \dots, \mathbf{a}_n\}$  and any set of indices  $\{t_1, t_2, \dots, t_n\} \in T$ .

**Proof.** The result can be obtained by evaluating the variance of

$$\sum_{j=1}^n \mathbf{a}'_j \mathbf{X}_{t_j} . \quad \blacktriangle$$

We define the *correlation matrix* of  $\mathbf{X}_t$  and  $\mathbf{X}_{t+h}$  by

$$\mathbf{P}(h) = \mathbf{D}_0^{-1} \boldsymbol{\Gamma}(h) \mathbf{D}_0^{-1}, \quad (1.7.5)$$

where  $\mathbf{D}_0$  is a diagonal matrix with the square root of the variances of the  $X_{jt}$  as diagonal elements; that is,

$$\mathbf{D}_0^2 = \text{diag}\{\gamma_{11}(0), \gamma_{22}(0), \dots, \gamma_{kk}(0)\}.$$

The  $ij$ th element of  $\mathbf{P}(h)$ ,  $\rho_{ij}(h)$ , is called the *cross correlation* of  $X_{it}$  and  $X_{jt}$ .

For the time series of (1.7.3) we have

$$\mathbf{P}(h) = \begin{cases} \begin{pmatrix} 1 & (1+\beta^2)^{-1/2} \\ (1+\beta^2)^{-1/2} & 1 \end{pmatrix}, & h=0, \\ \begin{pmatrix} 0 & (1+\beta^2)^{-1/2}\beta \\ 0 & (1+\beta^2)^{-1}\beta \end{pmatrix}, & h=1, \\ \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix}, & |h|>1. \end{cases}$$

## REFERENCES

Chung (1968), Gnedenko (1967), Loève (1963), Rao (1973), Tucker (1967), Yaglom (1962).

## EXERCISES

1. Determine the mean and covariance function for Example 1 of Section 1.3.
2. Discuss the stationarity of the following time series:

$$\begin{aligned}
 \text{(a) } \{X_t : t \in (0, \pm 1, \pm 2, \dots)\} &= \left. \begin{array}{l} \text{Value of a randomly chosen} \\ \text{observation from a normal} \\ \text{distribution with mean } \frac{1}{2} \\ \text{and variance } \frac{1}{4} \end{array} \right\} t \text{ odd} \\
 &= \left. \begin{array}{l} = 1 \text{ if toss of a true coin} \\ \text{results in a head} \\ = 0 \text{ if toss of a true coin} \\ \text{results in a tail} \end{array} \right\} t \text{ even}
 \end{aligned}$$

(b)  $\{X_t : t \in (0, \pm 1, \pm 2, \dots)\}$  is a time series of independent identically distributed random variables whose distribution function is that of Student's  $t$ -distribution with one degree of freedom.

$$\text{(c) } X_t = \begin{cases} ce_0, & t = 0, \\ \rho X_{t-1} + e_t, & t = 1, 2, \dots, \end{cases}$$

where  $c$  is a constant,  $|\rho| < 1$ , and the  $e_t$  are iid(0, 1) random variables.

3. Which of the following processes is covariance stationary for  $T = \{t : t \in (0, 1, 2, \dots)\}$ , where the  $e_t$  are independent identically distributed (0, 1) random variables and  $a_1, a_2$  are fixed real numbers?
- (a)  $e_1 + e_2 \cos t$ .
  - (b)  $e_1 + e_2 \cos t + e_3 \sin t$ .
  - (c)  $a_1 + e_1 \cos t$ .
  - (d)  $a_1 + e_1 \cos t + e_2 \sin t$ .
  - (e)  $e_t + a_1 \cos t$ .
  - (f)  $a_1 + e_1 a_2' + e_2, 0 < a_2 < 1$ .

4. (a) Prove the following theorem.

**Theorem.** If  $X_t$  and  $Y_t$  are independent covariance stationary time series, then  $aX_t + bY_t$ , where  $a$  and  $b$  are real numbers, is a covariance stationary time series.

- (b) Let  $Z_t = X_t + Y_t$ . Give an example to show that  $Z_t$  covariance stationary does not imply  $X_t$  is covariance stationary.

5. Give a covariance stationary time series such that  $\rho(h) \neq 0$  for all  $h$ .
6. Which of the following functions is the covariance function of a stationary time series? Explain why or why not.

(a)  $g(h) = 1 + |h|, \quad h = 0, \pm 1, \pm 2, \dots;$

(b)  $g(h) = \begin{cases} 1, & h = 0, \\ -\frac{1}{2}, & h = \pm 1, \\ 0 & \text{otherwise;} \end{cases}$

(c)  $g(h) = 1 + \frac{1}{4} \sin 4h, \quad h = 0, \pm 1, \pm 2, \dots;$

(d)  $g(h) = 1 + \frac{1}{4} \cos 4h, \quad h = 0, \pm 1, \pm 2, \dots;$

(e)  $g(h) = \begin{cases} 1, & h = 0, \pm 1, \\ 0 & \text{otherwise.} \end{cases}$

7. Let  $e_t$  be time series of normal independent  $(0, \sigma^2)$  random variables. Define

$$Y_t = \begin{cases} e_t, & \text{if } t \text{ is odd,} \\ -e_t, & \text{if } t \text{ is even.} \end{cases}$$

Is the complex valued time series  $Z_t = e_t + iY_t$  stationary?

8. Let

$$\rho(h) = \begin{cases} 1, & h = 0, \\ a, & h = \pm 1, \\ 0 & \text{otherwise.} \end{cases}$$

- (a) For what values of  $a$  is  $\rho(h)$  the autocorrelation function of a stationary time series?  
 (b) Compute the partial autocorrelation function associated with  $\rho(h)$ , given that  $a$  is such that  $\rho(h)$  is an autocorrelation function.

9. Let the time series  $Y_t$  be defined by

$$Y_t = e_1 \cos 0.5\pi t + e_2 \sin 0.5\pi t + e_t$$

for  $t = 3, 4, \dots$ , where  $e_t \sim NI(0, 1)$  and  $\sim NI(0, 1)$  denotes normally and independently distributed with zero mean and unit variance. Compute the partial autocorrelation function of  $Y_t$  for  $h = 1, 2, 3, 4$ .

10. Show that if  $\beta_{12,3}$  of (1.4.4) is zero, then  $\beta_{21,3}$  is also zero. Extend this result to the  $p$ th order partial regression coefficients  $\beta_{1,p-2,\dots,p-1}$  and  $\beta_{p+2,\dots,p-1}$ .  
 11. Assume that for the stationary time series  $Y_t$ , there is some  $r$  such that  $R_{1(2,3,\dots,r)}^2 = 1$ , and  $R_{1(2,3,\dots,r-1)}^2 < 1$ , where  $R_{1(2,3,\dots,r)}^2$  is the squared multiple correlation between  $Y_1$  and  $(Y_2, \dots, Y_r)$ . Show that  $|\phi(r-1)| = 1$  and that the correlation matrix in (1.4.11) is singular for  $h \geq r$ .

12. Let  $\{Z_t\}$  be a sequence of normal independent  $(0, 1)$  random variables defined on the integers.

(a) Give the mean and covariance function of the time series (i)  $Y_t = Z_t Z_{t-1}$  and (ii)  $X_t = Z_t + Z_{t-1}$ .

(b) Is the time series

$$W_t = \begin{cases} Z_t, & t \text{ even}, \\ 2^{-1/2}(Z_t^2 - 1), & t \text{ odd} \end{cases}$$

covariance stationary? Is it strictly stationary?

## CHAPTER 2

# Moving Average and Autoregressive Processes

In any study, one of the first undertakings is to describe the item under investigation. Typically there is no unique description and, as the discipline develops, alternatives appear that prove more useful in some applications.

The joint distribution function (1.2.1) gives a complete description of the distributional properties of a time series. The covariance function provides a less complete description, but one from which useful conclusions can be drawn. As we have noted, the correlation function has functional properties analogous to those of the characteristic function of a statistical distribution function. Therefore, the distribution function associated with the correlation function provides an alternative representation for a time series. This representation will be introduced in Chapter 3 and discussed in Chapter 4.

Another method of describing a time series is to express it as a function of more elementary time series. A sequence of independent identically distributed random variables is a very simple type of time series. In applications it is often possible to express the observed time series as a relatively simple function of this elementary time series. In particular, many time series have been represented as linear combinations of uncorrelated  $(0, \sigma^2)$  or of independent  $(0, \sigma^2)$  random variables. We now consider such representations.

### 2.1. MOVING AVERAGE PROCESSES

We shall call the time series  $\{X_t : t \in (0, \pm 1, \pm 2, \dots)\}$ , defined by

$$X_t = \sum_{j=-M}^M \alpha_j e_{t-j}, \quad (2.1.1)$$

where  $M$  is a nonnegative integer,  $\alpha_j$  are real numbers,  $\alpha_{-M} \neq 0$ , and the  $e_j$  are uncorrelated  $(0, \sigma^2)$  random variables, a *finite moving average time series* or a *finite moving average process*.

If there exists no finite  $M$  such that  $\alpha_j = 0$  for all  $j$  with absolute value greater than  $M$ , then  $\{X_t\}$  is an *infinite moving average process* and is given the representation

$$X_t = \sum_{j=-\infty}^{\infty} \alpha_j e_{t-j}. \quad (2.1.2)$$

The time series defined by

$$X_t = \sum_{j=0}^M \alpha_j e_{t-j}, \quad (2.1.3)$$

where  $\alpha_0 \neq 0$  and  $\alpha_M \neq 0$ , is called a *one-sided moving average of order  $M$* . Note that the number of terms in the sum on the right-hand side of (2.1.3) is  $M + 1$ . In particular, (2.1.3) is a *left-moving average*, since nonzero weights are applied only to  $e_t$  and to  $e_s$ 's with indexes less than  $t$ . We shall most frequently use the one-sided representation of moving average time series. Note that there would be no loss of generality in considering only the one-sided representation. If in the representation (2.1.1) we define the random variable  $\epsilon_i$  by  $\epsilon_i = e_{i+M}$ , we have, for (2.1.1),

$$\begin{aligned} X_t &= \sum_{j=-M}^M \alpha_j e_{t-j} = \sum_{s=0}^{2M} \alpha_{s-M} \epsilon_{t-s} \\ &= \sum_{s=0}^{2M} \beta_s \epsilon_{t-s}, \end{aligned} \quad (2.1.4)$$

where  $\beta_s = \alpha_{s-M}$ .

Likewise we can, without loss of generality, take  $\beta_0$  in (2.1.4) to be one. If  $\beta_0$  is not one, we simply define

$$\begin{aligned} u_{t-s} &= \beta_0 \epsilon_{t-s}, \\ \delta_s &= \beta_s \beta_0^{-1}, \quad s = 0, 1, 2, \dots, 2M, \end{aligned}$$

and obtain the representation

$$X_t = \sum_{s=0}^{2M} \delta_s u_{t-s}, \quad (2.1.5)$$

where  $\delta_0 = 1$  and the  $u_s$  are uncorrelated  $(0, \beta_0^2 \sigma^2)$  random variables.

The covariance function of an  $M$ th order moving average is distinctive in that it is zero for all  $h$  satisfying  $|h| > M$ . For the time series (2.1.3) we have

$$\begin{aligned} E\{X_t\} &= 0, \\ \gamma_X(h) &= E\{X_t X_{t+h}\} = \begin{cases} \sum_{i=0}^{M-|h|} \alpha_i \alpha_{i+|h|} \sigma^2, & 0 \leq |h| \leq M, \\ 0, & |h| > M. \end{cases} \end{aligned} \quad (2.1.6)$$

As an example, consider the simple time series

$$X_t = \sum_{i=0}^4 e_{t-i}.$$

Clearly,

$$E\{X_t\} = \sum_{i=0}^4 E\{e_{t-i}\} = 0 \quad \text{for all } t$$

and

$$\begin{aligned} E\{X_t X_{t+h}\} &= E\left\{\left(\sum_{i=0}^4 e_{t-i}\right)\left(\sum_{i=0}^4 e_{t+h-i}\right)\right\} \\ &= \begin{cases} (5 - |h|)\sigma^2, & |h| \leq 4, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

By (2.1.6) only the first autocorrelation of a first order moving average is nonzero, while all autocorrelations of order greater than two for a second order moving average are zero. For a  $q$ th order moving average,  $\rho(q) \neq 0$ , and all higher order autocorrelations are zero. There are further restrictions on the correlation function of moving average time series. Consider the first order moving average

$$X_t = e_t + \alpha e_{t-1}. \quad (2.1.7)$$

Then

$$\rho(1) = \frac{\alpha}{1 + \alpha^2}. \quad (2.1.8)$$

Using elementary calculus, it is easy to prove that  $\rho(1)$  achieves a maximum of 0.5 for  $\alpha = 1$  and a minimum of -0.5 for  $\alpha = -1$ .

Thus, the first autocorrelation of a first order moving average process will always fall in the interval  $[-0.5, 0.5]$ . For any  $\rho$  in  $(0, 0.5)$  [or in  $(-0.5, 0)$ ] there are two  $\alpha$ -values yielding such a  $\rho$ , one in the interval  $(0, 1)$  [or in  $(-1, 0)$ ] and one in the interval  $(1, \infty)$  [or  $(-\infty, -1)$ ].

**Example 2.1.1.** In Figure 2.1.1 we display 100 observations from a realization of the time series

$$X_t = e_t + 0.7e_{t-1},$$

where the  $e_t$  are normal independent  $(0, 1)$  random variables. The positive correlation between adjacent observations is clear to the eye and is emphasized by the plot of  $X_t$  against  $X_{t-1}$  of Figure 2.1.2. The correlation between  $X_t$  and  $X_{t-1}$  is  $(1.49)^{-1}(0.7) = 0.4698$ , and this is the slope of the line plotted on the figure. On the other hand, since the  $e_t$  are independent,  $X_t$  is independent of  $X_{t-2}$ , and this is clear from the plot of  $X_t$  against  $X_{t-2}$  in Figure 2.1.3.

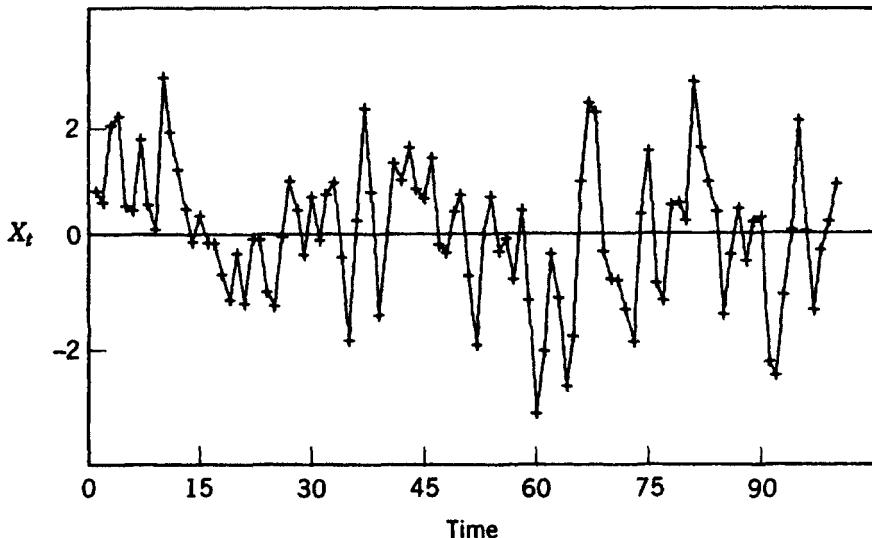


Figure 2.1.1. One hundred observations from the time series  $X_t = e_t + 0.7e_{t-1}$ .

While the simple correlation between  $X_t$  and  $X_{t-2}$  is zero, the partial correlation between  $X_t$  and  $X_{t-2}$  adjusted for  $X_{t-1}$  is not zero. The correlation between  $X_t$  and  $X_{t-1}$  is  $\rho(1) = (1 + \alpha^2)^{-1}\alpha$ , which is the same as that between  $X_{t-1}$  and  $X_{t-2}$ . Therefore, the partial correlation between  $X_t$  and  $X_{t-2}$  adjusted for  $X_{t-1}$  is

$$\frac{\text{Cov}\{X_t - cX_{t-1}, X_{t-2} - cX_{t-1}\}}{\text{Var}\{X_t - cX_{t-1}\}} = -\frac{\alpha^2}{1 + \alpha^2 + \alpha^4},$$

where  $c = (1 + \alpha^2)^{-1}\alpha$ .

In Figure 2.1.4,  $X_t - 0.7(1.49)^{-1}X_{t-1}$  is plotted against  $X_{t-2} - 0.7(1.49)^{-1}X_{t-1}$  for the time series of Figure 2.1.1. For this time series the theoretical partial autocorrelation is  $-0.283$ , and this is the slope of the line plotted in Figure 2.1.4.

▲▲

**Example 2.1.2.** One hundred observations from the second order moving average

$$X_t = e_t - 1.40e_{t-1} + 0.48e_{t-2}$$

are displayed in Figure 2.1.5. The correlation function of this process is

$$\begin{aligned} \rho(h) &\doteq -0.65, & h = 1, \\ &\doteq 0.15, & h = 2, \\ &= 0, & h \geq 3. \end{aligned}$$

Once again the nature of the correlation can be observed from the plot. Adjacent

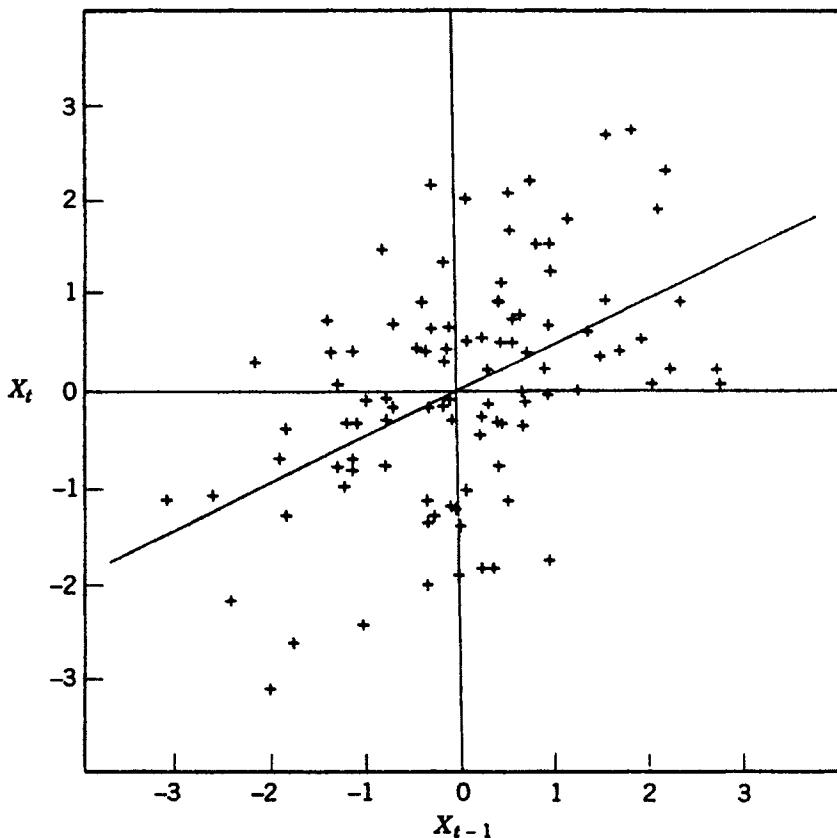


Figure 2.1.2. Plot of  $X_t$  against  $X_{t-1}$  for 100 observations from  $X_t = \epsilon_t + 0.7\epsilon_{t-1}$ .

observations are often on opposite sides of the mean, indicating that the first order autocorrelation is negative. This correlation is also apparent in Figure 2.1.6, where  $X_t$  is plotted against  $X_{t-1}$ . ▲▲

At the beginning of this section we defined a moving average time series to be a linear combination of a sequence of uncorrelated random variables. It is also common to speak of forming a moving average of an arbitrary time series  $X_t$ . Thus,  $Y_t$  defined by

$$Y_t = \sum_{i=-M}^M \alpha_i X_{t+i}$$

is a moving average of the time series  $X_t$ . The reader may prove:

**Lemma 2.1.1.** If  $\{X_t : t \in (0, \pm 1, \pm 2, \dots)\}$  is a stationary time series with mean zero and covariance function  $\gamma_X(h)$ , then

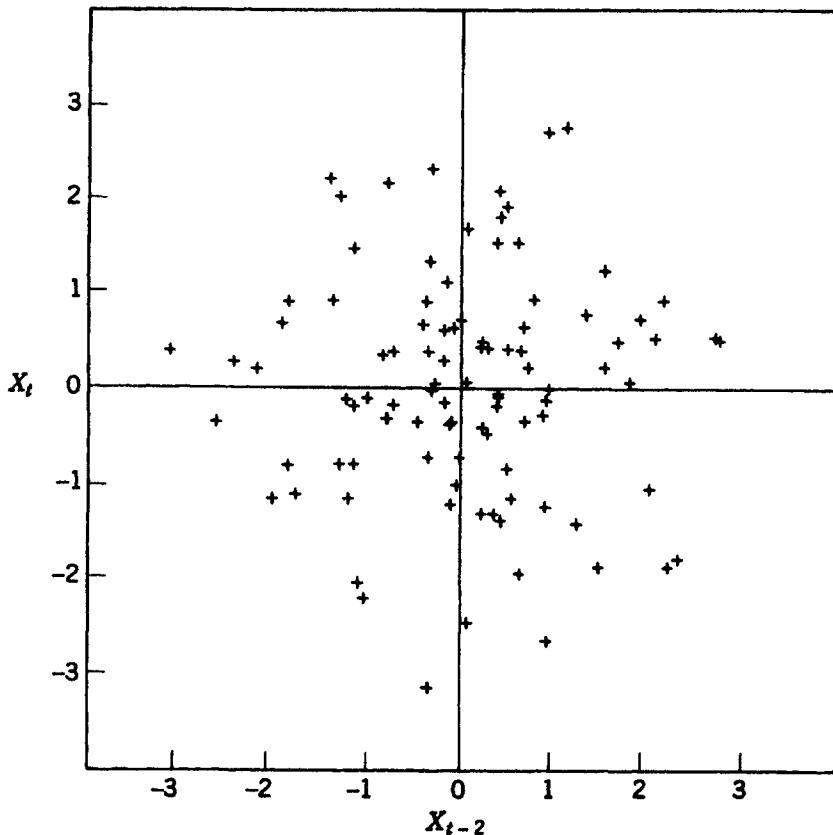


Figure 2.1.3. Plot of  $X_t$  against  $X_{t-2}$  for 100 observations from  $X_t = e_t + 0.7e_{t-1}$ .

$$Y_t = \sum_{i=-M}^M \alpha_i X_{t+i},$$

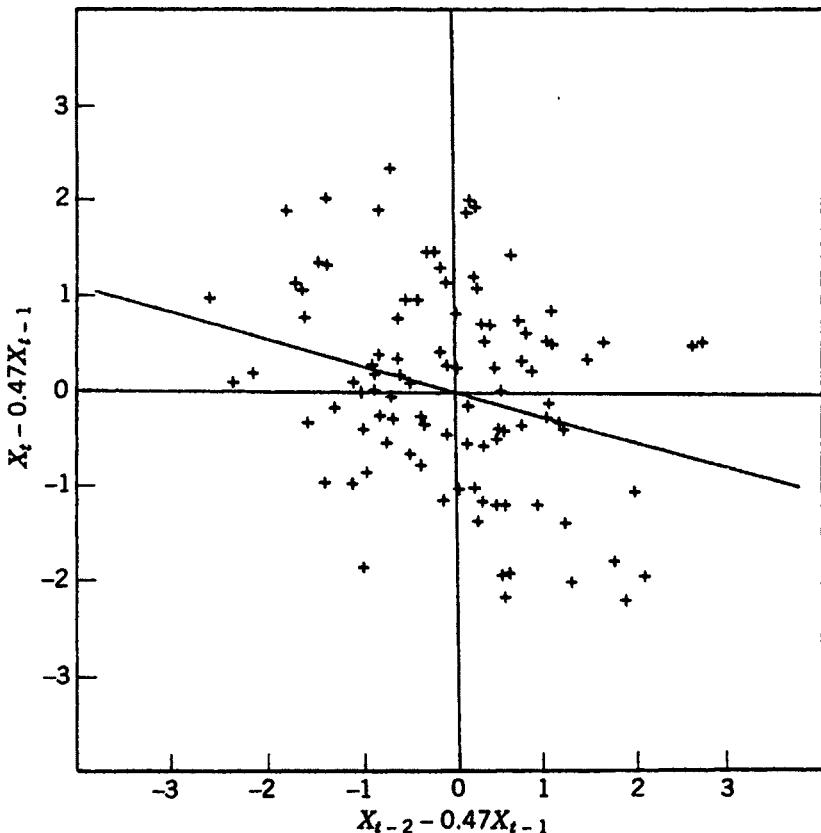
where  $M$  is finite and the  $\alpha_i$  are real numbers, is a stationary time series with mean zero and covariance function

$$\gamma_Y(h) = \sum_{i=-M}^M \sum_{j=-M}^M \alpha_i \alpha_j \gamma_X(h+j-i).$$

## 2.2. ABSOLUTELY SUMMABLE SEQUENCES AND INFINITE MOVING AVERAGES

Before investigating the properties of infinite moving average time series, we review some results on the convergence of partial sums of infinite sequences of real numbers.

Recall that an *infinite sequence* is a function whose domain is the set



**Figure 2.1.4.** Plot of  $X_t - 0.47X_{t-1}$  against  $X_{t-2} - 0.47X_{t-1}$ , for 100 observations from  $X_t = e_t + 0.7e_{t-1}$ .

$T_1 = (1, 2, \dots)$  of natural numbers. A *doubly infinite sequence* is a function whose domain is the set  $T_2 = (0, \pm 1, \pm 2, \dots)$  of all integers. When no confusion will result, we shall use the word sequence for a function whose domain is either  $T_1$  or  $T_2$ . We denote the infinite sequence by  $\{a_j\}_{j=1}^{\infty}$  and the doubly infinite sequence by  $\{a_j\}_{j=-\infty}^{\infty}$ . When the domain is clear, the notation will be abbreviated to  $\{a_j\}$  or perhaps to  $a_j$ .

The *infinite series* created from the sequence  $\{a_j\}_{j=1}^{\infty}$  is the sequence of partial sums  $\{s_j\}_{j=1}^{\infty}$  defined by

$$s_j = \sum_{i=1}^j a_i, \quad j = 1, 2, \dots \quad (2.2.1)$$

The symbols  $\sum_{j=1}^{\infty} a_j$  and  $\Sigma\{a_j\}$  are often used to represent both the infinite sequence  $\{s_j\}_{j=1}^{\infty}$  generated by the sequence  $\{a_j\}$  and the limit

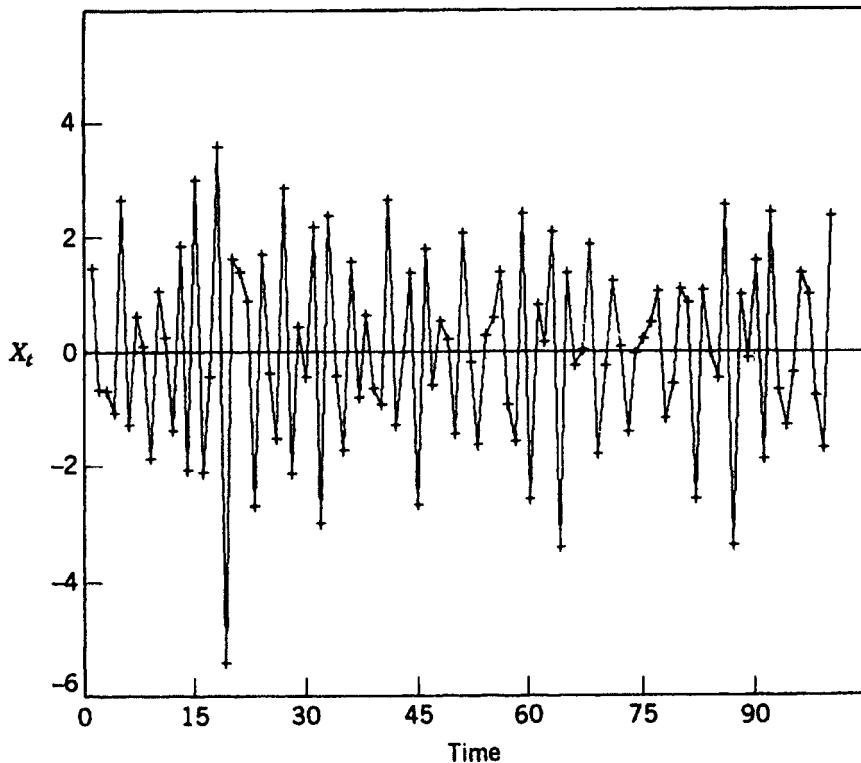


Figure 2.1.5. One hundred observations from the time series  $X_t = e_t - 1.40e_{t-1} + 0.48e_{t-2}$ .

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n a_j = \lim_{n \rightarrow \infty} s_n$$

when the limit is defined. If the sequence  $\{s_j\}$  has a finite limit, we say that the series  $\Sigma\{a_j\}$  is *convergent*. If the limit is not defined, the series is *divergent*.

If the limit

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n |a_j|$$

is finite, the series  $\Sigma\{a_j\}$  is said to be *absolutely convergent*, and we write  $\sum_{j=1}^{\infty} |a_j| < \infty$ . We also describe this situation by saying the sequence  $\{a_j\}$  is *absolutely summable*. For a doubly infinite sequence, if the limit

$$\lim_{n \rightarrow \infty} \sum_{j=-n}^n |a_j|$$

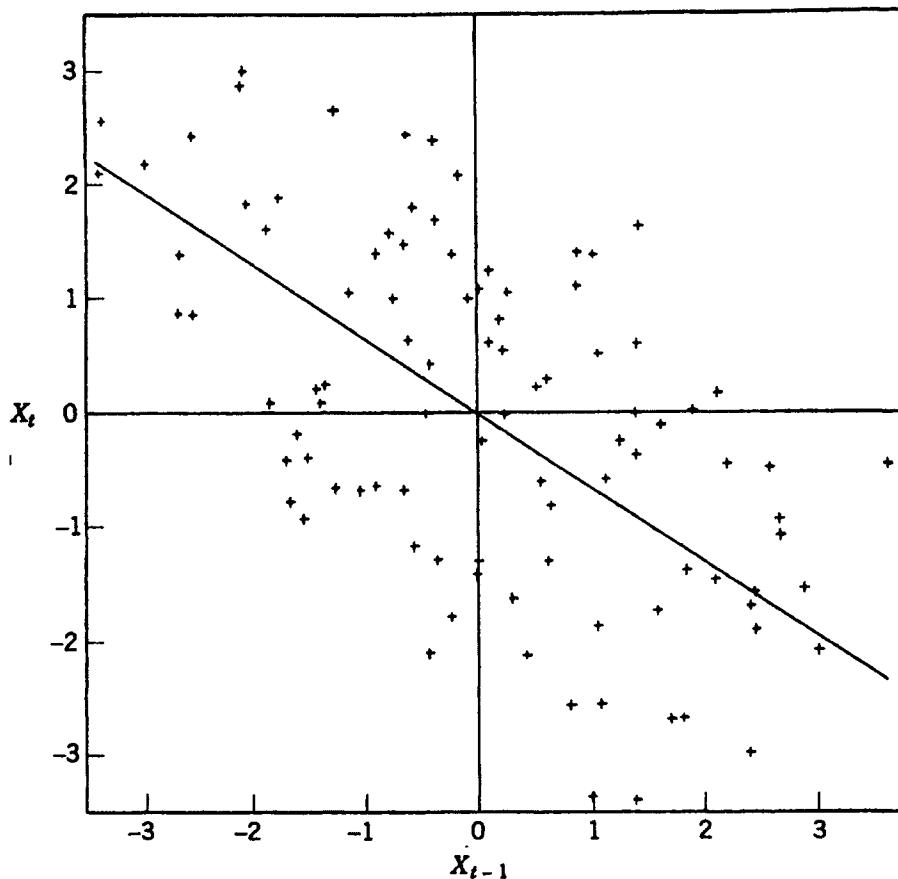


Figure 2.1.6. Plot of  $X_t$  against  $X_{t-1}$  for 100 observations from  $X_t = e_t - 1.40e_{t-1} + 0.48e_{t-2}$ .

exists and is finite, then the series is absolutely convergent, and we write

$$\sum_{j=-\infty}^{\infty} |a_j| < \infty$$

and say that the sequence  $\{a_j\}$  is absolutely summable.

The following elementary properties of absolutely summable sequences will be useful in our investigation of moving average time series.

1. If  $\{a_j\}$  is absolutely summable, then

$$\sum_{j=-\infty}^{\infty} a_j^2 < \infty.$$

The reader can verify that the converse is not true by considering  $\sum_{j=1}^{\infty} j^{-2}$  and  $\sum_{j=1}^{\infty} j^{-1}$ .

2. Given two absolutely summable sequences  $\{a_j\}$  and  $\{b_j\}$ , then the sequences  $\{a_j + b_j\}$  and  $\{a_j b_j\}$  are absolutely summable:

$$\begin{aligned}\sum_{j=-\infty}^{\infty} |a_j + b_j| &\leq \sum_{j=-\infty}^{\infty} (|a_j| + |b_j|) = \sum_{j=-\infty}^{\infty} |a_j| + \sum_{j=-\infty}^{\infty} |b_j| < \infty, \\ \sum_{j=-\infty}^{\infty} |a_j b_j| &= \sum_{j=-\infty}^{\infty} (|a_j| |b_j|) \leq \sum_{j=-\infty}^{\infty} (|a_j| + |b_j|)^2 < \infty.\end{aligned}$$

Of course,  $\sum_{j=-\infty}^{\infty} (|a_j| + |b_j|)^2 < \infty$ , since  $\{(|a_j| + |b_j|)^2\}$  is the square of an absolutely summable sequence.

3. The *convolution* of two absolutely summable sequences  $\{a_j\}$  and  $\{b_j\}$ , defined by

$$c_j = \sum_{k=-\infty}^{\infty} a_k b_{j-k},$$

is absolutely summable:

$$\sum_{j=-\infty}^{\infty} |c_j| \leq \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |a_k| |b_{j-k}| \leq \sum_{k=-\infty}^{\infty} |a_k| \sum_{j=-\infty}^{\infty} |b_j| < \infty.$$

This result generalizes, and, for example, if we define

$$d_j = \sum_{k=-\infty}^{\infty} a_k \sum_{m=-\infty}^{\infty} b_m c_{j-k-m},$$

where  $\{a_j\}$ ,  $\{b_j\}$ , and  $\{c_j\}$  are absolutely summable, then

$$\sum_{j=-\infty}^{\infty} |d_j| < \infty.$$

The infinite moving average time series

$$X_t = \sum_{j=-\infty}^{\infty} \alpha_j e_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots,$$

was introduced in equation (2.1.2) of Section 2.1. In investigating the behavior of infinite moving averages we will repeatedly interchange the summation and expectation operations. As justification for this procedure, we give the following theorems. The theorems may be skipped by those not familiar with real analysis. The organization of the proofs follows suggestions made by Torres (1986) and Pantula (1988a). We begin with a lemma on almost sure convergence of a sequence of random functions.

**Lemma 2.2.1.** Let  $\{Z_j\}$  be a sequence of random variables defined on the probability space  $(\Omega, \mathcal{A}, P)$ . Assume

$$\sum_{j=-\infty}^{\infty} E\{|Z_j|\} < \infty.$$

Then  $\sum_{j=-\infty}^{\infty} Z_j$  is defined as an almost sure limit and

$$E\left\{ \sum_{j=-\infty}^{\infty} Z_j \right\} = \sum_{j=-\infty}^{\infty} E\{Z_j\}.$$

**Proof.** Let a nondecreasing sequence of random variables be defined by

$$Y_k(\omega) = \sum_{j=-k}^k |Z_j(\omega)|.$$

Therefore, the limit as  $k \rightarrow \infty$  is defined. By assumption,  $E\{|Z_j|\}$  is finite for each  $j$  and it follows that

$$E\{Y_k\} < M < \infty$$

for some  $M$  and all  $k$ . Therefore, by the monotone convergence theorem,

$$E\left\{ \sum_{j=-\infty}^{\infty} |Z_j| \right\} = \sum_{j=-\infty}^{\infty} E\{|Z_j|\} < \infty.$$

See, for example, Royden (1989, p. 265). Also  $\sum_{j=-\infty}^{\infty} |Z_j| < \infty$  a.s. Therefore,

$$\lim_{n \rightarrow \infty} S_n = \sum_{j=-\infty}^{\infty} Z_j \quad \text{a.s.}$$

and

$$|S_n| < \sum_{j=-\infty}^{\infty} |Z_j| < \infty \quad \text{a.s.,}$$

where  $S_n = \sum_{j=-n}^n Z_j$ . It follows, by the dominated convergence theorem, that

$$E\left\{ \sum_{j=-\infty}^{\infty} Z_j \right\} = \sum_{j=-\infty}^{\infty} E\{Z_j\}. \quad \blacktriangle$$

**Theorem 2.2.1.** Let the sequence of real numbers  $\{a_j\}$  and the sequence of random variables  $\{Z_j\}$  satisfy

$$\sum_{j=-\infty}^{\infty} |a_j| = M < \infty$$

and  $E\{Z_t^2\} \leq K$ ,  $t = 0, \pm 1, \pm 2, \dots$ , for some finite  $K$ . Then there exists a sequence of random variables  $\{X_t\}$  such that, for  $t = 0, \pm 1, \pm 2, \dots$ ,

$$X_t = \sum_{j=-\infty}^{\infty} a_j Z_{t-j} \quad \text{a.s.,}$$

$$\lim_{n \rightarrow \infty} E\left\{ \left| X_t - \sum_{j=-n}^n a_j Z_{t-j} \right|^2 \right\} = 0,$$

and  $E\{X_t^2\} \leq M^2 K$ .

**Proof.** By Lemma 2.2.1,  $X_t$  is well defined as an almost sure limit and

$$E\left\{ \sum_{j=-\infty}^{\infty} a_j Z_j \right\} = \sum_{j=-\infty}^{\infty} E\{a_j Z_j\}.$$

Because  $(|Z_t| - |Z_j|)^2 \geq 0$ , we have  $E\{|Z_t| |Z_j|\} \leq K$  for all  $t, j$ . It follows that

$$\sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} E\{|a_i Z_{t-i} a_j Z_{t-j}|\} \leq K \left( \sum_{j=-\infty}^{\infty} |a_j| \right)^2.$$

Letting

$$T_n = \sum_{i=-n}^n \sum_{j=-n}^n a_i a_j Z_{t-i} Z_{t-j}$$

and applying Lemma 2.2.1, we have that

$$T_n \rightarrow \left( \sum_{i=-\infty}^{\infty} a_i Z_{t-i} \right)^2 \quad \text{a.s.}$$

and

$$E\left\{ \left( \sum_{i=-\infty}^{\infty} a_i Z_{t-i} \right)^2 \right\} \leq K \left( \sum_{j=-\infty}^{\infty} |a_j| \right)^2$$

for all  $t = 0, \pm 1, \pm 2, \dots$ . It follows that

$$E\left[ \left| X_t - \sum_{j=-n}^n a_j Z_{t-j} \right|^2 \right] = E\left[ \left| \sum_{j=n+1}^{\infty} a_j Z_{t-j} + \sum_{j=-\infty}^{-n-1} a_j Z_{t-j} \right|^2 \right] \leq K \left( \sum_{|j|>n} |a_j| \right)^2$$

and  $\sum_{|j|>n} |a_j|$  converges to zero as  $n \rightarrow \infty$ . ▲

Because almost sure convergence implies convergence in probability, the sequence of random variables  $\{X_t\}$  of Theorem 2.2.1 is defined as a limit in probability, as a limit in mean square, and as an almost sure limit. The second moment condition is required for convergence in mean square, and the first

absolute moment is required for almost sure convergence. Hereafter, we shall often use the infinite sum with no modifiers to represent the limit random variable.

**Theorem 2.2.2.** Let the sequence of real numbers  $\{a_j\}$ ,  $\{b_j\}$  be absolutely summable. Let  $\{Z_t\}$  be a sequence of random variables such that  $E\{Z_t^2\} \leq K$  for  $t = 0, \pm 1, \pm 2, \dots$ , and for some finite  $K$ . Define

$$(X_t, Y_t) = \left( \sum_{j=-\infty}^{\infty} a_j Z_{t-j}, \sum_{k=-\infty}^{\infty} b_k Z_{t-k} \right).$$

Then

$$E\{X_t\} = \lim_{n \rightarrow \infty} \sum_{j=-n}^n a_j E\{Z_{t-j}\}$$

and

$$E\{X_t Y_t\} = \lim_{n \rightarrow \infty} \sum_{j=-n}^n \sum_{k=-n}^n a_j b_k E\{Z_{t-j} Z_{t-k}\}.$$

**Proof.** The expectation result follows from Lemma 2.2.1 and the proof of Theorem 2.2.1. By assumption,

$$\sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} |a_j b_k| E\{|Z_{t-j} Z_{t-k}|^2\} \leq \left( \sum_{j=-\infty}^{\infty} |a_j| \right) \left( \sum_{k=-\infty}^{\infty} |b_k| \right) K < \infty$$

and the conclusion follows by Lemma 2.2.1.  $\blacktriangle$

**Corollary 2.2.2.1.** Let the sequences of real numbers  $\{a_j\}$  and  $\{b_j\}$  be absolutely summable. Let  $\{e_t\}$  be a sequence of uncorrelated  $(0, \sigma^2)$  random variables, and define for  $t = 0, \pm 1, \pm 2, \dots$

$$(X_t, Y_t) = \left( \sum_{j=-\infty}^{\infty} a_j e_{t-j}, \sum_{k=-\infty}^{\infty} b_k e_{t-k} \right).$$

Then,  $E\{X_t\} = 0$  and

$$E\{X_t Y_t\} = \sigma^2 \sum_{j=-\infty}^{\infty} a_j b_j.$$

**Proof.** The sequence  $\{e_t\}$  satisfies the hypotheses of Theorem 2.2.2 with  $K = \sigma^2$ . Since  $E\{e_t\} = 0$  and  $E\{e_t e_j\} = 0$  for  $t \neq j$ , we have the results.  $\blacktriangle$

**Corollary 2.2.2.2.** Let  $\{X_t\}$  be given by

$$X_t = \sum_{j=-\infty}^{\infty} a_j e_{t-j},$$

and define

$$Y_t = \sum_{j=-\infty}^{\infty} b_j X_{t-j},$$

where  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables and  $\{a_j\}$  and  $\{b_j\}$  are absolutely summable sequences of real numbers. Then  $\{Y_t\}$  is a stationary moving average time series and

$$\begin{aligned} \gamma_Y(h) &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} b_j b_k \gamma_X(k-j+h) \\ &= \sum_{j=-\infty}^{\infty} c_j c_{j-h} \sigma^2, \end{aligned} \quad (2.2.2)$$

where  $c_j = \sum_{k=-\infty}^{\infty} b_k a_{j-k}$ . Furthermore,

$$\sum_{h=0}^{\infty} |\gamma_Y(h)| < \infty.$$

**Proof.** By definition,

$$Y_t = \sum_{j=-\infty}^{\infty} b_j X_{t-j} = \sum_{j=-\infty}^{\infty} b_j \sum_{k=-\infty}^{\infty} a_k e_{t-j-k}.$$

Since  $\{a_j\}$  and  $\{b_j\}$  are absolutely summable, we may interchange the order of summation (Fubini's theorem). Setting  $m = j + k$ , we have

$$\begin{aligned} Y_t &= \sum_{m=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_j a_{m-j} e_{t-m} \\ &= \sum_{m=-\infty}^{\infty} c_m e_{t-m}, \end{aligned}$$

where  $c_m = \sum_{j=-\infty}^{\infty} b_j a_{m-j}$ . By result 3 on the convolution of absolutely summable sequences,  $\{c_m\}$  is absolutely summable. Therefore,

$$\gamma_Y(h) = \sum_{m=-\infty}^{\infty} c_m c_{m-h} \sigma^2$$

is also absolutely summable.

Using Theorem 2.2.2,

$$\begin{aligned} \gamma_Y(h) &= \sum_{m=-\infty}^{\infty} c_m c_{m-h} \sigma^2 \\ &= \sum_{m=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} b_j a_{m-j} \sum_{k=-\infty}^{\infty} b_k a_{m-k-h} \sigma^2 \end{aligned}$$

$$\begin{aligned}
 &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} b_j b_k \sum_{m=-\infty}^{\infty} a_{m-j} a_{m-k-h} \sigma^2 \\
 &= \sum_{j=-\infty}^{\infty} \sum_{k=-\infty}^{\infty} b_j b_k \gamma_X(k-j+h). \quad \blacktriangle
 \end{aligned}$$

We may paraphrase this corollary by saying that an infinite moving average (where the coefficients are absolutely summable) of an infinite moving average stationary time series is itself an infinite moving average stationary time series. If the  $\{a_j\}$  and  $\{b_j\}$  are exponentially declining, then  $\{c_m\}$  of Corollary 2.2.2.2 is exponentially declining.

**Corollary 2.2.2.3.** Let the  $\{a_j\}$  and  $\{b_j\}$  of Corollary 2.2.2.2 satisfy

$$|a_j| < M\lambda^{|j|} \quad \text{and} \quad |b_j| < M\lambda^{|j|}$$

for some  $M < \infty$  and  $0 < \lambda < 1$ . Let  $Y_t$  be as defined in Corollary 2.2.2.2. Then

$$Y_t = \sum_{m=-\infty}^{\infty} c_m e_{t-m},$$

where  $c_m = \sum_{j=-\infty}^{\infty} b_j a_{m-j}$  and  $|c_m| < M_c \lambda^{|m|}$  for some  $M_c < \infty$ .

**Proof.** We have

$$|c_m| = \left| \sum_{j=-\infty}^{\infty} b_j a_{m-j} \right| \leq M^2 \lambda^{|m|} \left( m + 1 + 2 \sum_{j=1}^{\infty} \lambda^{2j} \right). \quad \blacktriangle$$

In Theorems 2.2.1 and 2.2.2, we made mild assumptions on the  $Z_t$  and assumed the sequence of coefficients,  $\{a_j\}$ , to be absolutely summable. In Theorem 2.2.3, we give a mean square result for the infinite sum of the elements of a time series under weaker assumptions on the coefficients.

**Theorem 2.2.3.** Let  $\{a_j\}$  be a sequence of real numbers such that  $\{a_j^2\}$  is summable. Let  $Z_t$  be the time series

$$Z_t = \sum_{j=-\infty}^{\infty} b_j e_{t-j},$$

where  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables and  $\{b_j\}$  is absolutely summable. Then there exists a sequence of random variables  $\{X_t\}$  such that for  $t = 0, \pm 1, \pm 2, \dots$ ,

$$\lim_{n \rightarrow \infty} E \left\{ \left( X_t - \sum_{j=-n}^n a_j Z_{t-j} \right)^2 \right\} = 0,$$

$E\{X_t\} = 0$ , and

$$E\{X_t X_{t+h}\} = \sigma^2 \sum_{j=-\infty}^{\infty} c_j c_{j+h}, \quad (2.2.3)$$

where  $c_j = \sum_{k=-\infty}^{\infty} a_k b_{j-k}$ . Furthermore,

$$\lim_{n \rightarrow \infty} E\left\{ \left( X_t - \sum_{j=-n}^n c_j e_{t-j} \right)^2 \right\} = 0.$$

**Proof.** By the Cauchy-Schwarz inequality, we have, for  $n > m$  and  $h > 0$ ,

$$\left| \sum_{j=m+1}^n a_j a_{j+h} \right| \leq \left[ \left( \sum_{j=m+1}^n a_j^2 \right) \left( \sum_{j=m+1}^n a_{j+h}^2 \right) \right]^{1/2}. \quad (2.2.4)$$

Since  $\sum a_j^2 < \infty$ , given  $\epsilon > 0$ , there exists an  $N_0$  such that, for all  $n > m > N_0$ ,

$$\left| \sum_{j=m+1}^n a_j a_{j+h} \right| < \epsilon \quad (2.2.5)$$

uniformly in  $h$ . Let  $Y_{nt} = \sum_{j=-n}^n a_j Z_{t-j}$ . For  $n > m > N_0$ ,

$$E\{(Y_{nt} - Y_{mt})^2\} \leq 4E\left\{ \left( \sum_{j=-n}^{-m-1} a_j Z_{t-j} \right)^2 \right\} + 4E\left\{ \left( \sum_{j=m+1}^n a_j Z_{t-j} \right)^2 \right\}. \quad (2.2.6)$$

Now

$$E\left\{ \left( \sum_{j=m+1}^n a_j Z_{t-j} \right)^2 \right\} = \left( \sum_{j=m+1}^n a_j^2 \right) \gamma_Z(0) + 2 \sum_{j=m+1}^{n-1} \sum_{k=j+1}^n a_j a_k \gamma_Z(k-j),$$

where

$$\left| \sum_{j=m+1}^{n-1} \sum_{k=j+1}^n a_j a_k \gamma_Z(k-j) \right| \leq \epsilon \sum_{h=1}^{\infty} |\gamma_Z(h)|$$

by (2.2.5). Because, by Corollary 2.2.2.2,  $\sum_{h=1}^{\infty} |\gamma_Z(h)| < \infty$ , it follows that the right side of (2.2.6) converges to zero in mean square as  $m \rightarrow \infty$  and  $n \rightarrow \infty$ . Therefore,  $Y_{nt}$  is Cauchy in squared mean and there exists a random variable  $X_t$  such that  $E\{X_t^2\} < \infty$  and

$$E\{(X_t - Y_{nt})^2\} \rightarrow 0 \quad \text{as } n \rightarrow \infty.$$

See, for example, Tucker (1967, p. 105). Also,  $E\{X_t\} = \lim_{n \rightarrow \infty} E\{Y_{nt}\} = 0$  and  $E\{X_t X_{t+h}\} = \lim_{n \rightarrow \infty} E\{Y_{nt} Y_{nt+h}\}$ .

We write

$$Y_{nt} = \sum_{j=-\infty}^{\infty} d_{jn} Z_{t-j} = \sum_{j=-\infty}^{\infty} d_{jn} \sum_{l=-\infty}^{\infty} b_l e_{t-j-l},$$

where

$$d_{jn} = \begin{cases} a_j & \text{if } |j| \leq n, \\ 0 & \text{if } |j| > n. \end{cases}$$

Since  $d_{jn}$  and  $b_l$  are absolutely summable, we have

$$Y_{nt} = \sum_{l=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} d_{jn} b_l e_{t-j-l} = \sum_{k=-\infty}^{\infty} c_{kn} e_{t-k},$$

where

$$c_{kn} = \sum_{j=-\infty}^{\infty} d_{jn} b_{k-j} = \sum_{j=-n}^{\infty} a_j b_{k-j}.$$

Also,

$$E\{Y_{nt} Y_{n,t+h}\} = \sum_{k=-\infty}^{\infty} c_{kn} c_{k+h,n} \sigma^2.$$

Therefore,

$$E\{X_t X_{t+h}\} = \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} c_{kn} c_{k+h,n} \sigma^2.$$

Now  $c_{kn}$  converges to  $c_k = \sum_{j=-\infty}^{\infty} a_j b_{k-j}$  for all  $k$ . Also,  $|c_{kn}| \leq m_k$  for all  $k$ , where  $m_k = \sum_{j=-\infty}^{\infty} |a_j| |b_{k-j}|$ , and

$$\sum_{k=-\infty}^{\infty} |m_k| |m_{k+h}| < \infty,$$

because  $\{m_k\}$  is square summable. See Exercise 2.45. Therefore, by the dominated convergence theorem,

$$E\{X_t X_{t+h}\} = \lim_{n \rightarrow \infty} \sum_{k=-\infty}^{\infty} c_{kn} c_{k+h,n} \sigma^2 = \sum_{k=-\infty}^{\infty} c_k c_{k+h} \sigma^2. \quad \blacktriangle$$

**Corollary 2.2.3.** Let  $\{c_j\}_{-\infty}^{\infty}$  be square summable and let  $\{d_j\}_{-\infty}^{\infty}$  be absolutely summable. Let

$$Z_t = \sum_{j=-\infty}^{\infty} c_j e_{t-j},$$

where  $e_i$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables. Then there exists a sequence of random variables  $\{X_i\}_{-\infty}^{\infty}$  such that

$$\lim_{n \rightarrow \infty} E \left\{ \left| X_t - \sum_{j=-n}^n d_j Z_{t-j} \right|^2 \right\} = 0,$$

$E\{X_i\} = 0$ , and

$$E\{X_t X_{t+h}\} = \sigma^2 \sum_{j=-\infty}^{\infty} g_j g_{j+h},$$

where  $g_j = \sum_{k=-\infty}^{\infty} c_k d_{j-k}$ . Furthermore,

$$\lim_{n \rightarrow \infty} E \left\{ \left| X_t - \sum_{j=-n}^n g_j e_{t-j} \right|^2 \right\} = 0.$$

**Proof.** From Theorem 2.2.3, we have that  $Z_t$  is well defined in mean square with  $E\{Z_t\} = 0$  and  $\gamma_Z(h) = \sigma^2 \sum_{j=-\infty}^{\infty} c_j c_{j+h}$ . Now, the existence of the sequence of random variables  $\{X_i\}_{-\infty}^{\infty}$  such that  $E\{X_i\} = 0$  and

$$\lim_{n \rightarrow \infty} E \left\{ \left| X_t - \sum_{j=-n}^n d_j Z_{t-j} \right|^2 \right\} = 0$$

follows from Theorem 2.2.1. Furthermore, from Theorem 2.2.2, we have

$$\begin{aligned} E\{X_t X_{t+h}\} &= \lim_{n \rightarrow \infty} \sum_{j=-n}^n \sum_{k=-n}^n d_j d_k E\{Z_{t-j} Z_{t+h-k}\} \\ &= \lim_{n \rightarrow \infty} \sum_{j=-n}^n \sum_{k=-n}^n d_j d_k \sum_{l=-\infty}^{\infty} c_l c_{l+j+h-k} \sigma^2 \\ &= \sigma^2 \sum_{j=-\infty}^{\infty} g_j g_{j+h}, \end{aligned}$$

where  $g_j = \sum_{k=-\infty}^{\infty} c_k d_{j-k}$ . Using the dominated convergence theorem, we have

$$\begin{aligned} \lim_{n \rightarrow \infty} E \left\{ X_t \sum_{i=-n}^n g_i e_{t-i} \right\} &= \lim_{n \rightarrow \infty} \sum_{i=-n}^n g_i E \left\{ \sum_{j=-\infty}^{\infty} d_j Z_{t-j} e_{t-i} \right\} \\ &= \lim_{n \rightarrow \infty} \sum_{i=-n}^n g_i \sum_{j=-\infty}^{\infty} d_j c_{i-j} \sigma^2 \\ &= \sigma^2 \sum_{i=-\infty}^{\infty} g_i^2 = E\{X_t^2\}. \end{aligned}$$

Therefore,

$$\lim_{n \rightarrow \infty} E\left\{\left(X_t - \sum_{i=-n}^n g_i e_{t-i}\right)^2\right\} = E\{X_t^2\} + \lim_{n \rightarrow \infty} \sum_{i=-n}^n g_i^2 \sigma^2 - 2 \lim_{n \rightarrow \infty} E\left\{X_t \sum_{i=-n}^n g_i e_{t-i}\right\} \\ = 0. \quad \blacktriangle$$

### 2.3. AN INTRODUCTION TO AUTOREGRESSIVE TIME SERIES

Many time series encountered in practice are well approximated by the representation

$$\sum_{i=0}^p \alpha_i X_{t-i} = e_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad (2.3.1)$$

where  $\alpha_0 \neq 0$ ,  $\alpha_p \neq 0$ , and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables. The sequence  $\{X_t\}$  is called a *pth order autoregressive* time series. The defining equation (2.3.1) is sometimes called a *stochastic difference equation*.

To study such processes, let us consider the first order autoregressive time series, which we express in its most common form,

$$X_t = \rho X_{t-1} + e_t. \quad (2.3.2)$$

By repeated substitution of

$$X_{t-i} = \rho X_{t-i-1} + e_{t-i}$$

for  $i = 1, 2, \dots, N$  into (2.3.2), we obtain

$$X_t = \rho^N X_{t-N} + \sum_{i=0}^{N-1} \rho^i e_{t-i}. \quad (2.3.3)$$

Under the assumptions that  $|\rho| < 1$  and  $E\{X_t^2\} < K < \infty$ , we have

$$\lim_{N \rightarrow \infty} E\left\{\left(X_t - \sum_{i=0}^N \rho^i e_{t-i}\right)^2\right\} = 0. \quad (2.3.4)$$

Thus, if  $e_t$  is defined for  $t \in \{0, \pm 1, \pm 2, \dots\}$  and  $X_t$  satisfies the difference equation (2.3.2) with  $|\rho| < 1$ , then we may express  $X_t$  as an infinite moving average of the  $e_t$ :

$$X_t = \sum_{i=0}^{\infty} \rho^i e_{t-i}.$$

It follows that  $E\{X_t\} = 0$  for all  $t$ , and

$$\begin{aligned}\gamma(h) &= E\{X_t X_{t+h}\} = E\left\{\left(\sum_{i=0}^{\infty} \rho^i e_{t-i}\right)\left(\sum_{j=0}^{\infty} \rho^j e_{t+h-j}\right)\right\} \\ &= \sigma^2 \sum_{i=0}^{\infty} \rho^i \rho^{i+h} = \frac{\rho^h}{1-\rho^2} \sigma^2\end{aligned}\quad (2.3.5)$$

for  $h = 0, 1, \dots$ . The covariance function for  $h \neq 0$  is also rapidly obtained by making use of a form of (2.3.3),

$$X_t = \rho^h X_{t-h} + \sum_{j=0}^{h-1} \rho^j e_{t-j}, \quad h = 1, 2, \dots \quad (2.3.6)$$

Since  $X_{t-h}$  is a function of  $e$ 's with subscript less than or equal to  $t-h$ ,  $X_{t-h}$  is uncorrelated with  $e_{t-h+1}, e_{t-h+2}, \dots$ . Therefore, we have, after multiplying both sides of (2.3.6) by  $X_{t-h}$  and taking expectations,

$$\begin{aligned}\gamma(h) &= E\{X_t X_{t-h}\} = E\{\rho^h X_{t-h}^2\} + E\left\{X_{t-h} \sum_{j=0}^{h-1} \rho^j e_{t-j}\right\} \\ &= \rho^h \gamma(0), \quad h = 1, 2, \dots\end{aligned}\quad (2.3.7)$$

The correlation function is seen to be a monotonically declining function for  $\rho > 0$ , while for  $\rho < 0$  the function is alternately positive and negative, the absolute value declining at a geometric rate.

**Example 2.3.1.** Figure 2.3.1 displays 100 observations from the first order process

$$X_t = 0.7X_{t-1} + e_t.$$

The current observation,  $X_t$ , is plotted against the preceding observation,  $X_{t-1}$ , in Figure 2.3.2 and against  $X_{t-2}$  in Figure 2.3.3. The correlation between  $X_t$  and  $X_{t-2}$  is  $(0.7)^2 = 0.49$ , but the partial correlation between  $X_t$  and  $X_{t-2}$  adjusted for  $X_{t-1}$  is zero, since

$$\text{Cov}\{X_t - \rho X_{t-1}, X_{t-2} - \rho X_{t-1}\} = (\rho^2 - \rho^2 - \rho^2 + \rho^2)\gamma(0) = 0.$$

The zero partial correlation is illustrated by Figure 2.3.4, which contains a plot of  $X_t - 0.7X_{t-1}$  against  $X_{t-2} - 0.7X_{t-1}$ . In fact, for any  $h \in (2, 3, \dots)$ , the partial correlation between  $X_t$  and  $X_{t-h}$  adjusted for  $X_{t-1}$  is zero. This follows because

$$\begin{aligned}\text{Cov}\{X_t - \rho X_{t-1}, X_{t-h} - \rho^{h-1} X_{t-1}\} \\ = (\rho^h - \rho^{h-1} \rho - \rho \rho^{h-1} + \rho^h)\gamma(0) = 0, \quad h = 2, 3, \dots\end{aligned}$$

This important property of the first order autoregressive time series can be

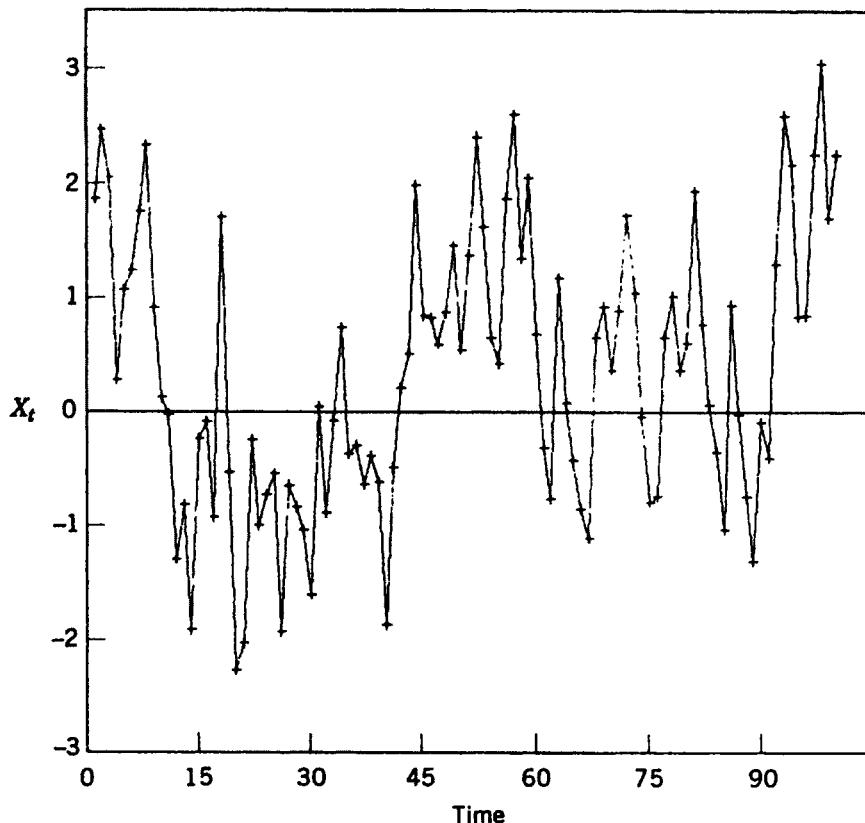


Figure 2.3.1. One hundred observations from the time series  $X_t = 0.7X_{t-1} + e_t$ .

characterized by saying that all the useful (correlational) information about  $X_t$  in the entire realization previous to  $X_t$  is contained in  $X_{t-1}$ . Compare this result with the partial correlational properties of the moving average process discussed in Section 2.1.  $\blacktriangle\blacktriangle$

Linear difference equations and their solutions play an important role in the study of autoregressive time series as well as in other areas of time series analysis. Therefore we digress to present the needed elementary results before studying higher order autoregressive processes. The initial development follows Goldberg (1958).

## 2.4. DIFFERENCE EQUATIONS

Given a sequence  $\{y_t\}_{t=0}^{\infty}$ , the *first difference* is defined by

$$\Delta y_t = y_t - y_{t-1}, \quad t = 1, 2, \dots, \quad (2.4.1)$$

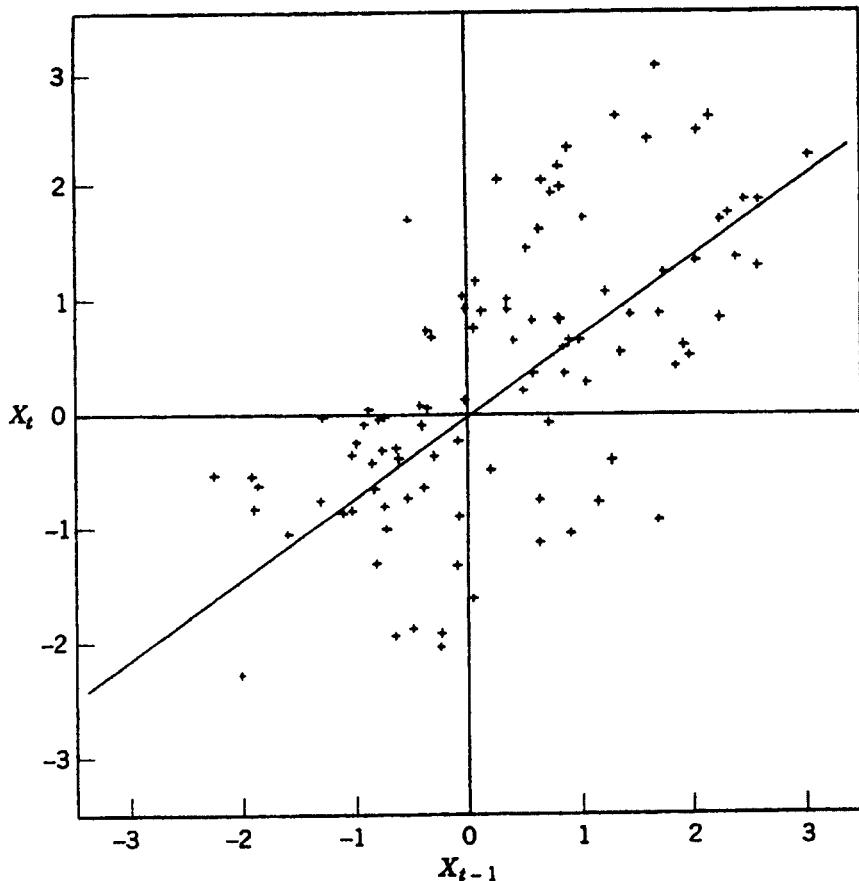


Figure 2.3.2. Plot of  $X_t$  against  $X_{t-1}$ , for 100 observations from  $X_t = 0.7X_{t-1} + \epsilon_t$ .

and the  $n$ th difference is defined by

$$\Delta^n y_t = \Delta^{n-1} y_t - \Delta^{n-1} y_{t-1} = \sum_{r=0}^n (-1)^r \binom{n}{r} y_{t-r}, \quad t = n, n+1, \dots,$$

where

$$\binom{n}{r} = \frac{n!}{r!(n-r)!}$$

are the binomial coefficients. The equation

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_n y_{t-n} = r_t, \quad t = n, n+1, \dots, \quad (2.4.2)$$

where the  $a_i$  are real constants,  $a_n \neq 0$ , and  $r_t$  is a real function of  $t$ , is a *linear difference equation of order  $n$*  with constant coefficients. The values  $y_{t-1}, y_{t-2}, \dots, y_{t-n}$  are sometimes called *lagged values* of  $y_t$ . Equation (2.4.2)

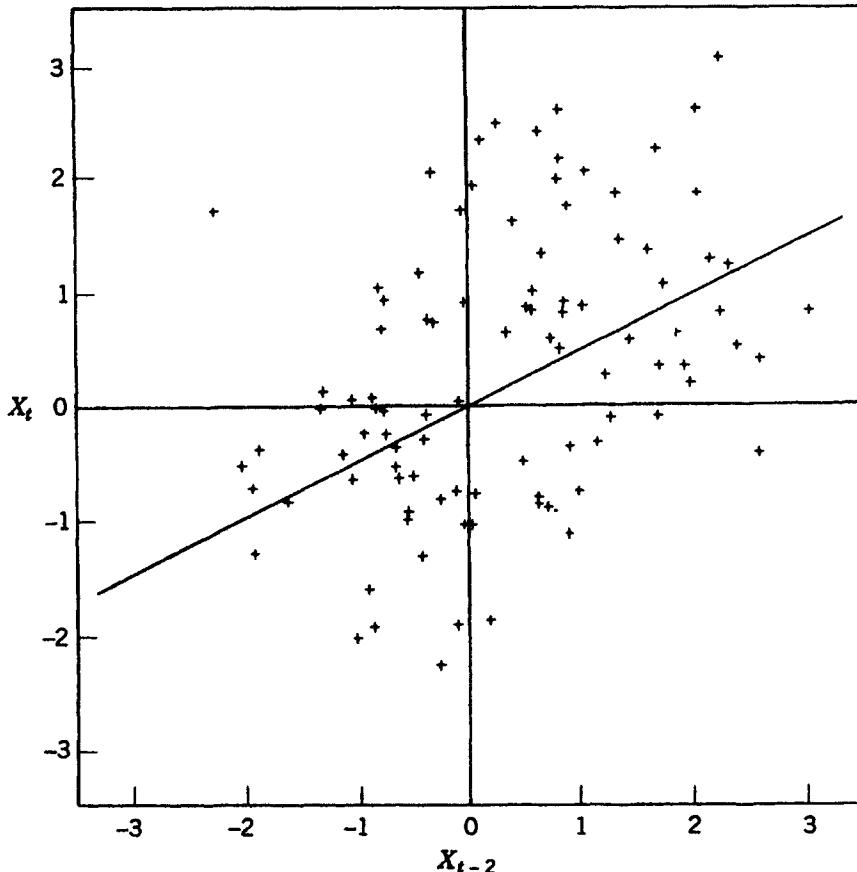


Figure 2.3.3. Plot of  $X_t$  against  $X_{t-2}$  for 100 observations from  $X_t = 0.7X_{t-1} + e_t$ .

could be expressed in terms of  $y_{t-n}$  and the differences of  $y_t$ ; for example,

$$\Delta^n y_t + b_1 \Delta^{n-1} y_{t-1} + b_2 \Delta^{n-2} y_{t-2} + \cdots + b_{n-1} \Delta y_{t-n+1} + b_n y_{t-n} = r_t, \quad (2.4.3)$$

where the  $b_i$ 's are linear functions of the  $a_i$ 's.

A third representation<sup>1</sup> of (2.4.2) is possible. Let the symbol  $\mathcal{B}$  denote the operation of replacing  $y_t$  by  $y_{t-1}$ ; that is,

$$\mathcal{B}y_t = y_{t-1}.$$

<sup>1</sup>The reader may wonder at the benefits associated with several alternative notations. All are heavily used in the literature and all have advantages for certain operations.

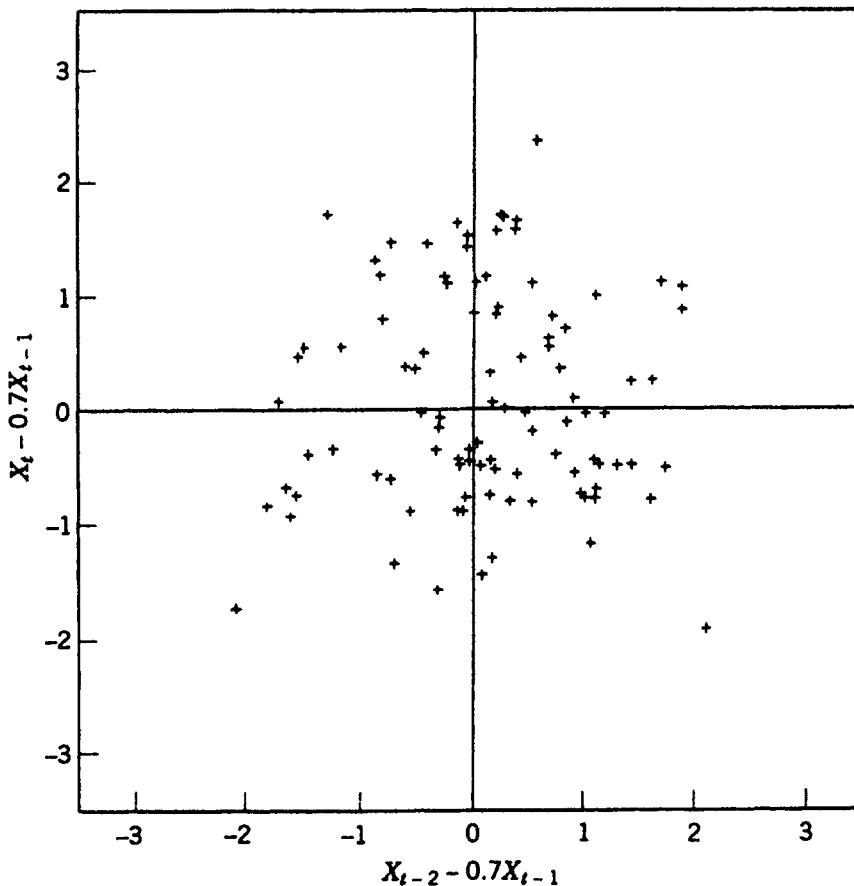


Figure 2.3.4. Plot of  $X_t - 0.7X_{t-1}$  against  $X_{t-2} - 0.7X_{t-1}$  for 100 observations from  $X_t = 0.7X_{t-1} + \epsilon_t$ .

Using the *backward shift operator*  $\mathcal{B}$ , we have

$$\Delta y_t = y_t - \mathcal{B}y_t = (1 - \mathcal{B})y_t.$$

As with the difference operator, we denote repeated application of the operator with the appropriate exponent; for example,

$$\mathcal{B}^3 y_t = \mathcal{B}\mathcal{B}\mathcal{B}y_t = y_{t-3}.$$

Therefore, we can use the operator symbol to write (2.4.2) as

$$(1 + a_1 \mathcal{B} + a_2 \mathcal{B}^2 + \cdots + a_n \mathcal{B}^n) y_t = r_t, \quad t = n, n+1, \dots$$

Associated with (2.4.2) is the *reduced* or *homogeneous* difference equation

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_n y_{t-n} = 0. \quad (2.4.4)$$

From the linearity of (2.4.2) and (2.4.4) we have the following properties.

1. If  $y^{(1)}$  and  $y^{(2)}$  are solutions of (2.4.4), then  $b_1 y^{(1)} + b_2 y^{(2)}$  is a solution of (2.4.4), where  $b_1$  and  $b_2$  are arbitrary constants.
2. If  $Y$  is a solution of (2.4.4) and  $y^t$  a solution of (2.4.2), then  $Y + y^t$  is a solution of (2.4.2).

$y^t$  is called a *particular solution* and  $Y + y^t$  a *general solution* of the difference equation (2.4.2).

Equations (2.4.2) and (2.4.4) specify  $y_t$  as a function of the preceding  $n$  values of  $y$ . By using an inductive proof it is possible to demonstrate that the linear difference equation of order  $n$  has one and only one solution for which values at  $n$  consecutive  $t$  values are arbitrarily prescribed. This important result means that when we obtain  $n$  linearly independent solutions to the difference equation, we can construct the general solution.

The first order homogeneous difference equation

$$y_t = ay_{t-1}, \quad t = 1, 2, \dots, \quad (2.4.5)$$

has the unique solution

$$y_t = a^t y_0, \quad (2.4.6)$$

where  $y_0$  is the value of  $y$  at  $t = 0$ . Note that if  $|a| < 1$ , the solution tends to zero as  $t$  increases. Conversely, if  $|a| > 1$ , the absolute value of the solution sequence increases without bound. If  $a$  is negative, the solution oscillates with alternately positive and negative values. The first order nonhomogeneous difference equation

$$y_t = ay_{t-1} + b \quad (2.4.7)$$

has the unique solution

$$y_t = \begin{cases} y_0 a^t + b \left( \frac{1 - a^t}{1 - a} \right), & a \neq 1, \\ y_0 + bt, & a = 1. \end{cases} \quad (2.4.8)$$

If the constant  $b$  in the difference equation (2.4.7) is replaced by a function of time, for example,

$$y_t = ay_{t-1} + b_t, \quad t = 1, 2, \dots, \quad (2.4.9)$$

then the solution is given by

$$y_t = y_0 a^t + \sum_{j=0}^{t-1} a^j b_{t-j}, \quad t = 1, 2, \dots, \quad (2.4.10)$$

where  $y_0$  is the value of  $y$  at time 0.

The solutions (2.4.6), (2.4.8), and (2.4.10) are readily verified by differencing. The fact that the linear function is reduced by differencing to the constant function is of particular interest. The generalization of this result is important in the analysis of nonstationary time series.

**Theorem 2.4.1.** Let  $y_t$  be a polynomial of degree  $n$  whose domain is the integers. Then the first difference  $\Delta y_t$  is expressible as a polynomial of degree  $n - 1$  in  $t$ .

**Proof.** Since

$$y_t = \sum_{p=0}^n \beta_p t^p, \quad \beta_n \neq 0,$$

and

$$y_{t-1} = \sum_{p=0}^n \beta_p (t-1)^p,$$

we have

$$\begin{aligned} \Delta y_t &= \sum_{p=0}^n \beta_p [t^p - (t-1)^p] \\ &= \sum_{p=0}^n \beta_p \left[ t^p - \sum_{s=0}^p \binom{p}{s} (-1)^s t^{p-s} \right] \\ &= \sum_{p=1}^n \beta_p \left[ - \sum_{s=1}^p \binom{p}{s} (-1)^s t^{p-s} \right] \\ &\stackrel{\text{(say)}}{=} \sum_{p=0}^{n-1} \gamma_p t^p, \end{aligned}$$

where  $\gamma_{n-1} = n\beta_n$ . ▲

It follows that by taking repeated differences, one can reduce a polynomial to the zero function. We state the result as a corollary.

**Corollary 2.4.1.** Let  $y_t$  be a polynomial of degree  $n$  whose domain is the integers. Then the  $n$ th difference  $\Delta^n y_t$  is a constant function, and the  $(n+1)$ st difference  $\Delta^{n+1} y_t$  is the zero function.

In investigating difference equations of higher order, the nature of the solution is a function of the *auxiliary equation*, sometimes called the *characteristic equation*. For equation (2.4.2) the auxiliary equation is

$$m^n + a_1 m^{n-1} + \cdots + a_n = 0. \quad (2.4.11)$$

This polynomial equation will have  $n$  (not necessarily distinct) roots, and the behavior of the solution is intimately related to these roots.

Consider the second order linear homogeneous equation

$$y_t + a_1 y_{t-1} + a_2 y_{t-2} = 0, \quad (2.4.12)$$

where  $a_2 \neq 0$ . The two roots of the auxiliary equation

$$m^2 + a_1 m + a_2 = 0 \quad (2.4.13)$$

will fall into one of three categories:

1. Real and distinct.
2. Real and equal.
3. A complex conjugate pair.

The general solution of the homogeneous difference equation (2.4.12) for the three categories is:

1.

$$y_t = b_1 m'_1 + b_2 m'_2, \quad (2.4.14)$$

where  $b_1$  and  $b_2$  are constants determined by the initial conditions, and  $m'_1$  and  $m'_2$  are the roots of (2.4.13).

2.

$$y_t = (b_1 + b_2 t) m', \quad (2.4.15)$$

where  $m'$  is the repeated root of (2.4.13).

3.

$$y_t = b_1^* m'_1 + b_2 m'_2, \quad (2.4.16)$$

where  $b_1^*$  is the complex conjugate of  $b_1$  and  $m_2 = m_1^*$  is the complex conjugate of  $m_1$ .

If the complex roots are expressed as

$$r(\cos \theta \pm i \sin \theta),$$

where  $r = a_2^{1/2} = |m_1|$  and

$$(\cos \theta, \sin \theta) = (2r)^{-1}(a_1, |a_1^2 - 4a_2|^{1/2}),$$

then the solution can be expressed as

$$\begin{aligned} y_t &= r'(d_1 \cos t\theta + d_2 \sin t\theta) \\ &= g_1 r' \cos(t\theta + g_2), \end{aligned} \quad (2.4.17)$$

where  $d_1$ ,  $d_2$ ,  $g_1$ , and  $g_2$  are real constants.

Substitution of (2.4.14) into (2.4.12) immediately establishes that (2.4.14) is a solution and, in fact, that  $m'_1$  and  $m'_2$  are independent solutions when  $m_1 \neq m_2$ . For a repeated root, substitution of (2.4.15) into (2.4.12) gives

$$(b_1 + b_2 t)m'^{-2}(m^2 + a_1 m + a_2) - b_2 m'^{-2}(a_1 m + 2a_2) = 0.$$

The first term is zero because  $m$  is a root of (2.4.13). For  $m$  a repeated root,  $ma_1 + 2a_2 = 0$ , proving that (2.4.15) is the general solution.

As with unequal real roots, substitution of (2.4.16) into (2.4.12) demonstrates that (2.4.16) is the solution in the case of complex roots.

The solution for a linear homogeneous difference equation of order  $n$  can be obtained using the roots of the auxiliary equation. The solution is a sum of  $n$  terms where:

1. For every real and distinct root  $m$ , a term of the form  $bm'$  is included.
2. For every real root of order  $p$  (a root repeated  $p$  times), a term of the form

$$(b_1 + b_2 t + b_3 t^2 + \cdots + b_p t^{p-1})m'$$

is included.

3. For each pair of unrepeated complex conjugate roots, a term of the form

$$\alpha r' \cos(t\theta + \beta)$$

is included, where  $r$  and  $\theta$  are defined following (2.4.16).

4. For a pair of complex conjugate roots occurring  $p$  times, a term of the form

$$r'[\alpha_1 \cos(t\theta + \beta_1) + \alpha_2 t \cos(t\theta + \beta_2) + \cdots + \alpha_p t^{p-1} \cos(t\theta + \beta_p)]$$

is included.

For the nonhomogeneous difference equation (2.4.2), with initial conditions  $y_0, y_1, \dots, y_{n-1}$ , a particular solution is

$$y_t^\dagger = \sum_{j=0}^t w_j r_{t-j}, \quad t = n, n+1, \dots,$$

where  $w_0 = 1$ ,

$$\begin{aligned} \sum_{i=0}^j a_i w_{j-i} &= 0, \quad j = 1, 2, \dots, n-1, \\ \sum_{i=0}^n a_i w_{j-i} &= 0, \quad j = n, n+1, \dots, \end{aligned}$$

and  $r_0, r_1, \dots, r_{n-1}$  are defined by

$$y_t = \sum_{j=0}^t w_j r_{t-j}, \quad t = 0, 1, \dots, n-1.$$

Observe that the weights  $w_j$  satisfy the same homogeneous difference equation as the  $y_i$ . Therefore,  $w_j$  can be expressed as a sum of the four types of terms that define the solution to the homogeneous difference equation.

Systems of difference equations in more than one function of time arise naturally in many models of the physical and economic sciences. To extend our discussion to such models, consider the system of difference equations in the two functions  $y_{1t}$  and  $y_{2t}$ :

$$\begin{aligned} y_{1t} &= a_{11}y_{1,t-1} + a_{12}y_{2,t-1} + b_{1t}, \\ y_{2t} &= a_{21}y_{1,t-1} + a_{22}y_{2,t-1} + b_{2t}, \end{aligned} \quad (2.4.18)$$

where  $a_{11}, a_{12}, a_{21}, a_{22}$  are real constants,  $b_{1t}$  and  $b_{2t}$  are real functions of time, and  $t = 1, 2, \dots$ . The system may be expressed in matrix notation as

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} + \mathbf{b}_t, \quad (2.4.19)$$

where

$$\mathbf{y}'_t = (y_{1t}, y_{2t}),$$

$$\mathbf{A} = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix},$$

and  $\mathbf{b}'_t = (b_{1t}, b_{2t})$ .

Proceeding from the solution (2.4.6) of the first order homogeneous scalar difference equation, we might postulate that the homogeneous equation

$$\mathbf{y}_t = \mathbf{A}\mathbf{y}_{t-1} \quad (2.4.20)$$

would have the solution

$$\mathbf{y}_t = \mathbf{A}'\mathbf{y}_0, \quad (2.4.21)$$

and that the equation (2.4.19) would have the solution

$$\mathbf{y}_t = \mathbf{A}'\mathbf{y}_0 + \sum_{j=0}^{t-1} \mathbf{A}'\mathbf{b}_{t-j}, \quad (2.4.22)$$

where we understand that  $\mathbf{A}^0 = \mathbf{I}$ . Direct substitution of (2.4.21) into (2.4.20) and (2.4.22) into (2.4.19) verifies that these are the solutions of the matrix analogs of the first order scalar equations.

In the scalar case the behavior of the solution depends in a critical manner on the magnitude of the coefficient  $a$  for the first order equation and on the magnitude of the roots of the characteristic equation for the higher order equations. If the

roots are all less than one in absolute value, then the effect of the initial conditions is transient, the initial conditions being multiplied by the product of powers of the roots and polynomials in  $t$ .

In the vector case the quantities analogous to the roots of the characteristic equation are the roots of the determinantal equation

$$|\mathbf{A} - m\mathbf{I}| = 0.$$

For the moment, we assume that the roots of the determinantal equation are distinct. Let the two roots of the matrix  $\mathbf{A}$  of (2.4.19) be  $m_1$  and  $m_2$ , and define the vectors  $\mathbf{q}_{.1}$  and  $\mathbf{q}_{.2}$  by the equations

$$\begin{aligned} (\mathbf{A} - m_1 \mathbf{I}) \mathbf{q}_{.1} &= 0, \\ (\mathbf{A} - m_2 \mathbf{I}) \mathbf{q}_{.2} &= 0, \\ \mathbf{q}'_{.1} \mathbf{q}_{.1} &= \mathbf{q}'_{.2} \mathbf{q}_{.2} = 1. \end{aligned}$$

By construction, the matrix  $\mathbf{Q} = (\mathbf{q}_{.1}, \mathbf{q}_{.2})$  is such that

$$\mathbf{AQ} = \mathbf{QM}, \quad (2.4.23)$$

where  $\mathbf{M} = \text{diag}(m_1, m_2)$ . Since the roots are distinct, the matrix  $\mathbf{Q}$  is nonsingular. Hence,

$$\mathbf{Q}^{-1} \mathbf{AQ} = \mathbf{M}. \quad (2.4.24)$$

Define  $\mathbf{z}'_t = (z_{1t}, z_{2t})$  and  $\mathbf{c}'_t = (c_{1t}, c_{2t})$  by

$$\begin{aligned} \mathbf{z}'_t &= \mathbf{Q}^{-1} \mathbf{y}_t, \\ \mathbf{c}'_t &= \mathbf{Q}^{-1} \mathbf{b}_t. \end{aligned}$$

Then, multiplying (2.4.19) by  $\mathbf{Q}^{-1}$ , we have

$$\mathbf{z}'_t = \mathbf{Q}^{-1} \mathbf{AQ} \mathbf{z}'_{t-1} + \mathbf{c}'_t$$

or

$$\begin{pmatrix} z_{1t} \\ z_{2t} \end{pmatrix} = \begin{pmatrix} m_1 & 0 \\ 0 & m_2 \end{pmatrix} \begin{pmatrix} z_{1,t-1} \\ z_{2,t-1} \end{pmatrix} + \begin{pmatrix} c_{1t} \\ c_{2t} \end{pmatrix}.$$

The transformation  $\mathbf{Q}^{-1}$  reduces the system of difference equations to two simple first order difference equations with solutions

$$\begin{aligned} z_{1t} &= \sum_{j=0}^t m_1^j c_{1,t-j}, \\ z_{2t} &= \sum_{j=0}^t m_2^j c_{2,t-j}, \end{aligned} \quad (2.4.25)$$

where we have set  $y_{10} = b_{10}$  and  $y_{20} = b_{20}$  for notational convenience. Therefore, the effect of the initial conditions is transient if  $|m_1| < 1$  and  $|m_2| < 1$ . The solutions (2.4.25) may be expressed in terms of the original variables as

$$\mathbf{Q}^{-1}\mathbf{y}_t = \sum_{j=0}^t \mathbf{M}^j \mathbf{Q}^{-1} \mathbf{b}_{t-j}$$

or

$$\begin{aligned} \mathbf{y}_t &= \sum_{j=0}^t \mathbf{Q} \mathbf{M}^j \mathbf{Q}^{-1} \mathbf{b}_{t-j} \\ &= \sum_{j=0}^t \mathbf{A}^j \mathbf{b}_{t-j}, \end{aligned} \quad (2.4.26)$$

where the last representation follows from  $\mathbf{A}^2 = \mathbf{QM} \mathbf{Q}^{-1} \mathbf{QM} \mathbf{Q}^{-1} = \mathbf{QM}^2 \mathbf{Q}^{-1}$ , etc.

If the  $n \times n$  matrix  $\mathbf{A}$  has multiple roots, it cannot always be reduced to a diagonal by a transformation matrix  $\mathbf{Q}$ . However,  $\mathbf{A}$  can be reduced to a matrix whose form is slightly more complicated.

**Theorem 2.4.2.** Let  $\Lambda_i$  be the  $k_i \times k_i$  matrix

$$\Lambda_i = \begin{bmatrix} m_i & 1 & 0 & \cdots & 0 & 0 \\ 0 & m_i & 1 & \cdots & 0 & 0 \\ 0 & 0 & m_i & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 0 & m_i \end{bmatrix},$$

where  $\Lambda_i = m_i$  when  $k_i = 1$ . Then there exists a matrix  $\mathbf{Q}$  such that

$$\mathbf{Q}^{-1} \mathbf{A} \mathbf{Q} = \begin{bmatrix} \Lambda_1 & 0 & \cdots & 0 \\ 0 & \Lambda_2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \Lambda_r \end{bmatrix} = \Lambda, \quad (2.4.27)$$

where  $\sum_{i=1}^r k_i = n$  and the  $m_i$ ,  $i = 1, 2, \dots, r$ , are the characteristic roots of  $\mathbf{A}$ . If  $m_i$  is a repeated root, it may appear in more than one block of  $\Lambda$ , but the total number of times it appears is equal to its multiplicity.

**Proof.** See Finkbeiner (1960) or Miller (1963). ▲

The representation  $\Lambda$  is called the *Jordan canonical form*. The powers of the matrix  $\Lambda$ , are given by

$$\Lambda_i^j = \begin{pmatrix} m_i^j & \binom{j}{1}m_i^{j-1} & \binom{j}{2}m_i^{j-2} & \cdots & \binom{j}{k_i-1}m_i^{j-k_i+1} \\ 0 & m_i^j & \binom{j}{1}m_i^{j-1} & \cdots & \binom{j}{k_i-2}m_i^{j-k_i+2} \\ 0 & 0 & m_i^j & \cdots & \binom{j}{k_i-3}m_i^{j-k_i+3} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & m_i^j \end{pmatrix}. \quad (2.4.28)$$

Therefore, the solution of the first order vector difference equation with initial conditions  $\mathbf{y}_0 = \mathbf{b}_0$  is given by

$$\mathbf{y}_t = \sum_{j=0}^t \mathbf{Q}\Lambda^j \mathbf{Q}^{-1} \mathbf{b}_{t-j}, \quad (2.4.29)$$

and the effect of the initial conditions goes to zero as  $t \rightarrow \infty$  if all of the roots of  $|\mathbf{A} - m\mathbf{I}| = 0$  are less than one in absolute value.

To further illustrate the solution, consider a system of dimension two with repeated root that has been reduced to the Jordan canonical form:

$$\begin{aligned} y_{1t} &= my_{1,t-1} + y_{2,t-1}, \\ y_{2t} &= my_{2,t-1}. \end{aligned}$$

It follows that

$$y_{2t} = y_{20}m^t$$

and  $y_{1t}$  is given as the solution of the nonhomogeneous equation

$$y_{1t} = my_{1,t-1} + y_{20}m^{t-1},$$

whence

$$y_{1t} = y_{10}m^t + y_{20}tm^{t-1}.$$

This approach may be extended to treat higher order vector difference equations. For example, consider the second order vector difference equation

$$\mathbf{y}_t = \mathbf{A}_1 \mathbf{y}_{t-1} + \mathbf{A}_2 \mathbf{y}_{t-2}, \quad (2.4.30)$$

where  $\mathbf{A}_1$  and  $\mathbf{A}_2$  are  $k \times k$  matrices and  $\mathbf{y}'_t = (y_{1t}, y_{2t}, \dots, y_{kt})$ . The system of equations may also be written as

$$\mathbf{x}_t = \mathbf{Ax}_{t-1}, \quad (2.4.31)$$

where  $\mathbf{x}'_t = (\mathbf{y}'_t, \mathbf{y}'_{t-1})$  and

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 \\ \mathbf{I} & \mathbf{0} \end{pmatrix}. \quad (2.4.32)$$

Thus we have converted the original second order vector difference equation of dimension  $k$  to a first order difference equation of dimension  $2k$ . The properties of the solution will depend on the nature of the Jordan canonical form of  $\mathbf{A}$ . In particular the effect of the initial conditions will be transient if all of the roots of  $|\mathbf{A} - m\mathbf{I}| = 0$  are less than one in absolute value. This condition is equivalent to the condition that the roots of

$$|\mathbf{Im}^2 - \mathbf{A}_1m - \mathbf{A}_2| = 0$$

are less than one in absolute value. To prove this equivalence, form the matrix product

$$\begin{aligned} & \begin{pmatrix} \mathbf{I} & \mathbf{A}_2m^{-1} \\ \mathbf{0} & \mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{A}_1 - m\mathbf{I} & \mathbf{A}_2 \\ \mathbf{I} & -m\mathbf{I} \end{pmatrix} \begin{pmatrix} \mathbf{I} & \mathbf{0} \\ m^{-1}\mathbf{I} & \mathbf{I} \end{pmatrix} \\ &= \begin{pmatrix} \mathbf{A}_1 - m\mathbf{I} + \mathbf{A}_2m^{-1} & \mathbf{0} \\ \mathbf{0} & -m\mathbf{I} \end{pmatrix} \end{aligned}$$

and take the determinant of both sides.

Let the  $n$ th order vector difference equation be given by

$$\mathbf{y}_t + \mathbf{A}_1\mathbf{y}_{t-1} + \mathbf{A}_2\mathbf{y}_{t-2} + \cdots + \mathbf{A}_n\mathbf{y}_{t-n} = \mathbf{r}_t, \quad t = 1, 2, \dots, \quad (2.4.33)$$

where  $\mathbf{y}_t$  is a  $k$ -vector and  $\mathbf{r}_t$  is a real vector valued function of time. The associated  $\mathbf{A}$ -matrix is

$$\mathbf{A} = \begin{pmatrix} -\mathbf{A}_1 & -\mathbf{A}_2 & \cdots & -\mathbf{A}_{n-1} & -\mathbf{A}_n \\ \mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I} & \mathbf{0} \end{pmatrix}.$$

Then the solution is

$$\mathbf{y}_t = \sum_{j=0}^{t-1} \mathbf{W}_j \mathbf{r}_{t-j} + \tau_t,$$

where the  $\mathbf{W}_j$  are  $k \times k$  matrices defined by

$$\begin{aligned}\mathbf{W}_0 &= \mathbf{I}, \\ \mathbf{W}_1 + \mathbf{A}_1 &= \mathbf{0}, \\ \mathbf{W}_2 + \mathbf{A}_1\mathbf{W}_1 + \mathbf{A}_2 &= \mathbf{0}, \\ &\vdots \\ \mathbf{W}_j + \mathbf{A}_1\mathbf{W}_{j-1} + \cdots + \mathbf{A}_n\mathbf{W}_{j-n} &= \mathbf{0}, \quad j = n, n+1, \dots,\end{aligned}$$

$\tau_t$  is the vector composed of the first  $k$  entries in the  $nk$ -vector  $\mathbf{x}_t = \mathbf{A}'\mathbf{x}_0$ , and  $\mathbf{x}'_0 = (\mathbf{y}'_0, \mathbf{y}'_{-1}, \dots, \mathbf{y}'_{1-n})$ . It follows that  $\mathbf{W}_j$  is the upper left block of the matrix  $\mathbf{A}'$ .

## 2.5. THE SECOND ORDER AUTOREGRESSIVE TIME SERIES

We consider the stationary second order autoregressive time series defined by the stochastic difference equation

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = e_t, \quad t \in (0, \pm 1, \pm 2, \dots), \quad (2.5.1)$$

where the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables,  $\alpha_2 \neq 0$ , and the roots of  $m^2 + \alpha_1 m + \alpha_2 = 0$  are less than one in absolute value. Recalling that we were able to express the first order autoregressive process as a weighted average of past values of  $e_t$ , we postulate that  $X_t$  can be expressed as

$$X_t = \sum_{j=0}^{\infty} w_j e_{t-j}, \quad (2.5.2)$$

where the  $w_j$  are functions of  $\alpha_1$  and  $\alpha_2$ . We display this representation as

$$\begin{aligned}X_t &= w_0 e_t + w_1 e_{t-1} + w_2 e_{t-2} + w_3 e_{t-3} + \cdots, \\ X_{t-1} &= w_0 e_{t-1} + w_1 e_{t-2} + w_2 e_{t-3} + \cdots, \\ X_{t-2} &= w_0 e_{t-2} + w_1 e_{t-3} + \cdots.\end{aligned}$$

If the  $X_t$  are to satisfy the difference equation (2.5.1), then we must have

$$\begin{aligned}w_0 &= 1, \\ w_1 + \alpha_1 w_0 &= 0, \\ w_j + \alpha_1 w_{j-1} + \alpha_2 w_{j-2} &= 0, \quad j = 2, 3, \dots.\end{aligned} \quad (2.5.3)$$

Thus the  $w_j$  are given by the general solution to the second order linear homogeneous difference equation subject to the initial conditions specified by the equations in  $w_0$  and  $w_1$ . Therefore, if the roots are distinct, we have

$$w_j = (m_1 - m_2)^{-1} m_1^{j+1} + (m_2 - m_1)^{-1} m_2^{j+1}, \quad j = 0, 1, 2, \dots,$$

and if the roots are equal,

$$w_j = (1 + j)m^j, \quad j = 0, 1, 2, \dots$$

In Theorem 2.6.1 of the next section, we prove that the representation (2.5.2) holds for a stationary time series satisfying (2.5.1).

The covariance function for the second order autoregressive time series defined in (2.5.2) is

$$\gamma(h) = \sigma^2 \sum_{j=h}^{\infty} w_j w_{j-h}, \quad h = 0, 1, 2, \dots \quad (2.5.4)$$

We shall investigate the covariance function in a slightly different manner. Multiply (2.5.1) by  $X_{t-h}$  ( $h \geq 0$ ) to obtain

$$X_t X_{t-h} + \alpha_1 X_{t-1} X_{t-h} + \alpha_2 X_{t-2} X_{t-h} = e_t X_{t-h}. \quad (2.5.5)$$

Because  $X_t$  is expressible as a weighted average of  $e_t$  and previous  $e$ 's,

$$E\{X_{t-j} e_i\} = 0, \quad j \geq 1.$$

Thus, taking the expectation of both sides of (2.5.5), we have

$$\gamma(h) + \alpha_1 \gamma(h-1) + \alpha_2 \gamma(h-2) = \begin{cases} 0, & h > 0, \\ \sigma^2, & h = 0. \end{cases} \quad (2.5.6)$$

That is, the covariance function, for  $h > 0$ , satisfies the homogeneous difference equation associated with the stochastic difference equation defining the time series.

The equations (2.5.6) are called the *Yule-Walker equations*. With the aid of these equations we can obtain a number of alternative expressions for the autocovariances, autocorrelations, and coefficients of the original representation. Using the two equations associated with  $h = 1$  and  $h = 2$ , we may solve for  $\alpha_1$  and  $\alpha_2$  as functions of the autocovariances:

$$\alpha_1 = \frac{\gamma(1)\gamma(2) - \gamma(0)\gamma(1)}{\gamma^2(0) - \gamma^2(1)}, \quad \alpha_2 = \frac{\gamma^2(1) - \gamma(0)\gamma(2)}{\gamma^2(0) - \gamma^2(1)}. \quad (2.5.7)$$

If we use the three equations (2.5.6) associated with  $h = 0, 1, 2$ , we can obtain expressions for the autocovariances in terms of the coefficients:

$$\begin{aligned} \gamma(0) &= \frac{(1 + \alpha_2)\sigma^2}{(1 - \alpha_2)[(1 + \alpha_2)^2 - \alpha_1^2]}, \\ \gamma(1) &= \frac{-\alpha_1\sigma^2}{(1 - \alpha_2)[(1 + \alpha_2)^2 - \alpha_1^2]}, \\ \gamma(2) &= \frac{[\alpha_1^2 - \alpha_2(1 + \alpha_2)]\sigma^2}{(1 - \alpha_2)[(1 + \alpha_2)^2 - \alpha_1^2]}. \end{aligned} \quad (2.5.8)$$

Using  $\gamma(0)$  and  $\gamma(1)$  of (2.5.8) as the initial conditions together with the general solution of the homogeneous equation, we can obtain the expression for the covariance function.

However, it is somewhat simpler to derive the correlation function first. To obtain  $\rho(1)$  as a function of the coefficients, we divide equation (2.5.6) for  $h = 1$  by  $\gamma(0)$ , giving

$$\rho(1) + \alpha_1 + \alpha_2 \rho(1) = 0$$

and

$$\rho(1) = -\frac{\alpha_1}{1 + \alpha_2} = \frac{m_1 + m_2}{1 + m_1 m_2}. \quad (2.5.9)$$

Using (2.5.9) and  $\rho(0) = 1$  as initial conditions, we have, for unequal roots (real or complex),

$$\begin{aligned} \rho(h) &= [(m_1 - m_2)(1 + m_1 m_2)]^{-1} [m_1^{h+1}(1 - m_2^2) - m_2^{h+1}(1 - m_1^2)], \\ h &= 0, 1, 2, \dots. \end{aligned} \quad (2.5.10)$$

If the roots are complex, the autocorrelation function may also be expressed as

$$\rho(h) = \frac{r^h \sin(h\theta + \delta)}{\sin \delta}, \quad (2.5.11)$$

where

$$r = \alpha_2^{1/2} = (m_1 m_2)^{1/2},$$

$$\begin{aligned} \cos \theta &= -\frac{\alpha_1}{2\alpha_2^{1/2}} = \frac{m_1 + m_2}{2(m_1 m_2)^{1/2}}, \\ \tan \delta &= \frac{1 + \alpha_2}{1 - \alpha_2} \tan \theta. \end{aligned}$$

For roots real and equal the autocorrelation function is given by

$$\rho(h) = \left[ 1 + h \left( \frac{1 - m^2}{1 + m^2} \right) \right] m^h, \quad h = 0, 1, 2, \dots. \quad (2.5.12)$$

If we substitute  $\alpha_1 = -(m_1 + m_2)$  and  $\alpha_2 = m_1 m_2$  into (2.5.8), we can express the variance as a function of the roots:

$$\gamma(0) = \frac{1 + m_1 m_2}{(1 - m_1 m_2)(1 - m_1^2)(1 - m_2^2)} \sigma^2. \quad (2.5.13)$$

Equation (2.5.13) together with (2.5.10) and (2.5.12) may be used to express the covariance function in terms of the roots.

**Example 2.5.1.** Figure 2.5.1 contains a plot of the correlation function, sometimes called the *correlogram*, for the time series

$$X_t = 1.40X_{t-1} - 0.48X_{t-2} + e_t.$$

The roots of the auxiliary equation are 0.8 and 0.6. The correlogram has much the same appearance as that of a first order autoregressive process with large  $\rho$ . The empirical correlograms of many economic time series have this general appearance, the first few correlations being quite large.

The correlogram in Figure 2.5.2 is that associated with the time series

$$X_t = X_{t-1} - 0.89X_{t-2} + e_t.$$

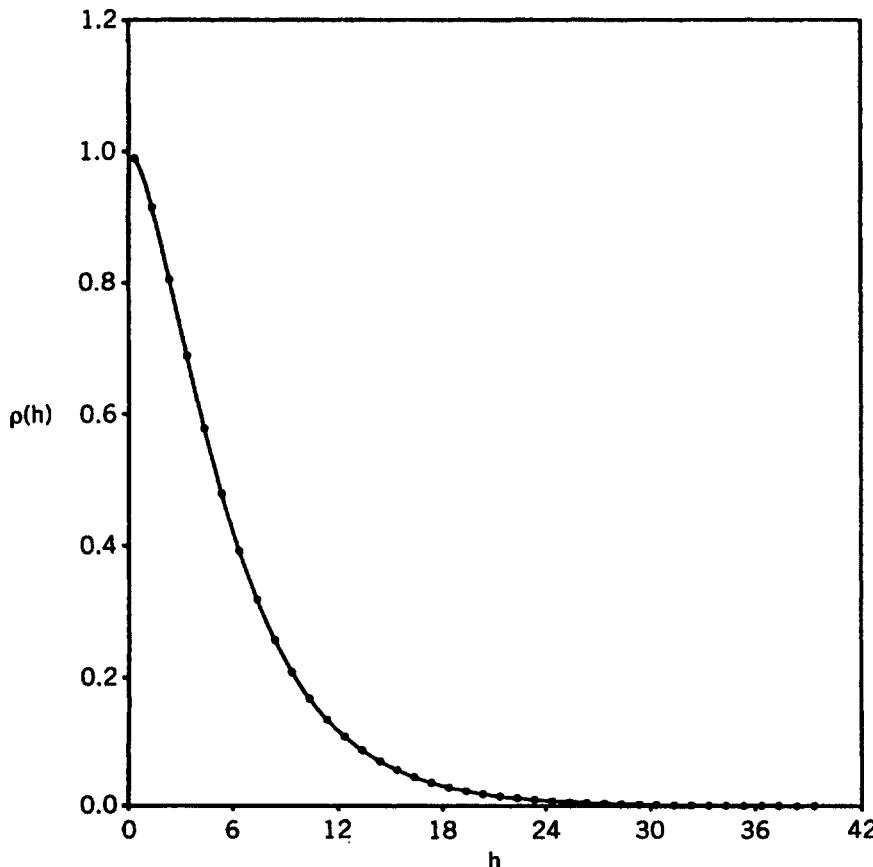


Figure 2.5.1. Correlogram for  $X_t = 1.40X_{t-1} - 0.48X_{t-2} + e_t$ .

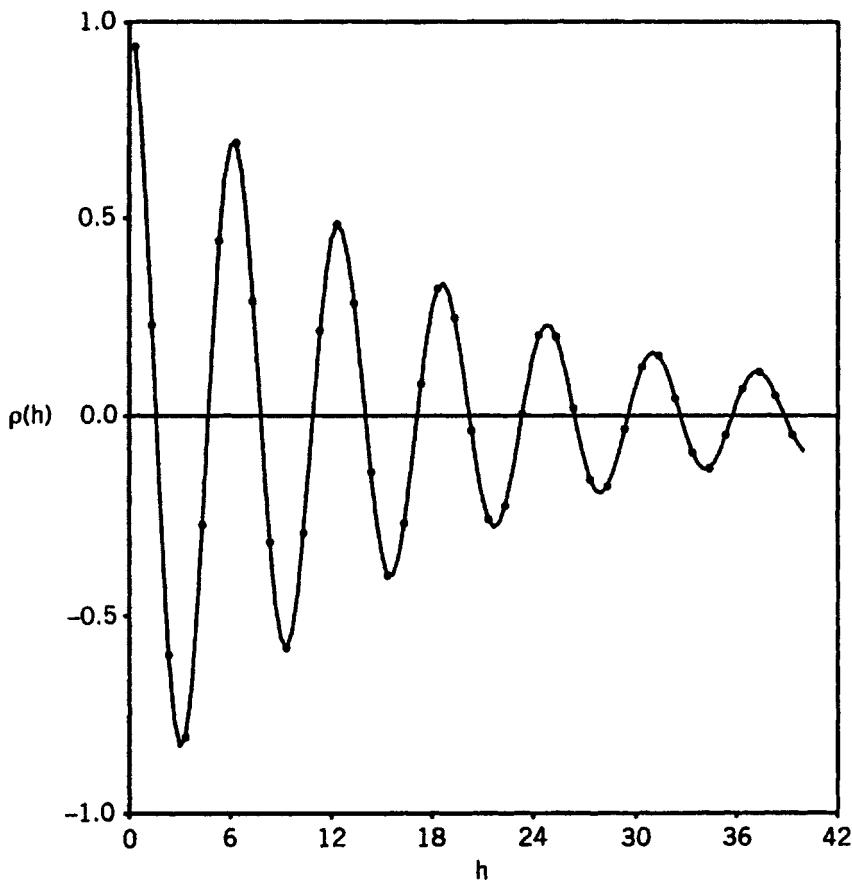


Figure 2.5.2. Correlogram for  $X_t = X_{t-1} - 0.89X_{t-2} + e_t$ .

The roots of the auxiliary equation are the complex pair  $0.5 \pm 0.8i$ . In this case the correlogram has the distinctive "declining cyclical" appearance associated with complex roots. This correlogram illustrates how such a process can generate a time series with the appearance of moderately regular "cycles." In the present case  $\cos \theta = 0.53$  and  $\theta = 0.322 \pi$ , where  $\cos \theta$  is defined in (2.5.11). Thus the apparent period of the cyclical behavior would be  $2(0.322)^{-1} \approx 6.2$  time units.  $\blacktriangle \blacktriangle$

## 2.6. ALTERNATIVE REPRESENTATIONS OF AUTOREGRESSIVE AND MOVING AVERAGE PROCESSES

We have argued that the stationary second order autoregressive time series can also be represented as an infinite moving average time series. We now prove that a time series satisfying a  $p$ th order difference equation has a representation as an infinite moving average.

**Theorem 2.6.1.** Let  $\{X_t\}$  be a time series defined on the integers with  $E\{X_t^2\} < K$  for all  $t$ . Suppose  $X_t$  satisfies

$$X_t + \sum_{j=1}^p \alpha_j X_{t-j} = e_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad (2.6.1)$$

where  $\{e_t\}$ ,  $t = 0, \pm 1, \pm 2, \dots$ , is a sequence of uncorrelated  $(0, \sigma^2)$  random variables. Let  $m_1, m_2, \dots, m_p$  be the roots of

$$m^p + \sum_{j=1}^p \alpha_j m^{p-j} = 0, \quad (2.6.2)$$

and assume  $|m_i| < 1$ ,  $i = 1, 2, \dots, p$ . Then  $X_t$  is covariance stationary. Furthermore,  $X_t$  is given as a limit in mean square by

$$X_t = \sum_{j=0}^{\infty} w_j e_{t-j}, \quad (2.6.3)$$

where  $\{w_j\}_{j=0}^{\infty}$  is the unique solution of the homogeneous difference equation

$$w_j + \alpha_1 w_{j-1} + \cdots + \alpha_p w_{j-p} = 0, \quad j = p, p+1, \dots, \quad (2.6.4)$$

subject to the boundary conditions  $w_0 = 1$  and

$$w_j + \sum_{i=1}^j \alpha_i w_{j-i} = 0, \quad j = 1, 2, \dots, p-1.$$

The limit (2.6.3) is also an almost sure limit.

**Proof.** Let  $\mathbf{Y}_t = (X_t, X_{t-1}, \dots, X_{t-p+1})'$ ,  $\boldsymbol{\epsilon}_t = (e_t, 0, \dots, 0)'$ , and

$$\mathbf{A} = \begin{pmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & \cdots & -\alpha_{p-1} & -\alpha_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}. \quad (2.6.5)$$

Then

$$\mathbf{Y}_t = \mathbf{A} \mathbf{Y}_{t-1} + \boldsymbol{\epsilon}_t = \mathbf{A}^n \mathbf{Y}_{t-n} + \sum_{j=0}^{n-1} \mathbf{A}^j \boldsymbol{\epsilon}_{t-j}$$

for  $n \geq 1$  and  $t = 0, \pm 1, \pm 2, \dots$ . Therefore,

$$\begin{aligned} X_t &= A_{(11)}^n X_{t-n} + A_{(12)}^n X_{t-n-1} + \cdots + A_{(1p)}^n X_{t-n-p+1} + \sum_{j=0}^{n-1} A_{(11)}^j e_{t-j}, \\ &= \sum_{j=1}^p A_{(1j)}^n X_{t-n-j+1} + \sum_{j=0}^{n-1} A_{(11)}^j e_{t-j}, \end{aligned}$$

where  $A_{(ij)}^r$  is the  $ij$ th element of the matrix  $\mathbf{A}^r$ . Consider the determinant

$$|\lambda \mathbf{I} - \mathbf{A}| = \begin{vmatrix} \lambda + \alpha_1 & \alpha_2 & \alpha_3 & \cdots & \alpha_{p-1} & \alpha_p \\ -1 & \lambda & 0 & \cdots & 0 & 0 \\ 0 & -1 & \lambda & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & \lambda \end{vmatrix}.$$

Beginning with column 1, successively multiply column  $i$  by  $\lambda$  and add the result to column  $i+1$  to obtain

$$\begin{aligned} |\lambda \mathbf{I} - \mathbf{A}| &= \begin{vmatrix} \lambda + \alpha_1 & \sum_{j=0}^2 \lambda^{2-j} \alpha_j & \sum_{j=0}^3 \lambda^{3-j} \alpha_j & \cdots & \sum_{j=0}^p \lambda^{p-j} \alpha_j \\ -1 & 0 & 0 & \cdots & 0 \\ 0 & -1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 0 \end{vmatrix} \\ &= (\lambda^p + \alpha_1 \lambda^{p-1} + \cdots + \alpha_{p-1} \lambda + \alpha_p) (-1)^{2p-1}. \end{aligned}$$

By the Cayley-Hamilton theorem,

$$\mathbf{A}^p + \alpha_1 \mathbf{A}^{p-1} + \cdots + \alpha_{p-1} \mathbf{A} + \alpha_p \mathbf{I} = \mathbf{0}. \quad (2.6.6)$$

For  $j \geq p$  multiply the matrix equation (2.6.6) by  $\mathbf{A}^{j-p}$  to obtain

$$\mathbf{A}^j + \alpha_1 \mathbf{A}^{j-1} + \cdots + \alpha_{p-1} \mathbf{A}^{j-p+1} + \alpha_p \mathbf{A}^{j-p} = \mathbf{0}, \quad j \geq p. \quad (2.6.7)$$

Set  $w_j = A_{(11)}^j$ , the first element of  $\mathbf{A}^j$ . Then

$$w_j + \alpha_1 w_{j-1} + \alpha_2 w_{j-2} + \cdots + \alpha_p w_{j-p} = 0, \quad j \geq p.$$

It is readily verified that  $w_j = A_{(1j)}^j$  satisfies the initial conditions associated with (2.6.4) in the statement of the theorem. Therefore,

$$X_t = \sum_{j=1}^p A_{(1j)}^n X_{t-n-j+1} + \sum_{j=0}^{n-1} w_j e_{t-j}$$

and

$$E\left[\left(X_t - \sum_{j=0}^{n-1} w_j e_{t-j}\right)^2\right] = E\left[\left(\sum_{j=1}^p A_{(1j)}^n X_{t-n-j+1}\right)^2\right]$$

for  $n = 1, 2, \dots$ . Because each element of  $\mathbf{A}^n$  satisfies the homogeneous difference equation (2.6.7), we have

$$A_{(1j)}^j + \alpha_1 A_{(1j)}^{j-1} + \cdots + \alpha_p A_{(1j)}^{j-p} = 0 \quad \text{for } j \geq p,$$

and there exists a  $c$  such that  $|A_{(1j)}^j| < c\lambda^j$  for  $i = 1, 2, \dots, p$ , where  $1 > \lambda > M$  and  $M$  is the largest of the absolute values of the roots  $m_i$ . See Exercise 2.24. It follows that

$$\begin{aligned} E\left[\left(X_t - \sum_{j=0}^{n-1} w_j e_{t-j}\right)^2\right] &\leq (cp\lambda^n)^2 K \\ &\rightarrow 0 \end{aligned}$$

as  $n \rightarrow \infty$ , because  $0 < \lambda < 1$ . Because  $\sum_{j=0}^{\infty} |w_j| < \infty$ ,  $X_t$  is covariance stationary. Now,

$$\begin{aligned} E\left\{\left|X_t - \sum_{j=0}^{n-1} w_j e_{t-j}\right|\right\} &\leq \sum_{j=1}^p |A_{(1j)}^n| E\{|X_{t-n-j+1}|\} \\ &\leq cp\lambda^n K^{1/2}, \end{aligned}$$

and it follows that the limit (2.6.3) holds almost surely. ▲

Because  $X_t$  can be written as a weighted sum of current and past  $e_t$ 's, it follows that  $e_t$  is uncorrelated with  $X_{t-h}$  for  $h = 1, 2, \dots$ . This enables us to derive many useful properties of the autoregressive time series and of the autocovariance function of the autoregressive time series.

**Corollary 2.6.1.1.** Let  $X_t$  be stationary and satisfy

$$X_t + \alpha_1 X_{t-1} + \cdots + \alpha_p X_{t-p} = e_t,$$

where the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables and the roots of the characteristic polynomial

$$m^p + \alpha_1 m^{p-1} + \cdots + \alpha_p = 0$$

are less than one in absolute value. Then

$$\gamma(0) + \alpha_1 \gamma(1) + \cdots + \alpha_p \gamma(p) = \sigma^2$$

and

$$\gamma(h) + \alpha_1 \gamma(h-1) + \cdots + \alpha_p \gamma(h-p) = 0, \quad h = 1, 2, \dots \quad (2.6.8)$$

**Proof.** Reserved for the reader. See equation (2.5.6).  $\blacktriangle$

The system (2.6.8) defines the  $\alpha$ 's in terms of the  $\gamma$ 's and vice versa. For example, for a stationary third order autoregressive process, we have

$$\begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) \\ \gamma(1) & \gamma(0) & \gamma(1) \\ \gamma(2) & \gamma(1) & \gamma(0) \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \\ \alpha_3 \end{pmatrix} = \begin{pmatrix} -\gamma(1) \\ -\gamma(2) \\ -\gamma(3) \end{pmatrix} \quad (2.6.9)$$

and

$$\begin{pmatrix} 1 & \alpha_1 & \alpha_2 & \alpha_3 \\ \alpha_1 & 1 + \alpha_2 & \alpha_3 & 0 \\ \alpha_2 & \alpha_1 + \alpha_3 & 1 & 0 \\ \alpha_3 & \alpha_2 & \alpha_1 & 1 \end{pmatrix} \begin{pmatrix} \gamma(0) \\ \gamma(1) \\ \gamma(2) \\ \gamma(3) \end{pmatrix} = \begin{pmatrix} \sigma^2 \\ 0 \\ 0 \\ 0 \end{pmatrix}. \quad (2.6.10)$$

Using the representation results, we show that the partial autocorrelation function, introduced in equation (1.4.12), of a  $p$ th order autoregressive process is zero for  $k > p$ .

**Corollary 2.6.1.2.** Let  $X_t$  be a stationary  $p$ th order autoregressive process with all of the roots of the characteristic polynomial less than one in absolute value, and let  $\phi(k)$  be the partial autocorrelation function. Then

$$\phi(k) = 0, \quad k = p + 1, p + 2, \dots$$

**Proof.** By the definition of the partial autocorrelation,  $\phi(p+1)$  is the correlation between the residual obtained in the population regression of  $X_t$  on  $X_{t-1}, X_{t-2}, \dots, X_{t-p}$  and the residual obtained in the population regression of  $X_{t-p-1}$  on  $X_{t-1}, X_{t-2}, \dots, X_{t-p}$ . By the definition of the autoregressive process, the residual obtained in the population regression of  $X_t$  on  $X_{t-1}, X_{t-2}, \dots, X_{t-p}$  is  $e_t$ , and the coefficients are  $-\alpha_i$ ,  $i = 1, 2, \dots, p$ . By the representation of Theorem 2.6.1,  $e_t$  is uncorrelated with  $X_{t-i}$ ,  $i \geq 1$  and hence with a linear combination of the

$X_{t-i}$ ,  $i \geq 1$ . Therefore,  $\theta_{kk}$  of (1.4.11), and hence  $\phi(k)$ , is zero for  $k = p + 1$ . If  $\theta_{ik} = -\alpha_i$  for  $i = 1, 2, \dots, p$  and  $\theta_{ik} = 0$  for  $i = p + 1, \dots, k$ , then  $\theta_{k+1,k+1} = 0$  and the result follows.  $\blacktriangle$

Whenever discussing stationary autoregressive time series, we have assumed that the roots of the auxiliary equation were less than one in absolute value. This is because we have explicitly or implicitly visualized the time series as being created in a forward manner. If a time series is created in a forward manner, the effect of the initial conditions will go to zero only if the roots are less than one in absolute value. For example, if we define

$$X_t = \begin{cases} a_0 e_0, & t = 0, \\ \rho X_{t-1} + e_t, & t = 1, 2, \dots, \end{cases} \quad (2.6.11)$$

where it is understood that  $X_0$  is formed by adding  $e_0$  to  $\rho X_{-1}$  and  $e_t$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables, then  $\{X_t : t = 0, 1, 2, \dots\}$  is stationary for  $|\rho| < 1$  and  $a_0 = (1 - \rho^2)^{-1/2}$ . However, for  $|\rho| \geq 1$ , the time series formed by adding  $e_t$  to  $\rho X_{t-1}$  is nonstationary for all  $a_0$ .

On the other hand, there is a stationary time series that satisfies the difference equation

$$X_t = \rho X_{t-1} + e_t, \quad t = 0, \pm 1, \pm 2, \dots, \quad (2.6.12)$$

for  $|\rho| > 1$ . To see this, let us consider the stationary time series

$$\begin{aligned} X_r &= 0.8X_{r-1} + \epsilon_r \\ &= \sum_{j=0}^{\infty} (0.8)^j \epsilon_{r-j}, \quad r = 0, \pm 1, \pm 2, \dots, \end{aligned} \quad (2.6.13)$$

where  $\epsilon_r$  are uncorrelated  $(0, \sigma^2)$  random variables. If we change the direction in which we count on the integers so that  $-r = t$  and  $r - 1 = -t + 1$ , and divide (2.6.13) by 0.8, we have

$$X_{t+1} = 1.25X_t - 1.25\epsilon_t \quad (2.6.14)$$

and

$$X_t = \sum_{j=0}^{\infty} (0.8)^j \epsilon_{t+j}. \quad (2.6.15)$$

By setting  $\epsilon_{t+1} = -1.25\epsilon_t$ , equation (2.6.14) can be written in the form (2.6.12) with  $\rho = 1.25$ .

The covariance function for the time series (2.6.15) is the same as the covariance function of the time series (2.6.13). Thus, if a time series  $\{X_t : t \in (0, \pm 1, \pm 2, \dots)\}$  has an autocorrelation function of the form  $\rho^{|h|}$ , where  $0 < |\rho| < 1$ , it can be written as a forward moving average of uncorrelated random variables

or as a backward moving average of uncorrelated random variables. Likewise,

$$X_t - \rho X_{t-1} = e_t, \quad (2.6.16)$$

defines a sequence of uncorrelated random variables, as does

$$X_t - \rho^{-1} X_{t-1} = Z_t. \quad (2.6.17)$$

From the representations (2.6.13) and (2.6.15), one sees that the only stationary time series with a unit root is the trivial time series that is a constant (with probability one) for all  $t$ . See Exercise 2.19.

While both the  $e_t$  of (2.6.16) and the  $Z_t$  of (2.6.17) are uncorrelated random variables, the variance of  $Z_t$  is larger than the variance of the  $e_t$ , by a factor of  $\rho^{-2}$ . This explains why the representation (2.6.16) is the one that appears in the applications of stationary autoregressive processes. That is, one typically chooses the  $\rho$  in an equation such as (2.6.16) to minimize the variance of  $e_t$ . It is worth mentioning that nonstationary autoregressive representations with roots greater than or equal to one in absolute value have appeared in practice. Estimation for such time series is discussed in Chapter 10. That the stationary autoregressive time series can be given either a forward or a backward representation also finds some use, and we state the result before proceeding.

**Corollary 2.6.1.3.** Let the covariance function of the time series  $\{X_t : t \in (0, \pm 1, \pm 2, \dots)\}$  with zero mean satisfy the difference equation

$$\sum_{j=0}^p \alpha_j \gamma_X(h-j) = 0, \quad h = 1, 2, \dots,$$

where  $\alpha_0 = 1$  and the roots of the characteristic equation

$$\sum_{j=0}^p \alpha_j m^{p-j} = 0$$

are less than one in absolute value. Then  $X_t$  satisfies the stochastic difference equation

$$\sum_{j=0}^p \alpha_j X_{t-j} = e_t,$$

where  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables, and also satisfies the stochastic difference equation

$$\sum_{j=0}^p \alpha_j X_{t+j} = v_t,$$

where  $\{v_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables.

**Proof.** Omitted. ▲

Having demonstrated that a stationary finite autoregressive time series can be given an infinite moving average representation, we now obtain an alternative representation for the finite moving average time series. Because the finite moving average time series can be viewed as a difference equation in  $e_t$ , we have a result parallel to that of Theorem 2.6.1.

**Theorem 2.6.2.** Let the time series  $\{X_t : t \in (0, \pm 1, \dots)\}$  be defined by

$$X_t = e_t + b_1 e_{t-1} + b_2 e_{t-2} + \cdots + b_q e_{t-q}, \quad t = 0, \pm 1, \pm 2, \dots,$$

where  $b_q \neq 0$ , the roots of the characteristic equation

$$m^q + b_1 m^{q-1} + b_2 m^{q-2} + \cdots + b_q = 0$$

are less than one in absolute value, and  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables. Then  $X_t$  can be expressed as an infinite autoregressive process

$$\sum_{j=0}^{\infty} c_j X_{t-j} = e_t, \quad (2.6.18)$$

where the coefficients  $c_j$  satisfy the homogeneous difference equation

$$c_j + b_1 c_{j-1} + b_2 c_{j-2} + \cdots + b_q c_{j-q} = 0, \quad j = q, q+1, \dots, \quad (2.6.19)$$

with the initial conditions  $c_0 = 1$ ,  $c_1 = -b_1$ ,

$$c_2 = -b_1 c_1 - b_2,$$

$$\vdots$$

$$c_{q-1} = -b_1 c_{q-2} - b_2 c_{q-3} - \cdots - b_{q-1}.$$

**Proof.** Reserved for the reader. ▲

The autocorrelation function of the  $p$ th autoregressive process is nonzero for some integer greater than  $N_0$  for all  $N_0$ . See Exercise 2.33. By Corollary 2.6.1.2, the partial autocorrelation function of the  $p$ th order autoregressive process is zero for argument greater than  $p$ . The opposite conditions hold for the autocorrelations of the moving average process.

**Corollary 2.6.2.1.** Let the time series  $X_t$  be the  $q$ th order moving average defined in Theorem 2.6.2. Let  $N_0$  be given. Then there is some  $k > N_0$  such that the partial autocorrelation function  $\phi(k) \neq 0$ .

**Proof.** By our definition,  $X_t = \sum_{i=0}^q b_i e_{t-i}$ , where  $b_0 = 1$  and  $b_q \neq 0$ . If  $\phi(k) =$

0 for all  $k > N_0$ , then the coefficients  $c_j$  in Theorem 2.6.2 must be zero for  $j > N_0$ . By Theorem 2.6.2 the coefficients  $c_j$  satisfy the difference equation (2.6.19). If all  $c_j$  are equal to zero for  $j > N_0$ , then  $c_{N_0}$  must be zero if it is to satisfy (2.6.19). Because  $b_q \neq 0$ , this leads to the conclusion that all  $c_j$  are zero, which contradicts the initial condition for  $(c_0, c_1, \dots, c_{q-1})$ .  $\blacktriangle$

In discussing finite moving averages we placed no restrictions on the coefficients, and, for example, the time series

$$Y_t = e_t - e_{t-1} \quad (2.6.20)$$

is clearly stationary. The root of the auxiliary equation for this difference equation is one, and therefore the condition of Theorem 2.6.2 is not met. An attempt to express  $e_t$  as an autoregressive process using that theorem will fail because the remainder associated with an autoregressive representation of order  $n$  will be  $e_{t-n-1}$ .

Time series satisfying the conditions of Theorem 2.6.2 are sometimes called *invertible* moving averages. From the example (2.6.20) we see that not all moving average processes are invertible.

In our earlier discussion of the moving average time series we demonstrated that we could always assign the value one to the coefficient of  $e_t$ . We were also able to obtain all autocorrelation functions of the first order moving average type for an  $\alpha$  restricted to the range  $[-1, 1]$ . We are now in a position to generalize this result.

**Theorem 2.6.3.** Given a time series  $X_t$  with zero mean and autocorrelation function

$$\rho_X(h) = \begin{cases} 1, & h = 0, \\ \rho(1), & h = 1, \\ 0, & h > 1, \end{cases}$$

where  $|\rho(1)| < 0.5$ , there exists an  $\alpha$ ,  $|\alpha| < 1$ , and a sequence of uncorrelated random variables  $\{e_t\}$  such that  $X_t$  is defined by

$$X_t = e_t + \alpha e_{t-1}.$$

**Proof.** The equation  $\rho(1) = (1 + \alpha^2)^{-1}\alpha$  in  $\alpha$  has one root that is less than one in absolute value and one that is greater than one in absolute value. The root of smaller absolute value is chosen, and we define  $e_t$  by

$$e_t = \sum_{j=0}^{\infty} (-\alpha)^j X_{t-j}.$$

By Theorem 2.2.1 this random variable is well defined as a limit in squared mean,

and by Theorem 2.2.2,

$$\begin{aligned}\gamma_e(0) &= E\{e_i^2\} = \sum_{j=0}^{\infty} (-\alpha)^{2j} \gamma_X(0) - 2\alpha \sum_{j=0}^{\infty} (-\alpha)^{2j} \gamma_X(1) \\ &= \left( \frac{1}{1-\alpha^2} - \frac{2\alpha}{1-\alpha^2} \frac{\alpha}{1+\alpha^2} \right) \gamma_X(0) = \frac{\gamma_X(0)}{1+\alpha^2}, \\ \gamma_e(h) &= \sum_{j=0}^{\infty} (-\alpha)^{h+2j} \gamma_X(0) + \sum_{j=0}^{\infty} (1+\alpha^2)(-\alpha)^{h-1+2j} \gamma_X(1) \\ &= \left( \frac{(-\alpha)^h}{1-\alpha^2} + \frac{(1+\alpha^2)(-\alpha)^{h-1}}{1-\alpha^2} \frac{\alpha}{1+\alpha^2} \right) \gamma_X(0) \\ &= 0, \quad h > 0.\end{aligned}$$

▲

The dualities between autoregressive and moving average representations established in Theorems 2.6.1 and 2.6.2 can also be described using formal operations with the backward shift operator. We recall that the first order autoregressive time series

$$Y_t - \rho Y_{t-1} = (1 - \rho \mathcal{B})Y_t = e_t, \quad |\rho| < 1,$$

can be written as

$$Y_t = \sum_{j=0}^{\infty} \rho^j e_{t-j} = \left[ \sum_{j=0}^{\infty} (\mathcal{B}\rho)^j \right] e_t = \frac{1}{1 - \rho \mathcal{B}} e_t,$$

where it is understood that  $\mathcal{B}^0$  is the identity operator, and we have written

$$(1 - \rho \mathcal{B})^{-1} = \sum_{j=0}^{\infty} (\mathcal{B}\rho)^j$$

in analogy to the expansion that holds for a real number  $|\rho \mathcal{B}| < 1$ .

The fact that the operator can be formally manipulated as if it were a real number with absolute value 1 furnishes a useful way of obtaining the alternative expressions for moving average and autoregressive processes whose characteristic equations have roots less than one in absolute value. Thus the second order autoregressive time series

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} = e_t,$$

can also be written as

$$(1 + a_1 \mathcal{B} + a_2 \mathcal{B}^2)X_t = e_t,$$

or

$$(1 - m_1 \mathcal{B})(1 - m_2 \mathcal{B})X_t = e_t$$

or

$$X_t = [(1 - m_1 \mathcal{B})(1 - m_2 \mathcal{B})]^{-1} e_t,$$

where  $m_1$  and  $m_2$  are the roots of the characteristic equation

$$m^2 + a_1 m + a_2 = 0.$$

Since the roots of the quadratic in  $\mathcal{B}$

$$1 + a_1 \mathcal{B} + a_2 \mathcal{B}^2 = 0$$

are the reciprocals of the roots of the characteristic equation, the restriction on the roots is often stated as the requirement that the roots of

$$1 + a_1 \mathcal{B} + a_2 \mathcal{B}^2 = 0$$

be greater than one in absolute value.

Using the backward shift operator, the invertible second order moving average

$$Y_t = e_t + b_1 e_{t-1} + b_2 e_{t-2}$$

can be given the representation

$$Y_t = (1 + b_1 \mathcal{B} + b_2 \mathcal{B}^2) e_t$$

or

$$Y_t = (1 - g_1 \mathcal{B})(1 - g_2 \mathcal{B}) e_t,$$

where  $g_1$  and  $g_2$ , the roots of

$$g^2 + b_1 g + b_2 = 0,$$

are less than one in absolute value. The autoregressive representation of the moving average process is then given by

$$e_t = [(1 - g_1 \mathcal{B})(1 - g_2 \mathcal{B})]^{-1} Y_t.$$

The backward shift representation of the moving average process is useful in illustrating the fact that any finite moving average process whose characteristic equation has some roots greater than one and some less than one can be given a representation whose characteristic equation has all roots less than one in absolute value.

**Theorem 2.6.4.** Let  $\{X_t : t \in (0, \pm 1, \pm 2, \dots)\}$  have the representation

$$X_t = \prod_{i=1}^p (1 - m_i \mathcal{B}) e_t,$$

where the  $\epsilon_i$  are uncorrelated  $(0, \sigma^2)$  random variables and none of the  $m_i$ ,  $i = 1, 2, \dots, p$ , are of unit absolute value. Let  $m_1, m_2, \dots, m_L$ ,  $0 < L \leq p$ , be greater than one in absolute value. Then  $X_t$  also has the representation

$$X_t = \begin{cases} \prod_{i=1}^L (1 - m_i^{-1} \mathcal{B}) \prod_{i=L+1}^p (1 - m_i \mathcal{B}) \epsilon_i, & 0 < L < p \\ \prod_{i=1}^p (1 - m_i^{-1} \mathcal{B}) \epsilon_i, & L = p, \end{cases}$$

where the  $\epsilon_i$  are uncorrelated  $(0, \sigma^2 [\prod_{i=1}^L m_i]^2)$ .

**Proof.** We define the polynomial  $\mathcal{Q}(\mathcal{B})$  in  $\mathcal{B}$  by

$$\mathcal{Q}(\mathcal{B}) = \begin{cases} \prod_{i=1}^L (1 - m_i^{-1} \mathcal{B}) \prod_{i=L+1}^p (1 - m_i \mathcal{B}), & 0 < L < p, \\ \prod_{i=1}^p (1 - m_i^{-1} \mathcal{B}), & L = p. \end{cases}$$

The roots of  $\mathcal{Q}(\mathcal{B})$  are greater than one in absolute value, and the roots of  $m^p \mathcal{Q}(m^{-1})$  are less than one in absolute value. Therefore the time series

$$\epsilon_t = \sum_{j=0}^{\infty} w_j X_{t-j},$$

where the  $w_j$  satisfy the difference equation

$$\mathcal{Q}(\mathcal{B})w_j = 0, \quad j = p, p+1, \dots,$$

and the initial conditions outlined in Theorem 2.6.2, is well defined. Let  $m_1$  be a real root greater than one in absolute value, and define  $Z_t$  by

$$Z_t = \prod_{i=L+1}^p (1 - m_i \mathcal{B}) e_t.$$

Then the two time series  $Z_t - m_1 Z_{t-1}$  and  $m_1(Z_t - m_1^{-1} Z_{t-1})$  have the same covariance function. Likewise if  $m_1$  and  $m_2 = m_1^*$  are a complex conjugate pair of roots with absolute value greater than one,  $Z_t - (m_1 + m_2) Z_{t-1} + |m_1|^2 Z_{t-2}$  has the same covariance function as  $|m_1|^2 [Z_t - (m_1^{-1} + m_2^{-1}) Z_{t-1} + |m_1|^{-2} Z_{t-2}]$ .

By repeated application of these arguments, we conclude that the original time series

$$X_t = \prod_{i=1}^p (1 - m_i \mathcal{B}) e_t$$

and the time series

$$Y_t = \left( \prod_{i=1}^L m_i \right) \mathcal{Q}(\mathcal{B}) e_t$$

have the same covariance function. Therefore,

$$\epsilon_t = \mathcal{Q}^{-1}(\mathcal{B}) X_t$$

and

$$\left( \prod_{i=1}^L m_i \right) e_t = \mathcal{Q}^{-1}(\mathcal{B}) Y_t$$

have the same covariance function. ▲

As an example of this theorem, the moving average

$$\begin{aligned} X_t &= e_t - 3.2e_{t-1} - 3.2e_{t-2} \\ &= (1 - 4.0\mathcal{B})(1 + 0.8\mathcal{B})e_t, \end{aligned}$$

where the  $e_t$  are uncorrelated  $(0, \sigma^2)$ , also has the representation

$$\begin{aligned} X_t &= \epsilon_t + 0.55\epsilon_{t-1} - 0.20\epsilon_{t-2} \\ &= (1 - 0.25\mathcal{B})(1 + 0.8\mathcal{B})\epsilon_t, \end{aligned}$$

where the  $\epsilon_t$  are uncorrelated  $(0, 16\sigma^2)$ . The reader may check that  $\gamma_X(0)$ ,  $\gamma_X(1)$ , and  $\gamma_X(2)$  are  $21.48\sigma^2$ ,  $7.04\sigma^2$ , and  $-3.20\sigma^2$ , respectively, for this moving average.

## 2.7. AUTOREGRESSIVE MOVING AVERAGE TIME SERIES

Having considered autoregressive and moving average processes, it is natural to investigate time series defined by the combination of low order autoregressive and moving average components. The sequence  $\{X_t : t \in (0, \pm 1, \pm 2, \dots)\}$ , defined by

$$X_t + a_1 X_{t-1} + \cdots + a_p X_{t-p} = e_t + b_1 e_{t-1} + \cdots + b_q e_{t-q}, \quad (2.7.1)$$

where  $a_p \neq 0$ ,  $b_q \neq 0$ , and  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables, is called an *autoregressive moving average time series of order  $(p, q)$* , which we shall abbreviate to autoregressive moving average  $(p, q)$ . The notation ARMA( $p, q$ ) is also commonly used.

The autoregressive moving average  $(1, 1)$ ,

$$X_t - \theta_1 X_{t-1} = e_t + b_1 e_{t-1}, \quad (2.7.2)$$

where  $|\theta_1| < 1$ , has furnished a useful approximation for some time series encountered in practice. Note that we have defined the coefficient on  $X_{t-1}$  as  $-\theta_1$  to simplify the representations to follow. If we let

$$u_t = e_t + b_1 e_{t-1}$$

and write

$$X_t = \theta_1 X_{t-1} + u_t,$$

we can express  $X_t$  of (2.7.2) as an infinite moving average of the  $u_t$ ,

$$X_t = \sum_{j=0}^{\infty} \theta_1^j u_{t-j},$$

and hence of the  $e_t$ ,

$$\begin{aligned} X_t &= \sum_{j=0}^{\infty} \theta_1^j e_{t-j} + b_1 \sum_{j=0}^{\infty} \theta_1^j e_{t-j-1} \\ &= e_t + (\theta_1 + b_1) \sum_{j=1}^{\infty} \theta_1^{j-1} e_{t-j}, \end{aligned} \quad (2.7.3)$$

where the random variables are defined by Theorem 2.2.1.

If  $|b_1| < 1$ , we can also express  $e_t$  as an infinite moving average of the  $u_t$ ,

$$e_t = \sum_{j=0}^{\infty} (-b_1)^j u_{t-j},$$

and hence of the  $X_t$ ,

$$e_t = X_t - (\theta_1 + b_1) \sum_{j=1}^{\infty} (-b_1)^{j-1} X_{t-j}, \quad (2.7.4)$$

If we let  $Z_t$  denote the first order autoregressive process

$$Z_t = \sum_{j=0}^{\infty} \theta_1^j e_{t-j},$$

then

$$X_t = Z_t + b_1 Z_{t-1}. \quad (2.7.5)$$

That is, the time series can be expressed as a first order moving average of a first order autoregressive time series. Using this representation, we express the autocovariance function of  $X_t$  in terms of that of  $Z_t$ ,

$$\gamma_X(h) = (1 + b_1^2) \gamma_Z(h) + b_1 \gamma_Z(h-1) + b_1 \gamma_Z(h+1),$$

to obtain

$$\gamma_X(h) = \left[ \frac{1+b_1^2}{1-\theta_1^2} \theta_1^{|h|} + \frac{b_1}{1-\theta_1^2} (\theta_1^{|h-1|} + \theta_1^{|h+1|}) \right] \sigma^2. \quad (2.7.6)$$

Hence,

$$\gamma_X(h) = \begin{cases} \frac{1+b_1^2+2b_1\theta_1}{1-\theta_1^2} \sigma^2, & h=0, \\ \frac{(1+b_1\theta_1)(\theta_1+b_1)}{1-\theta_1^2} \theta_1^{h-1} \sigma^2, & h=1,2,\dots, \end{cases} \quad (2.7.7)$$

and

$$\rho_X(h) = \begin{cases} 1, & h=0, \\ \frac{(1+b_1\theta_1)(\theta_1+b_1)}{1+b_1^2+2b_1\theta_1} \theta_1^{h-1}, & h=1,2,\dots. \end{cases}$$

Thus, for  $h \geq 1$ , the autocorrelation function of the autoregressive moving average  $(1, 1)$  has the same appearance as that of a first order autoregressive time series in that it is declining at a geometric rate where the rate is  $\theta_1$ .

From equation (2.7.7) we see that  $\gamma_X(h)$  is zero for  $h \geq 1$  if  $b_1 = -\theta_1$ . The autoregressive moving average  $(1, 1)$  process reduces to a sequence of uncorrelated random variables in such a case, which is also clear from the representation in (2.7.3).

To avoid this kind of degeneracy, the autoregressive moving average of order  $(p, q)$  is often defined with the condition that none of the roots of

$$m^p + a_1 m^{p-1} + \cdots + a_p = 0 \quad (2.7.8)$$

are roots of

$$m^q + b_1 m^{q-1} + \cdots + b_q = 0. \quad (2.7.9)$$

The autoregressive and moving average representations of (2.7.3) and (2.7.4) generalize to higher order processes in the obvious manner. We state the generalizations as theorems.

**Theorem 2.7.1.** Let the stationary time series  $X_t$  satisfy

$$\sum_{j=0}^p a_j X_{t-j} = \sum_{i=0}^q b_i e_{t-i}, \quad (2.7.10)$$

where  $a_0 = b_0 = 1$ , the roots of (2.7.8) are less than one in absolute value, and  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables. Then  $X_t$  has the

representation

$$X_t = \sum_{j=0}^{\infty} \upsilon_j e_{t-j}, \quad (2.7.11)$$

where  $\upsilon_0 = 1$ ,  $\upsilon_j = 0$  if  $j < 0$ , and

$$\upsilon_j = b_j - \sum_{i=1}^p a_i \upsilon_{j-i}$$

with  $b_j = 0$  for  $j > q$ .

**Proof.** Writing

$$\sum_{j=0}^p a_j X_{t-j} = u_t,$$

we obtain

$$\begin{aligned} X_t &= \sum_{j=0}^{\infty} w_j u_{t-j} = \sum_{j=0}^{\infty} w_j \left( \sum_{i=0}^q b_i e_{t-j-i} \right) \\ &\stackrel{(say)}{=} \sum_{r=0}^{\infty} v_r e_{t-r}, \end{aligned}$$

where  $u_t = \sum_{i=0}^q b_i e_{t-i}$  and the  $w_j$  are defined in Theorem 2.6.1. Since the  $w_j$  are absolutely summable, this representation is well defined as a limit in mean square. We then write

$$\sum_{i=0}^q b_i e_{t-i} = \sum_{j=0}^p a_j X_{t-j} = \sum_{j=0}^p a_j \left( \sum_{r=0}^{\infty} v_r e_{t-j-r} \right)$$

and obtain the result by equating coefficients of  $e_{t-i}$ ,  $i = 0, 1, \dots$ . ▲

The first few  $v_r$  of Theorem 2.7.1 can be used to compute the first few covariances of the process. If we multiply the defining equation of Theorem 2.7.1 by  $X_{t-r}$ , for  $r = 0, 1, \dots, p$  and take expectations, we have

$$\sum_{j=0}^p a_j \gamma_X(r-j) = \sum_{i=r}^q b_i v_{i-r} \sigma^2,$$

where it is understood that the summation on the right of the equality is zero for  $r > q$ .

**Theorem 2.7.2.** Let the stationary time series  $X_t$  satisfy

$$\sum_{j=0}^p a_j X_{t-j} = \sum_{i=0}^q b_i e_{t-i},$$

where  $a_0 = b_0 = 1$ , the roots of (2.7.8) and of (2.7.9) are less than one in absolute value, and  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables. Then  $X_t$  has the representation

$$e_t = \sum_{j=0}^{\infty} d_j X_{t-j}, \quad (2.7.12)$$

where  $d_0 = 1$ ,  $d_j = 0$  for  $j < 0$ , and

$$d_j = - \sum_{i=1}^q b_i d_{j-i} + a_j,$$

with  $a_j = 0$  for  $j > p$ .

**Proof.** Reserved for the reader. ▲

The backward shift operator is often used to give a compact representation for the autoregressive moving average. Let

$$A(m) = 1 + a_1 m + \cdots + a_p m^p$$

and

$$B(m) = 1 + b_1 m + \cdots + b_q m^q$$

be the polynomials associated with the autoregressive moving average (2.7.10). Then we can write

$$A(\mathcal{B})X_t = B(\mathcal{B})e_t, \quad (2.7.13)$$

or, if the roots of  $A(m) = 0$  are greater than one in absolute value,

$$X_t = A^{-1}(\mathcal{B})B(\mathcal{B})e_t, \quad (2.7.14)$$

or, if the roots of  $B(m) = 0$  are also greater than one in absolute value,

$$e_t = B^{-1}(\mathcal{B})A(\mathcal{B})X_t. \quad (2.7.15)$$

The representation (2.7.14) corresponds to (2.7.11), and the representation (2.7.15) corresponds to (2.7.12). Also see Exercise 2.26.

The equation (2.7.13) defines an autoregressive moving average constructed with  $e_t$  variables. By (2.7.14), we could apply the same moving average operator to any sequence. Thus, if  $C(m)$  and  $D(m)$  are polynomials whose roots are greater

than one in absolute value, and if  $X_t$  is the time series of (2.7.13), we might define the new time series  $Y_t$  by

$$C(\mathcal{B})Y_t = D(\mathcal{B})X_t$$

or

$$\begin{aligned} Y_t &= C^{-1}(\mathcal{B})D(\mathcal{B})X_t \\ &= C^{-1}(\mathcal{B})D(\mathcal{B})A^{-1}(\mathcal{B})B(\mathcal{B})e_t \\ &= G^{-1}(\mathcal{B})H(\mathcal{B})e_t, \end{aligned} \quad (2.7.16)$$

where  $G(\mathcal{B}) = C(\mathcal{B})A(\mathcal{B})$  and  $H(\mathcal{B}) = D(\mathcal{B})B(\mathcal{B})$ . If there are any common factors in  $G(\mathcal{B})$  and  $H(\mathcal{B})$ , they can be removed to obtain the standard representation. On the basis of this discussion, we might say that an autoregressive moving average of an autoregressive moving average is an autoregressive moving average.

## 2.8. VECTOR PROCESSES

From an operational standpoint the generalizations of moving average and autoregressive representations to the vector case are obtained by substituting matrix and vector expressions for the scalar expressions of the preceding sections.

Consider a first order moving average process of dimension two:

$$\begin{aligned} X_{1t} &= b_{11}e_{1,t-1} + b_{12}e_{2,t-1} + e_{1t}, \\ (2.8.1) \end{aligned}$$

$$X_{2t} = b_{21}e_{1,t-1} + b_{22}e_{2,t-1} + e_{2t}.$$

The system may be written in matrix notation as

$$\mathbf{X}_t = \mathbf{Be}_{t-1} + \mathbf{e}_t,$$

where  $\{\mathbf{e}_t\}$  is a sequence of uncorrelated  $(\mathbf{0}, \Sigma)$  vector random variables,

$$\Sigma = E\{\mathbf{e}_t \mathbf{e}'_t\} = \begin{pmatrix} \sigma_{11} & \sigma_{12} \\ \sigma_{21} & \sigma_{22} \end{pmatrix},$$

and we assume  $\mathbf{B}$  contains at least one nonzero element. Then the autocovariance function of  $\mathbf{X}_t$  is

$$\Gamma(h) = E\{\mathbf{X}_t \mathbf{X}'_{t+h}\} = E\{(\mathbf{Be}_{t-1} + \mathbf{e}_t)(\mathbf{e}'_{t+h-1} \mathbf{B}' + \mathbf{e}'_{t+h})\}.$$

and it follows that

$$\Gamma(h) = \begin{cases} \mathbf{B}\Sigma\mathbf{B}' + \Sigma, & h = 0, \\ \Sigma\mathbf{B}', & h = 1, \\ \mathbf{B}\Sigma, & h = -1, \\ \mathbf{0}, & \text{otherwise.} \end{cases}$$

Note that  $\Gamma(h) = \Gamma'(-h)$ , as we would expect from Lemma 1.7.1.

The  $q$ th order moving average time series of dimension  $k$  is defined by

$$\mathbf{X}_t = \sum_{j=0}^q \mathbf{B}_j \mathbf{e}_{t-j},$$

where the  $\mathbf{B}_j$  are  $k \times k$  matrices with  $\mathbf{B}_0 = \mathbf{I}$  and at least one element of  $\mathbf{B}_q$  not zero, and the  $\mathbf{e}_t$  are uncorrelated  $(\mathbf{0}, \Sigma)$  random variables. Then

$$\Gamma(h) = \begin{cases} \sum_{j=0}^{q-h} \mathbf{B}_j \Sigma \mathbf{B}'_{j+h}, & 0 \leq h \leq q, \\ \sum_{j=0}^{q+h} \mathbf{B}_{j-h} \Sigma \mathbf{B}'_j, & -q \leq h \leq 0, \\ \mathbf{0}, & |h| > q. \end{cases} \quad (2.8.2)$$

As with the scalar process, the autocovariance matrix is zero for  $|h| > q$ .

In investigating the properties of infinite moving average and autoregressive processes in the scalar case, we made extensive use of absolutely summable sequences. We now extend this concept to matrix sequences.

**Definition 2.8.1.** Let  $\{\mathbf{G}_j\}_{j=1}^\infty$  be a sequence of  $k \times r$  matrices where the elements of  $\mathbf{G}_j$  are  $g_{im}(j)$ ,  $i = 1, 2, \dots, k$ ,  $m = 1, 2, \dots, r$ . If each of the  $kr$  sequences  $\{g_{im}(j)\}_{j=1}^\infty$  is absolutely summable, we say that the sequence  $\{\mathbf{G}_j\}_{j=1}^\infty$  is absolutely summable.

The infinite moving average

$$\mathbf{X}_t = \sum_{j=0}^\infty \mathbf{B}_j \mathbf{e}_{t-j},$$

where the  $\mathbf{e}_t$  are uncorrelated  $(\mathbf{0}, \Sigma)$  random variables, will be well defined as an almost sure limit if  $\{\mathbf{B}_j\}$  is absolutely summable. Thus, for example, if  $\mathbf{X}_t$  is a vector stationary time series of dimension  $k$  with zero mean and absolutely summable covariance function and  $\{\mathbf{W}_j\}_{j=-\infty}^\infty$  is an absolutely summable sequence of  $k \times k$  matrices, then

$$\mathbf{Y}_t = \sum_{j=-\infty}^\infty \mathbf{W}_j \mathbf{X}_{t-j}$$

is well defined as an almost sure limit and

$$\Gamma_Y(h) = \sum_{j=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \mathbf{W}_j \Gamma_X(h-s+j) \mathbf{W}'_s.$$

The vector autoregressive process of order  $p$  and dimension  $k$  is defined by

$$\sum_{j=0}^p \mathbf{A}_j \mathbf{X}_{t-j} = \mathbf{e}_t, \quad (2.8.3)$$

where the  $\mathbf{e}_t$  are uncorrelated  $k$ -dimensional  $(\mathbf{0}, \Sigma)$  random variables and the  $\mathbf{A}_j$  are  $k \times k$  matrices with  $\mathbf{A}_0 = \mathbf{I}$  and  $\mathbf{A}_p \neq \mathbf{0}$ .

Any autoregressive process can be expressed as a first order vector process. Let  $\mathbf{Y}_t = (\mathbf{X}'_t, \mathbf{X}'_{t-1}, \dots, \mathbf{X}'_{t-p+1})'$  and  $\mathbf{e}_t = (\mathbf{e}'_t, \mathbf{0}, \dots, \mathbf{0})'$ , where the  $\mathbf{X}_t$  satisfy (2.8.3). Then we can write

$$\mathbf{Y}_t + \mathbf{A}\mathbf{Y}_{t-1} = \mathbf{e}_t,$$

where

$$\mathbf{A} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{A}_2 & \cdots & \mathbf{A}_{p-1} & \mathbf{A}_p \\ -\mathbf{I} & \mathbf{0} & \cdots & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & -\mathbf{I} & \cdots & \mathbf{0} & \mathbf{0} \\ \vdots & \vdots & & \vdots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & -\mathbf{I} & \mathbf{0} \end{pmatrix}.$$

This representation for the scalar autoregressive process was used in the proof of Theorem 2.6.1.

The process (2.8.3) can be expressed as an infinite one-sided moving average of the  $\mathbf{e}_t$ , when the roots of the characteristic equation are less than one in absolute value.

**Theorem 2.8.1.** Let the stationary time series  $\{\mathbf{X}_t : t \in (0, \pm 1, \pm 2, \dots)\}$  satisfy (2.8.3), and let the roots of the determinantal equation

$$|\mathbf{A}_0 m^p + \mathbf{A}_1 m^{p-1} + \cdots + \mathbf{A}_p| = 0 \quad (2.8.4)$$

be less than one in absolute value. Then  $\mathbf{X}_t$  can be expressed as an infinite moving average of the  $\mathbf{e}_t$ , say

$$\mathbf{X}_t = \sum_{j=0}^{\infty} \mathbf{K}_j \mathbf{e}_{t-j},$$

where  $\mathbf{K}_0 = \mathbf{I}$ ,  $\mathbf{K}_1 = -\mathbf{A}_1$ ,

$$\begin{aligned}\mathbf{K}_2 &= -\mathbf{A}_1\mathbf{K}_1 - \mathbf{A}_2, \\ &\vdots \\ \mathbf{K}_s &= -\sum_{j=1}^p \mathbf{A}_j \mathbf{K}_{s-j}, \quad s = p, p+1, \dots\end{aligned}$$

**Proof.** Essentially the same argument as used at the beginning of Section 2.5 may be used to demonstrate that  $\sum_{j=0}^{\infty} \mathbf{K}_j \mathbf{e}_{t-j}$  is the solution of the difference equation. Since the roots of the determinantal equation are less than one in absolute value, the sequence of matrices  $\{\mathbf{K}_j\}$  is absolutely summable and the random variables are well defined as almost sure limits. ▲

Given Theorem 2.8.1, we can construct the multivariate Yule–Walker equations for the autocovariances of the process. Multiplying (2.8.3) on the right by  $\mathbf{X}'_{t-s}$ ,  $s \geq 0$  and taking expectations, we obtain

$$\begin{aligned}\sum_{j=0}^p \mathbf{A}_j \Gamma(j) &= \Sigma, \quad s = 0, \\ (2.8.5) \quad \sum_{j=0}^p \mathbf{A}_j \Gamma(j-s) &= \mathbf{0}, \quad s = 1, 2, \dots,\end{aligned}$$

where we have used

$$E\{\mathbf{e}_t \mathbf{X}'_{t-s}\} = \begin{cases} \Sigma, & s = 0, \\ \mathbf{0} & \text{otherwise.} \end{cases}$$

These equations are of the same form as those given in equation (2.5.6) and Corollary 2.6.1.1, but they contain a larger number of elements.

**Theorem 2.8.2.** Let the time series  $\mathbf{X}_t$  be defined by

$$\mathbf{X}_t = \sum_{j=0}^q \mathbf{B}_j \mathbf{e}_{t-j},$$

where the  $\mathbf{e}_t$  are uncorrelated  $k$ -dimensional vector random variables with zero mean and covariance matrix  $\Sigma$ , the  $\mathbf{B}_j$  are  $k \times k$  matrices with  $\mathbf{B}_0 = \mathbf{I}$ ,  $\mathbf{B}_q \neq 0$ , and the  $kq$  roots of the determinantal equation

$$|\mathbf{B}_0 \mathbf{m}^q + \mathbf{B}_1 \mathbf{m}^{q-1} + \cdots + \mathbf{B}_q| = 0 \quad (2.8.6)$$

are less than one in absolute value. Then  $\mathbf{X}_t$  can be expressed as an infinite

autoregressive process

$$\mathbf{e}_t = \sum_{j=0}^{\infty} \mathbf{C}_j \mathbf{X}_{t-j},$$

where the  $\mathbf{C}_j$  satisfy  $\mathbf{C}_0 = \mathbf{I}$ ,  $\mathbf{C}_1 = -\mathbf{B}_1$ ,

$$\mathbf{C}_2 = -\mathbf{B}_1 \mathbf{C}_1 - \mathbf{B}_2,$$

⋮

$$\mathbf{C}_s = -\sum_{j=1}^q \mathbf{B}_j \mathbf{C}_{s-j}, \quad s = q, q+1, \dots.$$

**Proof.** Reserved for the reader. ▲

It can also be demonstrated that an autoregressive moving average of the form

$$\sum_{j=0}^p \mathbf{A}_j \mathbf{X}_{t-j} = \sum_{j=0}^q \mathbf{B}_j \mathbf{e}_{t-j},$$

where the roots of (2.8.4) and of (2.8.6) are less than one in absolute value, can be given either an infinite autoregressive or an infinite moving average representation with matrix weights of the same form as the scalar weights of Theorems 2.7.1 and 2.7.2.

## 2.9. PREDICTION

One of the important problems in time series analysis is the following: Given  $n$  observations on a realization, predict the  $(n+s)$ th observation in the realization, where  $s$  is a positive integer. The prediction is sometimes called the *forecast* of the  $(n+s)$ th observation. Because of the functional nature of a realization, prediction is also called *extrapolation*.

In some areas of statistics the term *predictor* is applied to a function of observations used to approximate an unknown *random* quantity, and the term *estimator* is applied to a function of observations used to approximate an unknown *fixed* quantity. Some authors use the term estimator to describe the function in both situations. In time series, the term *filter* is often used for the function of observations used to approximate an unknown value of the random process at time  $t$ . If only values of the process prior to  $t$  are used in the approximating function for the time series, the function is called a predictor. We adopt this use and apply the term predictor to functions used to approximate a future term of the realization. We will use estimator as a generic term for a function of observations used to approximate either a fixed or a random quantity.

In order to discuss alternative predictors, it is necessary to have a criterion by which the performance of a predictor is measured. The usual criterion, and the one that we adopt, is the mean square error of the predictor. For example, if  $\hat{Y}_{n+s}(Y_1, \dots, Y_n)$  is the predictor of  $Y_{n+s}$  based on the  $n$  observations  $Y_1, Y_2, \dots, Y_n$ , then the mean square error (MSE) of the predictor is

$$\text{MSE}\{\hat{Y}_{n+s}(Y_1, \dots, Y_n)\} = E\{[Y_{n+s} - \hat{Y}_{n+s}(Y_1, \dots, Y_n)]^2\}. \quad (2.9.1)$$

Generally, the problem of determining optimal predictors requires that the class of predictors be restricted. We shall investigate *linear* predictors.

The best linear predictor for a stationary time series with known covariance function and known mean is easily obtained.

**Theorem 2.9.1.** Given  $n$  observations on a realization of a time series with zero mean and known covariance matrix, the minimum mean square error linear predictor of the  $(n+s)$ th,  $s = 1, 2, \dots$ , observation is

$$\hat{Y}_{n+s}(Y_1, \dots, Y_n) = \mathbf{y}' \mathbf{b}_{ns}, \quad (2.9.2)$$

where  $\mathbf{y}' = (Y_1, Y_2, \dots, Y_n)$ ,  $\mathbf{V}_{nn} = E\{\mathbf{y}\mathbf{y}'\}$ ,

$$\begin{aligned} \mathbf{b}_{ns} &= \mathbf{V}_{nn}^+ \mathbf{V}_{ns}, \\ \mathbf{V}_{ns}' &= E\{Y_{n+s} \mathbf{y}'\}, \end{aligned}$$

and  $\mathbf{V}_{nn}^+$  is a generalized inverse of  $\mathbf{V}_{nn}$ . [See Rao (1973) for a discussion of generalized inverses.] The mean of the prediction error is zero, and the variance of the prediction error is

$$\tau_{ns}^2 = V\{Y_{n+s}\} - \mathbf{b}_{ns}' \mathbf{V}_{ns}. \quad (2.9.3)$$

**Proof.** To find the minimum mean square error linear predictor, we minimize

$$\begin{aligned} E\{[Y_{n+s} - \hat{Y}_{n+s}(Y_1, \dots, Y_n)]^2\} &= E\{[Y_{n+s} - \mathbf{y}' \mathbf{b}]^2\} \\ &= (1, -\mathbf{b}') \begin{pmatrix} V\{Y_{n+s}\} & \mathbf{V}_{ns}' \\ \mathbf{V}_{ns} & \mathbf{V}_{nn} \end{pmatrix} \begin{pmatrix} 1 \\ -\mathbf{b} \end{pmatrix} \end{aligned}$$

with respect to the vector of coefficients  $\mathbf{b}$ . Differentiating the quadratic form with respect to  $\mathbf{b}$  and equating the derivative to zero, we find that a  $\mathbf{b}$  that minimizes the expectation is  $\mathbf{b}_{ns} = \mathbf{V}_{nn}^+ \mathbf{V}_{ns}$ . While  $\mathbf{b}_{ns}$  is not necessarily unique, the predictor  $\mathbf{y}' \mathbf{b}_{ns}$  is unique. See Exercise 2.42. The mean of the prediction error is zero because the mean of the process is zero and

$$E\{Y_{n+s} - \hat{Y}_{n+s}(Y_1, \dots, Y_n)\} = E\{Y_{n+s} - \mathbf{y}' \mathbf{b}_{ns}\} = 0.$$

The variance of the prediction error follows immediately from least squares regression theory. ▲

To understand why the generalized inverse was required in Theorem 2.9.1, consider the time series

$$X_t = e_1 \cos \pi t, \quad (2.9.4)$$

where  $e_1$  is normally distributed with zero mean and variance  $\sigma^2$ . This time series is perfectly periodic with period two and covariance function  $\gamma_X(h) = \sigma^2 \cos \pi h$ . The matrix  $V_{nn}$  for a sample of three observations is

$$V_{33} = \begin{pmatrix} 1 & -1 & 1 \\ -1 & 1 & -1 \\ 1 & -1 & 1 \end{pmatrix} \sigma^2 \quad (2.9.5)$$

which is easily seen to be singular. Thus there are many possible linear combinations of past values that can be used to predict future values. For example, one can predict  $X_3$  by  $-X_{t-1}$  or by  $X_{t-2}$  or by  $-\frac{1}{2}X_{t-1} + \frac{1}{2}X_{t-2}$ .

The time series of equation (2.9.4) is a member of the class of time series defined by

$$X_t = e_1 \cos \lambda t + e_2 \sin \lambda t, \quad (2.9.6)$$

where  $e_i$ ,  $i = 1, 2$ , are independent  $(0, \sigma^2)$  random variables and  $\lambda \in (0, \pi]$ . Given  $\lambda$  and two observations, it is possible to predict perfectly all future observations in a given realization. That is, for  $n \geq 2$ ,

$$\begin{pmatrix} e_1 \\ e_2 \end{pmatrix} = \begin{bmatrix} \sum_{t=1}^n \cos^2 \lambda t & \sum_{t=1}^n \cos \lambda t \sin \lambda t \\ \sum_{t=1}^n \cos \lambda t \sin \lambda t & \sum_{t=1}^n \sin^2 \lambda t \end{bmatrix}^{-1} \begin{bmatrix} \sum_{t=1}^n X_t \cos \lambda t \\ \sum_{t=1}^n X_t \sin \lambda t \end{bmatrix},$$

and all future observations can be predicted by substituting these  $e$ 's into the defining equation (2.9.6).

This example illustrates that in predicting future observations of a realization from past observations of that realization, the element  $\omega \in \Omega$  is fixed. In the current example,  $\omega \in \Omega$  is a two-dimensional vector and can be perfectly identified for a particular realization by a few observations on that realization. In the more common situation,  $\omega \in \Omega$  is infinite-dimensional, and we can never determine it perfectly from a finite number of observations on the realization.

Time series for which it is possible to use the observations  $Y_t, Y_{t-1}, \dots$  of a realization to predict future observations with zero mean square error are called *deterministic*. They are also sometimes called *singular*, a term that is particularly meaningful when one looks at the covariance matrix (2.9.5). Time series that have a nonzero lower bound on the prediction error are called *regular* or *nondeterministic*.

**Definition 2.9.1.** A time series is *nonsingular (regular, nondeterministic)* if the sequence of mean square errors of one-period prediction,  $\tau_{n1}^2$ , is bounded away from zero. A time series is *singular (deterministic)* if

$$\lim_{n \rightarrow \infty} \tau_{n1}^2 = 0.$$

The vector of coefficients in Theorem 2.9.1 for one-period predictions can be computed recursively for nondeterministic time series. The recursive procedure is known as the Durbin–Levinson algorithm. See Durbin (1960) and Morettin (1984).

**Theorem 2.9.2.** Let  $Y_t$  be a zero mean, covariance stationary, nondeterministic time series defined on the integers. Then the coefficients defining the predictor of  $Y_{n+1}$  given in (2.9.2) of Theorem 2.9.1 satisfy

$$\begin{pmatrix} b_{n,1,2} \\ b_{n,1,3} \\ \vdots \\ b_{n,1,n} \end{pmatrix} = \begin{pmatrix} b_{n-1,1,1} \\ b_{n-1,1,2} \\ \vdots \\ b_{n-1,1,n-1} \end{pmatrix} - b_{n,1,1} \begin{pmatrix} b_{n-1,1,n-1} \\ b_{n-1,1,n-2} \\ \vdots \\ b_{n-1,1,1} \end{pmatrix}, \quad (2.9.7)$$

where  $b_{0,1,1} = 0$ ,  $\tau_{11}^2 = \gamma(0)$ ,  $\gamma(h) = \text{Cov}\{Y_t, Y_{t+h}\}$ ,

$$b_{n,1,1} = \left[ \gamma(n) - \sum_{j=1}^{n-1} b_{n-1,1,j} \gamma(n-j) \right] \tau_{n-1,1}^{-2}, \quad (2.9.8)$$

and

$$\tau_{n1}^2 = \tau_{n-1,1}^2 (1 - b_{n,1,1}^2), \quad (2.9.9)$$

for  $n = 1, 2, \dots$ .

**Proof.** It follows from the symmetry of the covariance function of a stationary process that the coefficients of the predictor of  $Y_n$  based upon  $(Y_1, \dots, Y_{n-1})$  are the coefficients of the predictor of  $Y_1$  based upon  $(Y_n, \dots, Y_2)$ . That is,

$$\hat{Y}_n(Y_1, \dots, Y_{n-1}) = \sum_{j=1}^{n-1} b_{n-1,1,j} Y_j$$

and

$$\hat{Y}_1(Y_n, \dots, Y_2) = \sum_{j=1}^{n-1} b_{n-1,1,j} Y_{n+1-j},$$

where  $\hat{Y}_1(Y_1, \dots, Y_n)$  is the predictor of  $Y_1$  based on  $(Y_2, \dots, Y_n)$  and

$$\mathbf{b}_{n-1,1} = (b_{n-1,1,1}, b_{n-1,1,2}, \dots, b_{n-1,1,n-1})'$$

is the vector of (2.9.2) for  $s = 1$ . The coefficient vector  $\mathbf{b}_{n-1,1}$  is unique because the covariance matrix of a nondeterministic time series is positive definite.

The best predictor of  $Y_{n+1}$  based on  $(Y_1, \dots, Y_n)$  can be constructed from the regression of  $Y_{n+1}$  on the  $n$ -dimensional vector  $[Y_1 - \hat{Y}_1(Y_1, \dots, Y_n), Y_2, \dots, Y_n]$ . By stationarity, the coefficients for the regression of  $Y_{n+1}$  on  $(Y_2, \dots, Y_n)$  are the same as the coefficients for the regression of  $Y_n$  on  $(Y_1, \dots, Y_{n-1})$ . Because  $\hat{Y}_1(Y_1, \dots, Y_n)$  is the best linear predictor,  $Y_1 - \hat{Y}_1(Y_1, \dots, Y_n)$  is uncorrelated with  $(Y_2, \dots, Y_n)$ , and the regression coefficient for  $Y_1 - \hat{Y}_1(Y_1, \dots, Y_n)$  is

$$b_{n,1,1} = \tau_{n-1,1}^{-2} [\gamma(n) - \sum_{j=1}^{n-1} b_{n-1,1,j} \gamma(n-j)].$$

Therefore,

$$\begin{aligned}\hat{Y}_{n+1}(Y_1, \dots, Y_n) &= \sum_{j=1}^{n-1} b_{n-1,1,j} Y_{1+j} + b_{n,1,1} \left( Y_1 - \sum_{j=1}^{n-1} b_{n-1,1,j} Y_{n+1-j} \right) \\ &= b_{n,1,1} Y_1 + \sum_{j=1}^{n-1} (b_{n-1,1,j} - b_{n,1,1} b_{n-1,1,n-j}) Y_{1+j},\end{aligned}$$

and we have the result (2.9.7). The error mean square for the prediction of  $Y_{n+1}$  based on  $Y_1, Y_2, \dots, Y_n$  is equal to the prediction variance for the prediction of  $Y_{n+1}$  based on  $Y_2, Y_3, \dots, Y_n$  reduced by the sum of squares due to  $Y_1$  after adjusting for  $Y_2, Y_3, \dots, Y_n$ . The variance of the prediction error for  $\hat{Y}_{n+1}(Y_2, \dots, Y_n)$  is  $\tau_{n-1,1}^2$ , and  $b_{n,1,1}^2$  is the squared correlation between  $Y_{n+1} - \hat{Y}_{n+1}(Y_2, \dots, Y_n)$  and  $Y_1 - \hat{Y}_1(Y_1, \dots, Y_n)$ , giving the result (2.9.9).  $\blacktriangle$

Theorem 2.9.1 gives a quite general solution to the problem of linear prediction, but the computation of  $n \times n$  generalized inverses can become inconvenient if one has a large number of predictions to make. Theorem 2.9.2 presents a recursive method of computing the coefficients of the lagged  $Y$ 's required for prediction. However, in general, these computations also become sizable as  $n$  increases. Fortunately, the forecast equations for autoregressive processes are very simple, and those for invertible moving average processes are fairly simple.

We begin with the first order autoregressive process, assuming the parameter to be known. Let the time series  $Y_t$  be defined by

$$Y_t = \rho Y_{t-1} + e_t, \quad |\rho| < 1,$$

where the  $e_t$  are independent  $(0, \sigma^2)$  random variables. We wish to construct a predictor of  $Y_{n+1} = \rho Y_n + e_{n+1}$ . Clearly, the "best predictor" of  $\rho Y_n$  is  $\rho Y_n$ .

Therefore, we need only predict  $e_{n+1}$ . But, by assumption,  $e_{n+1}$  is independent of  $Y_n, Y_{n-1}, \dots$ . Hence, the best predictor of  $e_{n+1}$  is zero, since predicting  $e_{n+1}$  is equivalent to predicting a random selection from a distribution with zero mean and finite variance. Therefore, we choose as a predictor of  $Y_{n+1}$  at time  $n$

$$\hat{Y}_{n+1}(Y_1, \dots, Y_n) = \rho Y_n. \quad (2.9.10)$$

Notice that we constructed this predictor by finding the expected value of  $Y_{n+1}$  given  $Y_n, Y_{n-1}, \dots, Y_1$ . Because of the nature of the time series, knowledge of  $Y_{n-1}, Y_{n-2}, \dots, Y_1$  added no information beyond that contained in knowledge of  $Y_n$ . It is clear that the mean square error of this predictor is

$$E\{[Y_{n+1} - \hat{Y}_{n+1}(Y_1, \dots, Y_n)]^2\} = E\{e_{n+1}^2\} = \sigma^2.$$

To predict more than one period into the future, we recall the representation

$$Y_{n+s} = \rho^s Y_n + \sum_{j=1}^s \rho^{s-j} e_{n+j}, \quad s = 1, 2, \dots \quad (2.9.11)$$

It follows that, for independent  $e_i$ , the conditional expectation of  $Y_{n+s}$  given  $Y_1, Y_2, \dots, Y_n$  is

$$E\{Y_{n+s} | Y_1, \dots, Y_n\} = \rho^s Y_n, \quad (2.9.12)$$

and this is the best predictor for  $Y_{n+s}$ .

These ideas generalize immediately to higher order processes. Let the  $p$ th order autoregressive process be defined by

$$Y_t = \sum_{j=1}^p \theta_j Y_{t-j} + e_t,$$

where the roots of

$$m^p - \sum_{j=1}^p \theta_j m^{p-j} = 0$$

are less than one in absolute value and the  $e_i$  are uncorrelated  $(0, \sigma^2)$  random variables. On the basis of the arguments for the first order process we choose as the one-period-ahead linear predictor for the stationary  $p$ th order autoregressive process

$$\hat{Y}_{n+1}(Y_1, \dots, Y_n) = \sum_{j=1}^p \theta_j Y_{n+1-j}. \quad (2.9.13)$$

It is clear that this predictor minimizes (2.9.1), since the prediction error is  $e_{n+1}$  and this is uncorrelated with  $Y_n$  and all previous  $Y_i$ 's. If the  $e_i$  are independent random variables, the predictor is the expected value of  $Y_{n+1}$  conditional on  $Y_n, Y_{n-1}, \dots, Y_1$ . If the  $e_i$  are only uncorrelated, we cannot make this statement. (See

Exercise 2.16). To obtain the best two-period linear prediction, we note that

$$\begin{aligned} Y_{n+2} &= \theta_1 \left[ \sum_{j=1}^p \theta_j Y_{n-j+1} + e_{n+1} \right] + \sum_{j=2}^p \theta_j Y_{n-j+2} + e_{n+2} \\ &= \theta_1 [\hat{Y}_{n+1}(Y_1, \dots, Y_n) + e_{n+1}] + \sum_{j=2}^p \theta_j Y_{n-j+2} + e_{n+2} \quad (2.9.14) \\ &= \sum_{j=1}^p (\theta_1 \theta_j + \theta_{j+1}) Y_{n-j+1} + \theta_1 e_{n+1} + e_{n+2}, \end{aligned}$$

where it is understood that  $\theta_{p+i} = 0$  for  $i = 1, 2, \dots$  and, to simplify the subscripting, we have taken  $p \geq 2$ . Since  $e_{n+2}$  and  $e_{n+1}$  are uncorrelated with  $(Y_1, Y_2, \dots, Y_n)$ , the two-period predictor can be constructed as

$$\begin{aligned} \hat{Y}_{n+2}(Y_1, \dots, Y_n) &= \theta_1 \hat{Y}_{n+1}(Y_1, \dots, Y_n) + \sum_{j=2}^p \theta_j Y_{n-j+2} \\ &= \sum_{j=1}^p (\theta_1 \theta_j + \theta_{j+1}) Y_{n-j+1}. \end{aligned}$$

In general, we can build predictors for  $s$  periods ahead by substituting the predictors for earlier periods into (2.9.13). Hence, for  $p \geq 2$ ,

$$\begin{aligned} \hat{Y}_{n+s}(Y_1, \dots, Y_n) &= \sum_{j=1}^{s-1} \theta_j \hat{Y}_{n-j+s}(Y_1, \dots, Y_n) + \sum_{j=s}^p \theta_j Y_{n-j+s}, \quad s = 2, 3, \dots, p, \\ &= \sum_{j=1}^p \theta_j \hat{Y}_{n-j+s}(Y_1, \dots, Y_n), \quad s = p+1, p+2, \dots \end{aligned} \quad (2.9.15)$$

We now turn to prediction for the  $q$ th order moving average time series

$$Y_t = \sum_{j=1}^q \beta_j e_{t-j} + e_t. \quad (2.9.16)$$

We assume  $\beta_j$ ,  $j = 1, 2, \dots, q$ , are known. The  $p$ th order autoregressive process expresses  $Y_t$  as a linear combination of  $p$  previous  $Y$ 's and  $e_t$ , while the  $q$ th order moving average process expresses  $Y_t$  as a linear combination of  $q$  previous  $e$ 's and  $e_t$ . In both cases,  $e_t$  is uncorrelated with previous  $Y$ 's and uncorrelated with previous  $e$ 's. Suppose we knew the  $e_t$  for  $t = n, n-1, \dots, n-q+1$  for the process (2.9.16). Then, by arguments analogous to those used for the autoregressive process, the best linear predictor for  $s$  periods ahead would be

$$\hat{Y}_{n+s}(e_1, \dots, e_n) = \begin{cases} \sum_{j=s}^q \beta_j e_{n+s-j}, & 1 \leq s \leq q, \\ 0, & s > q. \end{cases} \quad (2.9.17)$$

If the  $e_t$ 's are independent,

$$\hat{Y}_{n+s}(e_1, \dots, e_n) = E\{Y_{n+s} | e_1, e_2, \dots, e_n\}.$$

Because we do not know the  $e_t$ , we must develop predictors based upon estimators of the  $e_t$ , where the estimators of the  $e_t$  are functions of the observed  $Y_t$ . We first show that the predictions of Theorem 2.9.1 can be expressed in terms of the one-period prediction errors of previous predictions when  $Y_t$  is nondeterministic. The recursive computation is based on the Gram-Schmidt orthogonalization. Brockwell and Davis (1991, p. 171) call the procedure the *innovation algorithm*, where the  $Z_t$  are the innovations. We give the proof for stationary time series, but the result holds for more general covariance functions.

**Theorem 2.9.3.** Let  $Y_t$  be a zero mean, stationary, nondeterministic time series defined on the integers. Let

$$Z_t = Y_t - \sum_{j=1}^{t-1} c_{ij} Z_j, \quad t = 2, 3, \dots, \quad (2.9.18)$$

where  $Z_1 = c_{11} Y_1$ ,  $c_{11} = 1$ ,  $c_{21} = \kappa_1^{-2} \gamma_Y(1)$ ,  $\kappa_1^2 = \gamma_Y(0)$ ,

$$c_{ii} = \kappa_i^{-2} \gamma_Y(t-1), \quad (2.9.19)$$

$$c_{ii} = \kappa_i^{-2} \left[ \gamma_Y(t-i) - \sum_{j=1}^{i-1} c_{ij} c_{ij} \kappa_j^2 \right], \quad i = 2, 3, \dots, t-1,$$

and

$$\kappa_t^2 = \gamma_Y(0) - \sum_{j=1}^{t-1} c_{jj}^2 \kappa_j^2 \quad t = 2, 3, \dots. \quad (2.9.20)$$

Then the best predictor of  $Y_{n+s}$  given  $(Y_1, \dots, Y_n)$  is

$$\hat{Y}_{n+s}(Y_1, \dots, Y_n) = \sum_{j=1}^n c_{n+s,j} Z_j, \quad (2.9.21)$$

and the variance of the prediction error is

$$\tau_{ns}^2 = \begin{cases} \kappa_{n+1}^2, & s = 1, \\ \kappa_{n+s}^2 + \sum_{j=1}^{s-1} c_{n+s,n+j}^2 \kappa_{n+j}^2, & s > 1. \end{cases} \quad (2.9.22)$$

**Proof.** Because  $Z_t$  is the residual from the regression of  $Y_t$  on  $(Z_{t-1},$

$Z_{t-2}, \dots, Z_1$ ,  $Z_t$  is uncorrelated with  $(Z_{t-1}, Z_{t-2}, \dots, Z_1)$ , and hence the  $Z$ 's are mutually uncorrelated. The covariance between  $Y_t$  and  $Z_i$  for  $t > i$  is

$$\begin{aligned}\text{Cov}(Y_t, Z_i) &= \text{Cov}\left(Y_t, Y_t - \sum_{j=1}^{t-1} c_{ij} Z_j\right) \\ &= \gamma_Y(t-i) - \text{Cov}\left(Y_t, \sum_{j=1}^{t-1} c_{ij} Z_j\right) \\ &= \gamma_Y(t-i) - \text{Cov}\left(Z_t + \sum_{r=1}^{t-1} c_{ir} Z_r, \sum_{j=1}^{t-1} c_{ij} Z_j\right) \\ &= \gamma_Y(t-i) - \sum_{j=1}^{t-1} c_{ij} c_{ij} \kappa_j^2,\end{aligned}$$

where we have used the zero correlation property of the  $Z_j$ . Dividing  $\text{Cov}(Y_t, Z_i)$  by the variance of  $Z_i$ , we obtain the expression for  $c_{ii}$ . The variance of the prediction error is the regression error mean square, and by the zero correlation property we have (2.9.20).

Because of the zero correlation property, the coefficients of  $(Z_n, \dots, Z_1)$  in the regression of  $Y_{n+s}$  on  $(Z_n, \dots, Z_1)$  are the same as the coefficients of  $(Z_n, \dots, Z_1)$  in the regression of  $Y_{n+s}$  on  $(Z_{n+s-1}, \dots, Z_1)$ . The result (2.9.21) follows. The result (2.9.22) is an immediate consequence of (2.9.21) and (2.9.20). ▲

The recursive equations of Theorem 2.9.3 are particularly effective for predictions of finite moving average processes. Because the autocovariances of the  $q$ th order moving average are zero for  $h > q$ , the  $c_{ii}$  of Theorem 2.9.3 are zero for  $i < t - q$ .

**Corollary 2.9.3.** Let  $Y_t$  be the  $q$ th order moving average

$$Y_t = e_t + \sum_{j=1}^q \beta_j e_{t-j},$$

where the  $e_j$  are uncorrelated  $(0, \sigma^2)$  random variables. Assume that the  $Z_j$ ,  $c_{ji}$ , and  $\kappa_j^2$  of Theorem 2.9.3 have been constructed for  $j = 1, 2, \dots, t-1$ , where  $t-1 > q$ . Then the  $c_{ii}$  of Theorem 2.9.3 are

$$\begin{aligned}c_{ii} &= 0, \quad i < t-q, \\ c_{i,t-q} &= \kappa_{t-q}^{-2} \gamma_Y(q),\end{aligned}$$

and

$$c_{ii} = \kappa_i^{-2} \left[ \gamma_Y(t-i) - \sum_{j=t-q}^{t-1} c_{ij} c_{ij} \kappa_j^2 \right]$$

for  $i = t-q+1, \dots, t-1$ .

**Proof.** Omitted. ▲

If the roots of the moving average process are less than one in absolute value, then the  $Z_t$  of Theorem 2.9.3 converges to  $e_t$ ,  $\kappa^2$  converges to  $\sigma^2$ , and  $c_{t,t-i}$  converges to  $\beta_i$  as  $t$  increases. For invertible moving averages, the coefficients are often approximated by the  $\beta_i$ . Let  $(\tilde{e}_1, \tilde{e}_2, \dots, \tilde{e}_q)$  be initial estimates of  $(e_1, e_2, \dots, e_q)$ , and let

$$\tilde{e}_t = Y_t - \sum_{j=1}^q \beta_j \tilde{e}_{t-j}, \quad t = q+1, q+2, \dots \quad (2.9.23)$$

Then  $\tilde{e}_n$  for  $n > q$  is

$$\tilde{e}_n = \sum_{j=n-q}^{n-1} c_j \tilde{e}_{n-j} + \sum_{j=0}^{n-1-q} c_j Y_{n-j},$$

where the  $c_j$  are defined in Theorem 2.6.2. For an invertible moving average, the  $|c_j|$  converge to zero as  $j$  increases, and  $\tilde{e}_n$  converges to  $e_n$  in mean square as  $n$  increases. The initial values of (2.9.23) can be computed using Theorem 2.9.3 or by setting  $\tilde{e}_{-q+1}, \tilde{e}_{-q+2}, \dots, \tilde{e}_0$  equal to zero and using (2.9.23) for  $t = 1, 2, \dots, q$ . Therefore, an easily computed  $s$ -period predictor for the  $q$ th order invertible moving average, appropriate for large  $n$ , is

$$\tilde{Y}_{n+s}(Y_1, \dots, Y_n) = \begin{cases} \sum_{j=s}^q \beta_j \tilde{e}_{n+s-j}, & 1 \leq s \leq q, \\ 0, & s > q, \end{cases} \quad (2.9.24)$$

where  $\tilde{e}_j$  is defined in (2.9.23). The prediction error for  $\tilde{Y}_{n+1}(Y_1, \dots, Y_n)$  goes to  $\sigma^2$  as  $n$  increases.

The procedures for autoregressive and moving average processes are easily extended to the autoregressive moving average time series defined by

$$Y_t = \sum_{j=1}^p \theta_j Y_{t-j} + \sum_{i=1}^p \beta_i e_{t-i} + e_t,$$

where the roots of

$$m^p - \sum_{j=1}^p \theta_j m^{p-j} = 0$$

and of

$$m^q + \sum_{i=1}^p \beta_i m^{q-i} = 0$$

are less than one in absolute value and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables. For such processes the predictor for large  $n$  is given by

$$\hat{Y}_{n+s}(Y_1, \dots, Y_n) = \begin{cases} \sum_{j=1}^p \theta_j Y_{n-j+1} + \sum_{i=1}^q \beta_i \tilde{e}_{n-i+1}, & s = 1, \\ \sum_{j=1}^{s-1} \theta_j \hat{Y}_{n-j+s}(Y_1, \dots, Y_n) + \sum_{j=s}^p \theta_j Y_{n-j+s} \\ \quad + \sum_{i=s}^q \beta_i \tilde{e}_{n-i+s}, & s = 2, 3, \dots, \end{cases} \quad (2.9.25)$$

where

$$\tilde{e}_t = \begin{cases} 0, & t = p, p-1, \dots, \\ Y_t - \sum_{j=1}^p \theta_j Y_{t-j} - \sum_{i=1}^q \beta_i \tilde{e}_{t-i}, & t = p+1, p+2, \dots, n, \end{cases}$$

and it is understood that  $\theta_j = 0$  for  $j > p$  and

$$\begin{aligned} \sum_{i=s}^q \beta_i \tilde{e}_{n-i+s} &= 0, \quad s = q+1, q+2, \dots, \\ \sum_{j=s}^p \theta_j Y_{n-j+s} &= 0, \quad s = p+1, p+2, \dots. \end{aligned}$$

The stationary autoregressive invertible moving average, which contains the autoregressive and invertible moving average processes as special cases, can be expressed as

$$Y_t = \sum_{j=0}^{\infty} \psi_j e_{t-j}, \quad (2.9.26)$$

where  $\psi_0 = 1$ ,  $\sum_{j=0}^{\infty} |\psi_j| < \infty$ , and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables. The variances of the forecast errors for predictions of more than one period are readily obtained when the time series is written in this form.

**Theorem 2.9.4.** Given a stationary autoregressive invertible moving average with the representation (2.9.26) and the predictor

$$\hat{Y}_{n+s}(Y_1, \dots, Y_n) = \sum_{j=s}^{n+s-1} \psi_j \tilde{e}_{n-j+s},$$

where  $\tilde{e}_t$ ,  $t = 1, 2, \dots, n$ , is defined following (2.9.25), then

$$\lim_{n \rightarrow \infty} E \left\{ \left[ Y_{n+s} - \hat{Y}_{n+s}(Y_1, \dots, Y_n) - \sum_{j=0}^{s-1} \psi_j e_{n-j+s} \right]^2 \right\} = 0.$$

**Proof.** The result follows immediately from Theorem 2.2.1.  $\blacktriangle$

Using Theorem 2.9.4, we can construct the covariance matrix of the  $s$  predictors for  $Y_{n+1}, Y_{n+2}, \dots, Y_{n+s}$ . Let  $\mathbf{Y}'_{n,s} = [Y_{n+1}, Y_{n+2}, \dots, Y_{n+s}]$ , and let

$$\hat{\mathbf{Y}}'_{n,s} = [\hat{Y}_{n+1}(Y_1, \dots, Y_n), \hat{Y}_{n+2}(Y_1, Y_2, \dots, Y_n), \dots, \hat{Y}_{n+s}(Y_1, \dots, Y_n)].$$

Then

$$\begin{aligned} & \lim_{n \rightarrow \infty} E\{(\mathbf{Y}_{n,s} - \hat{\mathbf{Y}}_{n,s})(\mathbf{Y}_{n,s} - \hat{\mathbf{Y}}_{n,s})'\} \\ &= \begin{pmatrix} 1 & u_1 & \cdots & u_{s-1} \\ u_1 & \sum_{j=0}^1 v_j^2 & \cdots & \sum_{j=0}^1 u_j u_{s+j-2} \\ \vdots & \vdots & & \vdots \\ u_{s-1} & \sum_{j=0}^{s-1} u_j u_{s+j-2} & \cdots & \sum_{j=0}^{s-1} v_j^2 \end{pmatrix} \sigma^2. \quad (2.9.27) \end{aligned}$$

Note that the predictors we have considered are unbiased: that is,

$$E\{Y_{n+s} - \hat{Y}_{n+s}(Y_1, \dots, Y_n)\} = 0.$$

Therefore, the mean square error of the predictor is the variance. To use this information to establish confidence limits for the predictor, we require knowledge of the distribution of the  $e_t$ . If the time series is normal, confidence intervals can be constructed using normal theory. That is, a  $1 - \eta$  level confidence interval is given by

$$\hat{Y}_{n+s}(Y_1, \dots, Y_n) \pm t_\eta [\text{Var}\{Y_{n+s} - \hat{Y}_{n+s}(Y_1, \dots, Y_n)\}]^{1/2},$$

where  $t_\eta$  is the value such that  $\eta$  of the probability of the standard normal distribution is beyond  $\pm t_\eta$ .

**Example 2.9.1.** In Example 6.3.1, we demonstrate that the U.S. quarterly seasonally adjusted unemployment rate for the period 1948–1972 behaves much like the time series

$$Y_t - 4.77 = 1.54(Y_{t-1} - 4.77) - 0.67(Y_{t-2} - 4.77) + e_t,$$

where the  $e_t$  are uncorrelated  $(0, 0.115)$  random variables. Let us assume that we know that this is the proper representation. The four observations for 1972 are 5.83, 5.77, 5.53, and 5.30. The predictions for 1973 are

$$\begin{aligned} \hat{Y}_{73-1}(Y_{72-1}, \dots, Y_{72-4}) &= 4.77 + 1.54(0.53) - 0.67(0.76) \\ &= 5.08, \end{aligned}$$

$$\begin{aligned}\hat{Y}_{73-2}(Y_{72-1}, \dots, Y_{72-4}) &= 4.77 + 1.54(0.3070) - 0.67(0.53) \\ &= 4.89,\end{aligned}$$

$$\begin{aligned}\hat{Y}_{73-3}(Y_{72-1}, \dots, Y_{72-4}) &= 4.77 + 1.54(0.1177) - 0.67(0.3070) \\ &= 4.75,\end{aligned}$$

$$\begin{aligned}\hat{Y}_{73-4}(Y_{72-1}, \dots, Y_{72-4}) &= 4.77 + 1.54(-0.0245) - 0.67(0.1177) \\ &= 4.65.\end{aligned}$$

The covariance matrix of the prediction errors can be obtained by using the representation  $Y_t = \sum_{j=0}^{\infty} w_j e_{t-j}$ , where the  $w$ 's are defined in Theorem 2.6.1. We have  $a_1 = -1.54$ ,  $a_2 = 0.67$ , and

$$w_0 = 1,$$

$$w_1 = 1.54,$$

$$w_2 = (1.54)^2 - 0.67 = 1.702,$$

$$w_3 = 1.54(1.702) - 0.67(1.54) = 1.589.$$

Therefore, the covariance matrix of equation (2.9.27) given by the product

$$0.115\mathbf{H}'\mathbf{H} = \begin{pmatrix} 0.115 & 0.177 & 0.196 & 0.183 \\ 0.177 & 0.388 & 0.478 & 0.477 \\ 0.196 & 0.478 & 0.720 & 0.789 \\ 0.183 & 0.477 & 0.789 & 1.011 \end{pmatrix},$$

where

$$\mathbf{H}' = \begin{pmatrix} 1.00 & 0 & 0 & 0 \\ 1.54 & 1.00 & 0 & 0 \\ 1.70 & 1.54 & 1.00 & 0 \\ 1.59 & 1.70 & 1.54 & 1.00 \end{pmatrix}.$$

We observe that there is a considerable increase in the variance of the predictor as  $s$  increases from 1 to 4. Also, there is a high positive correlation between the predictors.  $\blacktriangle\blacktriangle$

**Example 2.9.2.** In Table 2.9.1 are seven observations from the time series

$$Y_t = 0.9Y_{t-1} + 0.7e_{t-1} + e_t,$$

where the  $e_t$  are normal independent  $(0, 1)$  random variables. The first 11 autocorrelations for this time series are, approximately, 1.00, 0.95, 0.85, 0.77, 0.69, 0.62, 0.56, 0.50, 0.45, 0.41, and 0.37. The variance of the time series is about 14.4737.

To construct the linear predictor for  $Y_8$ , we obtain the seven weights by solving

**Table 2.9.1.** Observations and predictions for the time series  $Y_t = 0.9Y_{t-1} + 0.7e_{t-1} + e_t$

<i>t</i>	Observation	Prediction
1	-1.58	—
2	-2.86	—
3	-3.67	—
4	-2.33	—
5	0.42	—
6	2.87	—
7	3.43	—
8	—	3.04
9	—	2.73
10	—	2.46

the system of equations

$$\begin{pmatrix} 1.00 & 0.95 & 0.85 & \cdots & 0.56 \\ 0.95 & 1.00 & 0.95 & \cdots & 0.62 \\ 0.85 & 0.95 & 1.00 & \cdots & 0.69 \\ \vdots & \vdots & \vdots & & \vdots \\ 0.56 & 0.62 & 0.69 & \cdots & 1.00 \end{pmatrix} \begin{pmatrix} b_{11} \\ b_{12} \\ b_{13} \\ \vdots \\ b_{17} \end{pmatrix} = \begin{pmatrix} 0.50 \\ 0.56 \\ 0.62 \\ \vdots \\ 0.95 \end{pmatrix}. \quad (2.9.28)$$

This system of equations gives the weights for the one-period-ahead forecast. The vector of weights  $b'_{7,1} = (0.06, -0.18, 0.32, -0.50, 0.75, -1.10, 1.59)$  is applied to  $(Y_1, Y_2, Y_3, Y_4, Y_5, Y_6, Y_7)$ . [In obtaining the solution more digits were used than are displayed in equation (2.9.28).] The predictor for two periods ahead is obtained by replacing the right side of (2.9.28) by  $(0.45, 0.50, 0.56, 0.62, 0.77, 0.85)'$ , and the weights for the three-period-ahead predictor are obtained by using  $(0.41, 0.45, 0.50, 0.56, 0.62, 0.69, 0.77)'$  as the right side. The predictions using these weights are displayed in Table 2.9.1.

The covariance matrix of the prediction errors is

$$E\{(\mathbf{Y}_{7,3} - \hat{\mathbf{Y}}_{7,3})(\mathbf{Y}_{7,3} - \hat{\mathbf{Y}}_{7,3})'\} = \begin{pmatrix} 1.005 & 1.604 & 1.444 \\ 1.604 & 3.563 & 3.907 \\ 1.444 & 3.907 & 5.637 \end{pmatrix},$$

where  $\hat{\mathbf{Y}}_{7,3}$  and  $\mathbf{Y}_{7,3}$  are defined following Theorem 2.9.4. The covariance matrix of prediction errors was obtained as  $\mathbf{AV}\mathbf{A}'$ , where  $\mathbf{V}$  is the  $10 \times 10$  covariance matrix for 10 observations and

$$\mathbf{A} = \begin{bmatrix} -0.06 & 0.18 & -0.32 & 0.50 & -0.75 & 1.10 & -1.59 & 1 & 0 & 0 \\ -0.05 & 0.16 & -0.29 & 0.45 & -0.68 & 0.99 & -1.44 & 0 & 1 & 0 \\ -0.05 & 0.14 & -0.26 & 0.41 & -0.61 & 0.89 & -1.29 & 0 & 0 & 1 \end{bmatrix}.$$

If we had an infinite past to use in prediction, the variance of the one-period-ahead prediction error would be one, the variance of  $e_t$ . Thus, the loss in prediction efficiency from having only seven observations is about  $\frac{1}{2}\%$ . Since the time series is normal, we can construct 95% confidence intervals for the predictions. We obtain the intervals (1.08, 5.00), (-0.97, 6.43), and (-2.19, 7.11) for the predictions one, two, and three periods ahead, respectively.

To construct a predictor using the  $\tilde{e}$ -method, we set  $\tilde{e}_1 = 0$  and calculate

$$\begin{aligned}\tilde{e}_2 &= Y_2 - 0.9Y_1 - 0.7\tilde{e}_1 = -2.86 - 0.9(-1.58) = -1.438, \\ \tilde{e}_3 &= Y_3 - 0.9Y_2 - 0.7\tilde{e}_2 = -3.67 - 0.9(-2.86) - 0.7(-1.438) = -0.089, \\ \tilde{e}_4 &= -2.33 - 0.9(-3.67) - 0.7(-0.089) = 1.035, \\ \tilde{e}_5 &= 0.42 - 0.9(-2.33) - 0.7(1.035) = 1.793, \\ \tilde{e}_6 &= 2.87 - 0.9(0.42) - 0.7(1.793) = 1.237, \\ \tilde{e}_7 &= 3.43 - 0.9(2.86) - 0.7(1.227) = -0.003.\end{aligned}$$

Therefore, the predictor of  $Y_8$  is

$$\hat{Y}_8(Y_1, \dots, Y_7) = 0.9Y_7 + 0.7\tilde{e}_7 = 3.085,$$

and

$$\begin{aligned}\hat{Y}_9(Y_1, \dots, Y_7) &= 0.9\hat{Y}_8(Y_1, \dots, Y_7) = 2.776, \\ \hat{Y}_{10}(Y_1, \dots, Y_7) &= 0.9\hat{Y}_9(Y_1, \dots, Y_7) = 2.498.\end{aligned}$$

In this case the prediction using the approximate method of setting  $\tilde{e}_1 = 0$  is quite close to the prediction obtained by the method of Theorem 2.9.1, even though the sample size is small. If the coefficient of  $e_{t-1}$  had been closer to one, the difference would have been larger.

By Theorem 2.7.1, we may write our autoregressive moving average process as

$$Y_t = \sum_{j=0}^{\infty} u_j e_{t-j},$$

where

$$\begin{aligned}u_0 &= b_0 = 1, \\ u_1 &= b_1 - a_1 = 0.7 + 0.9 = 1.6, \\ u_j &= -a_1 u_{j-1} = (0.9)u_{j-1}, \quad j = 2, 3, \dots.\end{aligned}$$

Therefore, the large sample covariance matrix of prediction errors for predictors one, two, and three periods ahead is

$$\begin{pmatrix} 1.00 & 1.60 & 1.44 \\ 1.60 & 3.56 & 3.90 \\ 1.44 & 3.90 & 5.63 \end{pmatrix},$$

which is very close to the covariance matrix of the best predictor based on  $n = 7$  observations.  $\blacktriangle \blacktriangle$

## 2.10 THE WOLD DECOMPOSITION

We obtained moving average representations for stationary autoregressive and autoregressive moving average processes in Sections 2.6 and 2.7. In this section, we give a moving average representation, due to Wold (1938), that is appropriate for any stationary process. We first investigate the properties of the sequence of one-period prediction errors of a stationary nondeterministic process. We are particularly interested in the limit as  $n$  becomes large.

**Theorem 2.10.1.** Let  $\{Y_t\}$  be a stationary, nondeterministic time series defined on the integers. Then there exists a unique (a.s.) time series  $\{e_t\}$  defined as the limit in mean square of the one-period prediction error as  $n$  becomes large. Furthermore,  $\{e_t\}$  is a sequence of uncorrelated random variables with mean zero and common variance  $\sigma^2$ , where

$$\sigma^2 = \lim_{n \rightarrow \infty} \tau_{n1}^2, \quad (2.10.1)$$

and  $\tau_{n1}^2$  is the variance of the one-period prediction error defined in Theorem 2.9.1.

**Proof.** Without loss of generality, let the mean of  $Y_t$  be zero. Define the new time series

$$\begin{aligned} W_{1t} &= Y_{t-1}, \\ W_{2t} &= Y_{t-2} - [\gamma_Y(0)]^{-1} \gamma_Y(1) Y_{t-1}, \\ &\vdots \\ W_{it} &= Y_{t-i} - \hat{Y}_{t-i}(Y_{t-i+1}, \dots, Y_{t-1}) \\ &= Y_{t-i} - \sum_{j=1}^{i-1} b_{i-1,1,j} Y_{t-j} \quad i = 2, 3, \dots, \end{aligned} \quad (2.10.2)$$

where

$$\mathbf{b}_{i-1,1} = (b_{i-1,1,1}, b_{i-1,1,2}, \dots, b_{i-1,1,i-1})', \quad (2.10.3)$$

$$E\{W_{it}^2\} = \tau_{i-1,1}^2, \quad i = 2, 3, \dots, \quad (2.10.4)$$

and  $\mathbf{b}_{n1}$  and  $\tau_{n1}^2$  are defined in Theorem 2.9.1. Because  $Y_t$  is regular,  $E\{W_{it}^2\}$  is bounded below and the vectors  $\mathbf{b}_{i-1,1}$  are unique. Each  $W_{it}$  is a stationary time

series, and

$$E\{W_{it}W_{jt}\} = 0, \quad i \neq j. \quad (2.10.5)$$

Let

$$e_t(n) = Y_t - \sum_{i=1}^n c_i W_{it}, \quad (2.10.6)$$

where

$$(c_1, c_2, \dots, c_n)' = [E\{\mathbf{w}_{nt}\mathbf{w}_{nt}'\}]^{-1} E\{\mathbf{w}_{nt}Y_t\},$$

$$\mathbf{w}'_{nt} = (W_{1t}, W_{2t}, \dots, W_{nt}). \quad (2.10.7)$$

Then

$$E\{e_t^2(n)\} = \gamma_Y(0) - \sum_{j=1}^n c_j^2 \tau_{j-1,1}^2 = \tau_{n1}^2. \quad (2.10.8)$$

By (2.10.8),  $\tau_{n1}^2$  is monotone decreasing in  $n$ , and by the assumption that  $Y_t$  is regular,  $\tau_{n1}^2$  is bounded below. Therefore, the limit

$$\lim_{n \rightarrow \infty} \tau_{n1}^2 = \sigma^2 \quad (2.10.9)$$

is well defined and

$$\lim_{n \rightarrow \infty} \sum_{j=1}^n c_j^2 \tau_{j-1,1}^2 = \gamma_Y(0) - \sigma^2.$$

By the definitions,

$$\begin{aligned} E\{|e_t(n) - e_t(n+m)|^2\} &= \sum_{j=1}^m c_{n+j}^2 \tau_{n+j-1,1}^2 \\ &\leq \sum_{j=n+1}^{\infty} c_j^2 \tau_{j-1,1}^2, \end{aligned}$$

which converges to zero uniformly in  $m$  as  $n \rightarrow \infty$ . Therefore, by Exercise 4.4 of Tucker (1967, p. 106) there exists a random variable  $e_t$  such that  $E\{e_t^2\} < \infty$  and  $e_t(n)$  converges to  $e_t$ , in mean square. For any fixed  $h$ ,

$$\begin{aligned} E\{|e_t e_{t+h} - e_t(n)e_{t+h}(n)|\} &= E\{|[e_t - e_t(n)]e_{t+h} + e_t(n)[e_{t+h} - e_{t+h}(n)]|\| \\ &\leq (E\{|e_t - e_t(n)|^2\} E\{e_{t+h}^2\})^{1/2} \\ &\quad + (E\{e_t^2(n)\} E\{|e_{t+h} - e_{t+h}(n)|^2\})^{1/2} \\ &\rightarrow 0 \quad \text{as } n \rightarrow \infty, \end{aligned}$$

because  $E\{e_i^2(n)\} < \gamma_r(0)$ . Therefore, by Exercise 4.14 of Royden (1989, p. 93),

$$E\{e_i e_{i+h}\} = \lim_{n \rightarrow \infty} E\{e_i(n) e_{i+h}(n)\}.$$

Similarly,

$$E\{e_i\} = \lim_{n \rightarrow \infty} E\{e_i(n)\} = 0.$$

The covariance between  $e_i(n)$  and  $e_{i+h}(n)$ , for  $n > h > 0$ , is

$$\begin{aligned} E\{e_i(n)e_{i+h}(n)\} &= E\left\{e_{i+h}(n) \left( Y_i - \sum_{i=1}^{n-h} c_i W_{it} - \sum_{i=n-h+1}^n c_i W_{it} \right) \right\} \\ &= -E\left\{e_{i+h}(n) \sum_{i=n-h+1}^n c_i W_{it}\right\}, \end{aligned}$$

where we have used the fact that  $e_{i+h}(n)$  is uncorrelated with  $Y_{i+h-1}, Y_{i+h-2}, \dots, Y_{i+h-n}$ . Because

$$\begin{aligned} |E\{e_i(n)e_{i+h}(n)\}|^2 &\leq \tau_{n1}^2 \sum_{i=n-h+1}^n c_i^2 \tau_{i-1,1}^2 \\ &\leq \gamma_r(0) \sum_{i=n-h+1}^{\infty} c_i^2 \tau_{i-1,1}^2 \end{aligned}$$

we have

$$\lim_{n \rightarrow \infty} E\{e_i(n)e_{i+h}(n)\} = E\{e_i e_{i+h}\} = 0.$$

Similarly,

$$E(e_i^2) = \lim_{n \rightarrow \infty} E[e_i^2(n)] = \sigma^2. \quad \blacktriangle$$

We now give Wold's theorem for the representation of a stationary time series. The theorem states that it is possible to express every stationary time series as the sum of two time series: an infinite moving average time series and a singular time series.

**Theorem 2.10.2** (Wold decomposition). Let  $Y_t$  be a covariance stationary time series defined on  $T = \{0, \pm 1, \pm 2, \dots\}$ . Then  $Y_t$  has the representation

$$Y_t = X_t + Z_t, \tag{2.10.10}$$

where

$$X_t = \sum_{i=0}^{\infty} a_i e_{t-i},$$

$a_0 = 1$ ,  $\{Z_t\}$  is singular,  $\{Z_t\}$  and  $\{X_t\}$  are uncorrelated,  $\{e_t\}$  is the sequence of

uncorrelated  $(0, \sigma^2)$  random variables defined in Theorem 2.10.1, and  $X_t$  is defined as a limit in mean square.

**Proof.** If  $Y_t$  is singular, then  $Y_t = Z_t$  and  $X_t = 0$ . Suppose  $Y_t$  is nonsingular. Let  $\{e_i\}$  be the time series defined by Theorem 2.10.1. Let

$$X_t(n) = \sum_{i=0}^n a_i e_{t-i},$$

where

$$(a_0, a_1, \dots, a_n) = [\gamma_e(0)]^{-1} [\gamma_{eY}(0), \gamma_{eY}(1), \dots, \gamma_{eY}(n)],$$

and

$$\gamma_{eY}(h) = \text{Cov}(e_i, Y_{i+h}) = \text{Cov}(Y_i, e_{i-h}).$$

We have

$$\gamma_Y(0) \geq \gamma_{X_t(n)}(0) = \sum_{i=0}^n a_i^2 \gamma_e(0) = \sigma^2 \sum_{i=0}^n a_i^2$$

and  $\sum_{i=0}^{\infty} a_i^2 < \infty$ . It follows that

$$\sum_{i=0}^n a_i e_{t-i} \rightarrow \sum_{i=0}^{\infty} a_i e_{t-i} = X_t$$

in mean square. Let  $Z_t = Y_t - X_t$ . Using the arguments of Theorem 2.10.1, it follows that

$$\text{Cov}(X_{t+h}, e_i) = a_h \sigma^2 = \text{Cov}(Y_{t+h}, e_i), \quad h \geq 0,$$

$$\text{Cov}(Y_t, e_{t+h}) = \lim_{n \rightarrow \infty} \text{Cov}\left(Y_t, Y_{t+h} - \sum_{i=1}^n c_i W_{i,t+h}\right) = 0, \quad h > 0,$$

$$\text{Cov}(X_t, e_{t+h}) = \text{Cov}\left(\sum_{i=0}^{\infty} a_i e_{t-i}, e_{t+h}\right) = 0, \quad h > 0.$$

Therefore, the processes  $\{Z_t\}$  and  $\{X_t\}$  are uncorrelated. For any set of real numbers  $\{\delta_i : i = 0, 1, \dots, p\}$  with  $\delta_0 = 1$ ,

$$V\left\{\sum_{i=0}^p \delta_i Y_{t-i}\right\} = V\left\{\sum_{i=0}^p \delta_i Z_{t-i}\right\} + V\left\{\sum_{i=0}^p \delta_i X_{t-i}\right\},$$

where we abbreviate  $\text{Var}\{X_t\}$  to  $V\{X_t\}$ . By (2.10.9), given  $\epsilon > 0$ , there is some  $p_\epsilon$

and a set of  $\delta_i$  defined by (2.10.6) such that

$$V\left\{\sum_{i=0}^{p_\epsilon} \delta_i Y_{t-i}\right\} < \sigma^2 + \epsilon.$$

Also,

$$\sum_{i=0}^p \delta_i X_{t-i} = e_t + (\delta_1 + a_1)e_{t-1} + (\delta_2 + \delta_1 a_1 + a_2)e_{t-2} + \dots$$

and  $V\{\sum_{i=0}^p \delta_i X_{t-i}\} \geq \sigma^2$ . Therefore, given  $\epsilon > 0$ , there exist  $p_\epsilon$  and  $\delta_i$  such that

$$V\left\{\sum_{i=0}^{p_\epsilon} \delta_i Z_{t-i}\right\} < \epsilon,$$

and we conclude that  $\{Z_t\}$  is singular. ▲

Note that the mean can be included in the singular part of the representation (2.10.10). Proofs of the Wold decomposition using Hilbert space arguments are available in Anderson (1971, p. 420) and Brockwell and Davis (1991, p. 187).

## 2.11. LONG MEMORY PROCESSES

In Section 2.7, we demonstrated that the autocorrelation function of an autoregressive moving average process declines at an exponential rate. By the results of Section 2.10, any nondeterministic stationary process can be represented as an infinite moving average of uncorrelated random variables. The coefficients of the sum decline exponentially if the original process is an autoregressive moving average. Because of this rapid rate of decline, finite autoregressive moving averages are sometimes called *short memory* processes.

Consider a weighted average of uncorrelated  $(0, \sigma^2)$  random variables

$$Y_t = \sum_{j=0}^{\infty} \nu_j e_{t-j} \quad (2.11.1)$$

in which the coefficients of  $e_{t-j}$  are declining at the rate  $j^{d-1}$  for some  $d \in (-0.5, 0.5)$ . Such a time series is well defined as a limit in mean square by Theorem 2.2.3. For the special case  $\nu_j = (j+1)^{d-1}$ , we have

$$E\{Y_t^2\} = \sum_{j=0}^{\infty} (j+1)^{2(d-1)} \sigma^2 = [1 - (2d-1)^{-1}(1.5)^{2d-1}] \sigma^2, \quad (2.11.2)$$

$$\gamma_Y(h) = E\{Y_t Y_{t+h}\} = \sum_{i=1}^{\infty} i^{d-1} (i+h)^{d-1} \sigma^2, \quad h > 0, \quad (2.11.3)$$

where we have used the integral approximation to the sum in (2.11.2). For large  $h$ ,  $\gamma_y(h)$  is approximately equal to  $ch^{2d-1}$ , where  $c$  is a constant. See Exercise 2.44. We say that a process  $X_t$  is a *long memory* process if  $\rho_X(h)$  is approximately proportional to  $h^{2d-1}$  for  $d \in (-0.5, 0.5)$ .

A time series that has received considerable attention is

$$X_t = \sum_{j=0}^{\infty} u_j e_{t-j}, \quad (2.11.4)$$

where  $u_0 = 1$ ,

$$\begin{aligned} u_j &= [\Gamma(j+1)\Gamma(d)]^{-1}\Gamma(j+d) \\ &= \prod_{i=1}^j i^{-1}(i-1+d), \quad j = 1, 2, \dots, \end{aligned} \quad (2.11.5)$$

and  $\Gamma(\cdot)$  is the gamma function. Using Stirling's approximation that  $\Gamma(s+1) \approx (2\pi s)^{1/2} e^{-s} s^s$  for large  $s$ , we have

$$u_j \approx [\Gamma(d)]^{-1} j^{d-1}$$

for large  $j$ . It can be shown that the autocovariance, autocorrelation, and partial autocorrelation of the time series (2.11.4) are

$$\gamma_X(h) = \sigma^2 [\Gamma(h-d+1)\Gamma(1-h-d)]^{-1} (-1)^h \Gamma(1-2d), \quad h = 0, 1, 2, \dots, \quad (2.11.6)$$

$$\begin{aligned} \rho_X(h) &= [\Gamma(h-d+1)\Gamma(d)]^{-1} \Gamma(h+d)\Gamma(1-d) \\ &= \prod_{i=1}^h (i-d)^{-1}(i-1+d), \quad h = 1, 2, \dots, \end{aligned} \quad (2.11.7)$$

and

$$\phi_X(h) = (h-d)^{-1} d, \quad h = 1, 2, \dots. \quad (2.11.8)$$

Using Stirling's approximation,

$$\rho_X(h) \approx [\Gamma(d)]^{-1} \Gamma(1-d) h^{2d-1}.$$

Thus, the moving average with weights (2.11.5) is a long memory process.

The time series (2.11.4) can be written

$$X_t = (1 - \mathcal{B})^{-d} e_t, \quad (2.11.9)$$

where  $(1 - \mathcal{B})^r$  is defined by the expansion

$$(1 - \mathcal{B})^r = 1 - r\mathcal{B} - (2!)^{-1}r(1-r)\mathcal{B}^2 - (3!)^{-1}r(1-r)(2-r)\mathcal{B}^3 - \dots$$

for  $-0.5 < r < 0$  and  $0 < r < 0.5$ . For  $r = 0$ , the operator is the identity operator. The operator  $(1 - \mathcal{B})^d$  for  $d \in (-0.5, 0.5)$  is called the *fractional difference operator*. The  $j$ th coefficient in the expansion of  $(1 - \mathcal{B})^{-d}$  is the  $\nu_j$  of (2.11.5).

For the time series (2.11.4), we can also write

$$(1 - \mathcal{B})^d X_t = e_t, \quad (2.11.10)$$

or

$$X_t + \sum_{j=1}^{\infty} \kappa_j(d) X_{t-j} = e_t, \quad (2.11.11)$$

where

$$\kappa_j(d) = [\Gamma(j+1)\Gamma(-d)]^{-1}\Gamma(j-d) = \prod_{i=1}^j i^{-1}(i-1-d). \quad (2.11.12)$$

If the time series satisfies (2.11.10) with  $d \in (-0.5, 0.5)$ , then it has an infinite moving average representation in which the coefficients decline at the rate  $j^{d-1}$  and an infinite autoregressive representation in which the coefficients decline at the

**Table 2.11.1.** Coefficients for fractional differenced time series

Index	$d = 0.4$			$d = -0.4$		
	$\nu_j$	$\kappa_j$	$\rho(j)$	$\nu_j$	$\kappa_j$	$\rho(j)$
1	0.4000	-0.4000	0.6667	-0.4000	0.4000	-0.2857
2	0.2800	-0.1200	0.5833	-0.1200	0.2800	-0.0714
3	0.2240	-0.0640	0.5385	-0.0640	0.2240	-0.0336
4	0.1904	-0.0416	0.5085	-0.0416	0.1904	-0.0199
5	0.1676	-0.0300	0.4864	-0.0300	0.1676	-0.0132
6	0.1508	-0.0230	0.4691	-0.0230	0.1508	-0.0095
7	0.1379	-0.0184	0.4548	-0.0184	0.1379	-0.0072
8	0.1275	-0.0152	0.4429	-0.0152	0.1275	-0.0057
9	0.1190	-0.0128	0.4326	-0.0128	0.1190	-0.0046
10	0.1119	-0.0110	0.4236	-0.0110	0.1119	-0.0038
15	0.0881	-0.0062	0.3906	-0.0062	0.0881	-0.0018
20	0.0743	-0.0041	0.3688	-0.0041	0.0743	-0.0011
25	0.0650	-0.0030	0.3527	-0.0030	0.0650	-0.0007
30	0.0583	-0.0023	0.3400	-0.0023	0.0583	-0.0005
35	0.0532	-0.0019	0.3297	-0.0019	0.0532	-0.0004
40	0.0491	-0.0015	0.3210	-0.0015	0.0491	-0.0003

rate  $j^{-d-1}$ . Thus, if  $d > 0$ , the moving average coefficients are square summable and the autoregressive coefficients are absolutely summable. If  $d < 0$ , the moving average coefficients are absolutely summable and the autoregressive coefficients are square summable. Examples of the coefficients are given in Table 2.11.1.

To generalize the class of time series with long memory properties, we assume  $Y_t$  satisfies

$$(1 - \mathcal{B})^d Y_t = Z_t, \quad (2.11.13)$$

where  $Z_t$  is the stationary autoregressive moving average satisfying

$$\sum_{j=0}^p \alpha_j Z_{t-j} = \sum_{i=0}^q \beta_i e_{t-i}, \quad (2.11.14)$$

$e_t$  are uncorrelated  $(0, \sigma^2)$  random variables, the roots of the characteristic equations associated with (2.11.14) are less than one in absolute value, and  $d \in (-0.5, 0.5)$ . The time series  $Y_t$  is sometimes called a fractionally differenced autoregressive moving average. The representation is sometimes abbreviated to ARIMA( $p, d, q$ ). By Theorem 2.2.2 and Theorem 2.2.3, the time series  $Y_t$  can be given an infinite moving average or an infinite autoregressive representation.

## REFERENCES

- Sections 2.1, 2.3, 2.5–2.9.** Anderson (1971), Bartlett (1966), Box, Jenkins and Reinsel (1994), Hannan (1970), Kendall and Stuart (1966), Pierce (1970a), Wold (1938), Yule (1927).
- Section 2.2.** Bartle (1964), Pantula (1988a), Royden (1989), Torres (1986), Tucker (1967).
- Section 2.4.** Bellman (1960), Finkbeiner (1960), Goldberg (1958), Hildebrand (1968), Kempthorne and Folks (1971), Miller (1963, 1968).
- Section 2.10.** Anderson (1971), Brockwell and Davis (1991), Royden (1989), Tucker (1967), Wold (1938).
- Section 2.11.** Beran (1992), Cox (1984), Deo (1995), Granger (1980), Granger and Joyeux (1980), Hosking (1981, 1982), Mandelbrot (1969).

## EXERCISES

1. Express the coefficients  $b_i$  of (2.4.3) in terms of the coefficients  $a_i$  of (2.4.2) for  $n = 3$ .
2. Given that  $\{e_t : t \in (0, \pm 1, \pm 2, \dots)\}$  is a sequence of independent  $(0, \sigma^2)$  random variables, define

$$(X_t, Y_t) = \left( \sum_{i=-2}^2 e_{t-i}, \sum_{i=-1}^1 X_{t-i} \right).$$

- (a) Derive and graph  $\gamma_X(h)$  and  $\gamma_Y(h)$ .
- (b) Express  $Y_t$  as a moving average of  $e_t$ .
- (c) Is it possible to express  $e_t$  as a finite moving average of  $X_t$ ?
3. For the second order autoregressive process, obtain an expression for  $\rho(h)$  as a function of  $\alpha_1$  and  $\alpha_2$  analogous to (2.5.10).
4. Given a second order moving average time series, what are the largest and smallest possible values for  $\rho(1)$ ? For  $\rho(2)$ ? For  $\rho(3)$ ? Give an example of a moving average time series such that  $\gamma(0) = 2$ ,  $\gamma(1) = 0$ ,  $\gamma(2) = -1$ , and  $\gamma(h) = 0$  for  $|h| > 2$ .
5. Consider the difference equation
- $$y_t - 1.6y_{t-1} + 0.89y_{t-2} = 0, \quad t = 2, 3, \dots$$
- (a) Give the characteristic equation, and find its roots.
- (b) Give the expression for  $y_t$  if  $y_0 = 0.8$  and  $y_1 = 0.9$ .
- (c) If  $e_t \sim NI(0, \sigma^2)$ ,  $\sigma^2 > 0$ ,  $t = \{0, \pm 1, \pm 2, \dots\}$ , is there a stationary time series satisfying
- $$Y_t - 1.6Y_{t-1} + 0.89Y_{t-2} = e_t?$$
- Explain.
- (d) Let the time series  $Y_t$ ,  $t = 2, 3, 4, \dots$ , be defined by
- $$Y_t = 1.6Y_{t-1} - 0.89Y_{t-2} + e_t,$$
- where  $(Y_0, Y_1) = (0, 0)$  and  $e_t \sim NI(0, \sigma^2)$ . Give the covariance matrix for  $(Y_2, Y_3, Y_4)$ .
- (e) Define  $(Y_0, Y_1)$  so that  $Y_t$  of (d) is a stationary time series.
6. Find four (admittedly similar) time series with the covariance function

$$\gamma(h) = \begin{cases} 1, & h = 0, \\ 0.3, & |h| = 1, \\ 0 & \text{otherwise.} \end{cases}$$

7. Draw pictures illustrating the differences among the theoretical correlation functions for
- (a) A (finite) moving average process.
- (b) A (finite) autoregressive process.

(c) A strictly periodic process of the form

$$X_t = \sum_{i=1}^M \{e_{1i} \cos \lambda_i t + e_{2i} \sin \lambda_i t\}, \quad M < \infty,$$

where  $e_{1i}$  and  $e_{2i}$  are independent  $(0, \sigma^2)$  random variables independent of  $e_{1j}$  and  $e_{2j}$  for all  $i \neq j$ , and  $\lambda_i$ ,  $i = 1, 2, \dots, M$ , are distinct frequencies.

8. Consider the time series  $\{X_t\}$  defined by

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} + \alpha_3 X_{t-3} = e_t,$$

where the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables and the roots of

$$m^3 + \alpha_1 m^2 + \alpha_2 m + \alpha_3 = 0$$

are less than one in absolute value. Develop a system of equations defining  $\gamma(0)$  as a function of the  $\alpha$ 's.

9. Let the time series  $Y_t$  be defined by

$$Y_t = \beta_0 + \beta_1 t + X_t, \quad t = 1, 2, \dots,$$

where

$$X_t = e_t + 0.6e_{t-1},$$

$\{e_t : t \in (0, 1, 2, \dots)\}$  is a sequence of normal independent  $(0, \sigma^2)$  random variables, and  $\beta_0$  and  $\beta_1$  are fixed numbers. Give the mean and covariance functions for the time series.

10. The second order autoregressive time series

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = e_t,$$

where the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables and the roots of  $m^2 + \alpha_1 m + \alpha_2 = 0$  are less than one in absolute value, can be expressed as the infinite moving average

$$X_t = \sum_{j=0}^{\infty} w_j e_{t-j}.$$

For the following three time series find and plot  $w_j$  for  $j = 0, 1, \dots, 8$  and  $\rho(h)$  for  $h = 0, 1, \dots, 8$ :

- (a)  $X_t + 1.6X_{t-1} + 0.64X_{t-2} = e_t$ .
- (b)  $X_t - 0.4X_{t-1} - 0.45X_{t-2} = e_t$ .
- (c)  $X_t - 1.2X_{t-1} + 0.85X_{t-2} = e_t$ .

**11. Given the time series**

$$Y_t - 0.8Y_{t-1} = e_t + 0.7e_{t-1} + 0.6e_{t-2},$$

where the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables, find and plot  $\rho(h)$  for  $h = 0, 1, 2, \dots, 8$ .

**12. Prove Theorem 2.1.1.**

- 13. Find a second order autoregressive time series with complex roots whose absolute value is 0.8 that will give realizations with an apparent cycle of 20 time periods.**
- 14. Given that  $X_t$  is a first order autoregressive process with parameter  $\rho = 0.8$ , use Theorem 2.9.1 to find the best one- and two-period-ahead predictions based on three observations.**

- 15. Let the time series  $X_t$  satisfy**

$$(X_t - 4) - 0.8(X_{t-1} - 4) = e_t,$$

where the  $e_t$  are uncorrelated  $(0, 7)$  random variables. Five observations from a realization of the time series are given below:

$t$	$X_t$
1	5.9
2	4.9
3	2.2
4	2.0
5	4.9

Predict  $X_t$  for  $t = 6, 7, 8$ , and estimate the covariance matrix for your predictions, that is, estimate the matrix  $E\{(\mathbf{X}_{5,3} - \hat{\mathbf{X}}_{5,3})(\mathbf{X}_{5,3} - \hat{\mathbf{X}}_{5,3})'\}$ .

- 16. Let  $Y_t$  be defined by**

$$Y_t = \rho Y_{t-1} + e_t, \quad |\rho| < 1,$$

where  $\{e_t, t \in (0, \pm 1, \pm 2, \dots)\}$  is obtained from the sequence  $\{u_t\}$  of normal independent  $(0, 1)$  random variables as follows:

$$e_t = \begin{cases} u_t, & t = 0, \pm 2, \pm 4, \dots, \\ 2^{-1/2}(u_{t-1}^2 - 1), & t = \pm 1, \pm 3, \dots. \end{cases}$$

Is this a strictly stationary process? Is it a covariance stationary process? Given a sample of ten observations  $(Y_1, Y_2, \dots, Y_{10})$ , what is the best linear predictor of  $Y_{11}$ ? Can you construct a better predictor of  $Y_{11}$ ?

17. Let  $Y_t$  be defined by

$$Y_t = e_t + \beta e_{t-1}, \quad |\beta| < 1.$$

Show that the predictor of  $Y_{n+1}$  given in (2.9.17) is equivalent to the predictor  $\hat{Y}_{n+1}(Y_1, \dots, Y_n) = -\sum_{j=1}^n (-\beta)^j Y_{n-j+1}$ .

18. Let  $\{e_t\}$  be a sequence of independent  $(0, 1)$  random variables. Let the time series  $X_t$  be defined by:

- (a)  $X_t = e_0$ .
- (b)  $X_t = e_t - 1.2e_{t-1} + 0.36e_{t-2}$ .
- (c)  $X_t = e_1 \cos 0.4\pi t + e_2 \sin 0.4\pi t + 6$ .
- (d)  $X_t = e_t + 1.2e_{t-1}$ .

Assume the covariance structure is known and we have available an infinite past for prediction, that is,  $\{X_t : t \in (\dots, -1, 0, 1, \dots, n)\}$  is known. What is the variance of the prediction error for a predictor of  $X_{n+1}$  for each case above?

19. Show that if a stationary time series satisfies the difference equation

$$Y_t - Y_{t-1} = e_t,$$

then  $E\{e_t^2\} = 0$ .

20. Let  $\{X_t : t \in (0, \pm 1, \pm 2, \dots)\}$  be defined by

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = e_t,$$

where the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables, and the roots  $r_1$  and  $r_2$  of

$$m^2 + \alpha_1 m + \alpha_2 = 0$$

are less than one in absolute value.

- (a) Let  $r_1$  and  $r_2$  be real. Show that  $X_t - r_1 X_{t-1}$  is a first order autoregressive process with parameter  $r_2$ .
- (b) If the roots  $r_1$  and  $r_2$  form a complex conjugate pair, how would you describe the time series  $Y_t = X_t - r_1 X_{t-1}$ ?

21. Assume that  $X_t$  is a stationary, zero mean, normal time series with autocorrelation function  $\rho_X(h)$ . Let  $Y_t = X_t^2$ . Show that the autocorrelation function of  $Y_t$  is  $\rho_Y(h) = \rho_X^2(h)$ .

22. Let  $\Delta y_t = a_1 + a_2 \cos \omega t$ ,  $t = 1, 2, \dots$ , where  $a_1$ ,  $a_2$ , and  $\omega$  are real numbers, and  $y_0 = 0$ . Find  $y_t$ .

23. Let  $Y_t = e_t - e_{t-1}$ , where the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables. Given  $n$  observations from a realization, predict the  $(n+1)$ st observation in the realization. What is the mean square error of your predictor?  
*Hint:* Consider the time series

$$Z_t = \sum_{j=1}^t Y_j$$

and predict  $Z_{n+1}$ . How would your answer change if  $e_0$  were considered fixed?

24. Prove the following lemma.

**Lemma.** Let  $\lambda > \beta \geq 0$  and let  $p$  be a nonnegative integer. Then there exists an  $M$  such that  $(t+1)^p \beta^t < M\lambda^t$  for all  $t \geq 0$ .

25. Let

$$Y_t - 0.8Y_{t-1} = e_t - 0.9e_{t-4},$$

where the  $e_t$  are uncorrelated  $(0, 1)$  random variables. Using Theorem 2.7.1 obtain  $E\{Y_t e_{t-4}\}$  and  $E\{Y_{t-1} e_{t-4}\}$ . Multiply the equation defining  $Y_t$  by  $Y_t$  and  $Y_{t-1}$ , and take expectations, to obtain a system of equations defining  $\gamma_Y(0)$  and  $\gamma_Y(1)$ . Solve for the variance of  $Y_t$ .

26. Using the backward shift operator, we can write the autoregressive moving average process of order  $(p, q)$  as

$$\left( \sum_{j=0}^p a_j \mathcal{B}^j \right) X_t = \left( \sum_{i=0}^q b_i \mathcal{B}^i \right) e_t.$$

Show that the coefficients of the moving average representation

$$X_t = \sum_{j=0}^{\infty} \psi_j e_{t-j} = \left( \sum_{j=0}^{\infty} \psi_j \mathcal{B}^j \right) e_t,$$

given in Theorem 2.7.1 can be obtained from the quotient  $(\sum_{j=0}^p a_j \mathcal{B}^j)^{-1} (\sum_{i=0}^q b_i \mathcal{B}^i)$  by long division. Obtain the autoregressive representation of Theorem 2.7.2 in the same manner.

27. Demonstrate how the  $n$ th order scalar difference equation  $y_t + a_1 y_{t-1} + a_2 y_{t-2} + \cdots + a_n y_{t-n} = 0$  can be written as a first order vector difference equation of dimension  $n$ . If  $A$  denotes the matrix of coefficients of the resulting first order difference equation, show that the roots of  $|A - mI| = 0$  are the same as the roots of

$$m^n + a_1 m^{n-1} + \cdots + a_n = 0.$$

- 28.** Find  $\Gamma(h)$  for the vector autoregressive time series

$$\begin{aligned} X_{1t} &= 0.9X_{1,t-1} + e_{1t}, \\ X_{2t} &= 0.8X_{1,t-1} + 0.3X_{2,t-1} + e_{2t}, \end{aligned}$$

where  $e_t = (e_{1t}, e_{2t})'$  is a sequence of normal independent  $(\mathbf{0}, \Sigma)$  random variables and

$$\Sigma = \begin{pmatrix} 2 & -1 \\ -1 & 2 \end{pmatrix}.$$

*Hint:* Express the covariance matrix as a linear function of the covariance functions of the two scalar autoregressive processes associated with the canonical form. Assume that we have observations  $\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_n$  with  $\mathbf{X}'_n = (3, 2)$ . Predict  $\mathbf{X}_{n+1}$  and  $\mathbf{X}_{n+2}$  and obtain the covariance matrix of the prediction errors.

- 29.** The cobweb theory has been suggested as a model for the price behavior of some agricultural commodities. The model is composed of a demand equation and a supply equation:

$$\begin{aligned} \text{demand } Q_t &= \alpha_1 + \beta_1 P_t + U_{1t}, \\ \text{supply } Q_t &= \alpha_2 + \beta_2 P_{t-1} + U_{2t}, \end{aligned}$$

where  $Q_t$  is the quantity produced for time  $t$ ,  $P_t$  is the price at time  $t$ , and  $\mathbf{U}' = (U_{1t}, U_{2t})$  is a stationary vector time series. Assuming  $\mathbf{U}_t$  is a sequence of uncorrelated  $(\mathbf{0}, \Sigma)$  vector random variables, express the model as a vector autoregressive process. Under what conditions will the process be stationary? If  $U_{1t} = e_{1t} + b_{11}e_{1,t-1}$  and  $U_{2t} = e_{2t} + b_{22}e_{2,t-1}$  where  $e_t = (e_{1t}, e_{2t})'$  is a sequence of uncorrelated  $(\mathbf{0}, \sigma^2 \mathbf{I})$  vector random variables, how would you describe the vector time series  $(Q_t, P_t)'$ ?

- 30.** Let  $e_t$  be a sequence of independent  $(\mathbf{0}, \sigma^2)$  random variables, and let  $\{a_j\}$  be an absolutely summable sequence of real numbers. Use Theorem 2.2.2 to show that

$$X_t = \sum_{j=0}^{\infty} a_j e_{t-j}$$

is defined as an almost sure limit and  $X_t$  is a stationary process with zero mean and

$$\gamma_X(h) = \sigma^2 \sum_{j=0}^{\infty} a_j a_{j+h}.$$

31. Show that the assumption that  $E\{W_{ii}^2\}$  is bounded below guarantees the existence of the inverse of  $\mathbf{V}_{ii}$  in Theorem 2.9.1 used to define (2.10.3).

32. Let the stationary  $p$ th order autoregressive process satisfy

$$\sum_{j=0}^p a_j X_{t-j} = e_t,$$

where  $a_0 = 1$  and  $e_t$  is a sequence of uncorrelated  $(0, \sigma^2)$  variables. The equation can also be written

$$\Phi(\mathcal{B})X_t = e_t,$$

where  $\mathcal{B}$  is the back shift operator defined in Section 2.4, and

$$\Phi(\mathcal{B}) = 1 + a_1 \mathcal{B} + a_2 \mathcal{B}^2 + \cdots + a_p \mathcal{B}^p.$$

Some authors define a stationary process  $Y_t$  to be *causal* if it is possible to represent  $Y_t$  as

$$Y_t = \sum_{j=0}^{\infty} w_j e_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots,$$

where  $\sum_{j=0}^{\infty} |w_j| < \infty$ . Show that the following statements are equivalent:

- (i)  $X_t$  is causal.
- (ii) The roots of the characteristic equation

$$m^p + a_1 m^{p-1} + \cdots + a_p = 0$$

are less than one in absolute value.

- (iii)  $\Phi(z) \neq 0$  for all complex numbers  $z$  such that  $|z| \leq 1$ .
- (iv) All roots of  $\Phi(z) = 0$  are greater than one in absolute value.

33. Using (2.6.8) and the fact that  $a_p \neq 0$ , show that the autocorrelation function of the  $p$ th order autoregressive process is nonzero for some integer greater than  $N_0$  for all  $N_0$ .

34. Let

$$Y_t = e_0 \cos \pi t + Z_t, \quad t = 1, 2, \dots,$$

where  $e_0 \sim N(0, \sigma_{ee})$ , independent of  $Z_t \sim NI(0, \sigma_{zz})$ . Let  $\sigma_{ee}$  and  $\sigma_{zz}$  be known.

- (a) Given three observations  $(Y_1, Y_2, Y_3)$ , what is the minimum mean square error predictor of  $Y_4$ ? Give the variance of the prediction error.
- (b) What is the limiting expression for the predictor of  $Y_{n+1}$  given  $(Y_1, Y_2, \dots, Y_n)$ , where the limit is taken as  $n$  increases? What is the limit of the variance of the prediction error?

35. Let  $(Y_1, \dots, Y_n)$  be observations on the first order moving average,

$$Y_t = e_t + \beta_1 e_{t-1},$$

where  $e_t \sim NI(0, \sigma^2)$  and  $|\beta_1| < 1$ . Construct the best predictor  $\hat{e}_0(Y_1, Y_2, Y_3)$  of  $e_0$ . What is the variance of the prediction error of  $\hat{e}_0(Y_1, Y_2, Y_3)$ ? What is the limit of the variance of the prediction error of  $\hat{e}_0(Y_1, \dots, Y_n)$  as  $n \rightarrow \infty$ ?

36. Let the stationary vector autoregressive process  $Y_t$  satisfy

$$Y_t + \sum_{j=1}^p A_j Y_{t-j} = e_t,$$

where the  $e_t$  are independent  $(\mathbf{0}, \Sigma_{ee})$  random vectors. Show that the process also satisfies an equation of the form

$$Y_t + \sum_{j=1}^p B_j Y_{t+j} = a_t,$$

where the  $a_t$  are uncorrelated  $(\mathbf{0}, \Sigma_{aa})$  random vectors. Give the equations defining the  $B_j$ ,  $j = 1, 2, \dots, p$ , as a function of the  $\Gamma(h)$ . Show that if  $p = 1$ ,  $A_1 = A'_1$ , and  $\Sigma_{ee} = I$ , then  $\Sigma_{aa} = I$ .

37. Let  $a_t$  be a sequence of uncorrelated  $(0, 1)$  random variables. Put the following time series in canonical form as moving averages of uncorrelated random variables  $e_t$ : (a)  $a_t + 4a_{t-1}$ , (b)  $a_{t-1} + 4a_t + a_{t+1}$ , (c)  $a_t - 2.5a_{t-1} + a_{t-2}$ . Give the variance of  $e_t$  in each case.

38. Let  $V_n$  denote the  $n \times n$  covariance matrix of  $n$  observations  $(Y_1, Y_2, \dots, Y_n)$  on the first order moving average process

$$Y_t = e_t + \beta e_{t-1},$$

where the  $e_t$  are uncorrelated  $(0, 1)$  random variables. Let  $D_n$  be the determinant of  $V_n$ . Show that

$$D_n = (1 + \beta^2)D_{n-1} - \beta^2 D_{n-2}$$

for  $n = 2, 3, \dots$ , with  $D_0 = 1$  and  $D_1 = 1 + \beta^2$ . Hence, show that  $D_n = (1 - \beta^2)^{-1}(1 - \beta^{2(n+1)})$ ,  $n = 1, 2, \dots$ , for  $|\beta| < 1$ , and  $D_n = n + 1$ ,  $n = 1, 2, \dots$ , for  $|\beta| = 1$ .

39. Let  $Y_t$  be a zero mean stationary time series with  $\gamma(0) > 0$ ,  $\lim_{n \rightarrow \infty} \gamma(n) = 0$ , and  $\phi(h) = 0$  for  $h > K$ . Show that  $Y_t$  is nondeterministic.

40. Consider the model  $Y_t = \mathbf{x}'\boldsymbol{\beta} + Z_t$ ,  $t = 1, \dots, n$ , where  $Z_t$  is a process with mean zero and known covariance function  $\text{Cov}(Z_t, Z_j) = V_{tj}$ .

- (a) Show that the best linear unbiased estimator for  $\beta$  is

$$\hat{\beta}_G = (\mathbf{X}' \mathbf{V}_{nn}^{-1} \mathbf{X})^{-1} \mathbf{X}' \mathbf{V}_{nn}^{-1} \mathbf{y},$$

where  $\mathbf{X} = (\mathbf{x}_1, \dots, \mathbf{x}_n)'$ ,  $\mathbf{y} = (Y_1, \dots, Y_n)'$ , and  $\mathbf{V}_{nn} = \mathbf{V}\{\mathbf{y}\}$  is nonsingular.

- (b) Given  $Y_1, \dots, Y_n$ ,  $\mathbf{X}'_{n+s}$ , and  $\mathbf{V}_{ns} = \text{Cov}(\mathbf{y}, Y_{n+s})$ , show that the best linear unbiased predictor of  $Y_{n+s} = \mathbf{x}'_{n+s} \beta + Z_{n+s}$  is

$$\mathbf{x}'_{n+s} \hat{\beta}_G + \mathbf{b}'_{ns} (\mathbf{y} - \mathbf{X} \hat{\beta}_G),$$

where  $\mathbf{b}_{ns} = \mathbf{V}_{nn}^{-1} \mathbf{V}_{ns}$ .

41. Consider a stationary and invertible time series  $Y_t$  given by

$$Y_t = \sum_{j=1}^{\infty} w_j e_{t-j} + e_t = \sum_{j=1}^{\infty} \pi_j Y_{t-j} + e_t,$$

where  $\sum_{j=1}^{\infty} |w_j| < \infty$ ,  $\sum_{j=1}^{\infty} |\pi_j| < \infty$ , and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables. Show that

$$\lim_{n \rightarrow \infty} E \left[ \left( \hat{Y}_{t,n} - \sum_{j=1}^n \pi_j Y_{t-j} \right)^2 \right] = 0,$$

where  $\hat{Y}_{t,n} = \hat{Y}_t(Y_{t-1}, \dots, Y_{t-n})$  is the best linear predictor of  $Y_t$  given  $Y_{t-1}, \dots, Y_{t-n}$ .

42. In Theorem 2.9.1, the fact that the system of equations  $\mathbf{V}_{nn} \mathbf{b} = \mathbf{V}_{ns}$  is consistent is used. Show that this system is always consistent and that  $\mathbf{y}' \mathbf{b}_{ns}$  in (2.9.2) is unique a.e. for different choices of generalized inverses of  $\mathbf{V}_{nn}$ .

43. Consider a sequence satisfying  $e_t = Z_t(\beta_0 + \beta_1 e_{t-1}^2)^{1/2}$ , where  $Z_t \sim NI(0, 1)$ ,  $\beta_0 > 0$  and  $0 \leq \beta_1 < 1$ . That is, the distribution of  $e_t$  given the past is  $N(0, h_t)$ , where  $h_t = \beta_0 + \beta_1 e_{t-1}^2$ . Thus, the conditional variance depends on the past errors. The model

$$Y_t = \alpha Y_{t-1} + e_t$$

is a special case of autoregressive conditionally heteroscedastic models called an ARCH(1) model. See Engle (1982). Note that

$$\begin{aligned} e_t^2 &= Z_t^2(\beta_0 + \beta_1 e_{t-1}^2) = Z_t^2[\beta_0 + \beta_1 Z_{t-1}^2(\beta_0 + \beta_1 e_{t-2}^2)] \cdots \\ &= \beta_0 \sum_{j=0}^{\infty} \beta_1^j \left( \prod_{i=0}^{j-1} Z_{t-i}^2 \right) \quad \text{a.s.}, \end{aligned}$$

where it is assumed that the process started with finite variance in the indefinite past.

- (a) Show that  $\{e_t\}$  is a sequence of uncorrelated  $[0, (1 - \beta_1)^{-1} \beta_0]$  random variables.
- (b) Show that if  $3\beta_1^2 < 1$ , then  $E\{e_t^4\}$  exists. Find  $E\{e_t^4\}$ .
- (c) Consider  $X_t = e_t^2$ . Assuming  $3\beta_1^2 < 1$ , show that  $X_t$  is stationary. Give its autocorrelation function.
- (d) Consider the stationary ARCH(1) model,  $Y_t = \alpha_1 Y_{t-1} + e_t$ , where  $|\alpha_1| < 1$ . Assume  $(\alpha_1, \beta_0, \beta_1)$  are known.
- What is the best predictor  $\hat{Y}_{n+s}$  for  $Y_{n+s}$  given  $Y_1, \dots, Y_n$ ?
  - Find  $V\{Y_{n+s} - \hat{Y}_{n+s}\}$ , the unconditional forecast error variance.
  - Find  $V\{(Y_{n+s} - \hat{Y}_{n+s}) | (e_1, \dots, e_n)\}$ , the conditional forecast error variance.
  - Show that  $V\{(Y_{n+2} - \hat{Y}_{n+2}) | (e_1, \dots, e_n)\}$  may be less than  $V\{(Y_{n+1} - \hat{Y}_{n+1}) | (e_1, \dots, e_n)\}$  for some  $\beta_0, \beta_1, \alpha_1$ , and  $e_n^2$ . (That is, the conditional forecast error variance for two-step-ahead forecasting may be less than that for one-step-ahead forecasting.)
  - Show that

$$\lim_{s \rightarrow \infty} V\{Y_{n+s} - \hat{Y}_{n+s} | (e_1, \dots, e_n)\} = (1 - \alpha_1^2)^{-1} \sigma^2 \quad \text{a.s.,}$$

$$\text{where } \sigma^2 = (1 - \beta_1)^{-1} \beta_0.$$

44. Use the facts that, for  $-0.5 < d < 0.5$  and  $h > 0$ ,

$$2^{d-1} i^{d-1} < i^{d-1} (1 + h^{-1} i)^{d-1} < i^{d-1} \quad \text{for } 1 \leq i \leq h,$$

$$(i + h)^{2d-2} < i^{d-1} (i + h)^{d-1} < 2^{d-1} i^{2d-2} \quad \text{for } i > h$$

to show that  $\gamma_r(h)$  of (2.11.3) is bounded above by  $c_1 h^{2d-1}$  and bounded below by  $c_2 h^{2d-1}$ , where  $c_1$  and  $c_2$  are constants.

45. Let  $\{a_j\}_{-\infty}^{\infty}$  and  $\{b_j\}_{-\infty}^{\infty}$  be sequences of real numbers satisfying

$$\sum_{j=-\infty}^{\infty} |a_j| = M < \infty \quad \text{and} \quad \sum_{j=-\infty}^{\infty} b_j^2 < \infty.$$

Show that  $\sum_{r=-\infty}^{\infty} c_r^2 < \infty$ , where

$$c_r = \sum_{i=-\infty}^{\infty} b_i a_{r-i} = \sum_{j=-\infty}^{\infty} a_r b_{r-j}.$$

## CHAPTER 3

## Introduction to Fourier Analysis

## 3.1. SYSTEMS OF ORTHOGONAL FUNCTIONS—FOURIER COEFFICIENTS

In many areas of applied mathematics it is convenient to approximate a function by a linear combination of elementary functions. The reader is familiar with the Weierstrass theorem, which states that any continuous function on a compact set may be approximated by a polynomial. Likewise, it may be convenient to construct a set of vectors called a *basis* such that all other vectors may be expressed as linear combinations of the elements of the basis. Often the basis vectors are constructed to be orthogonal, that is, the sum of the products of the elements of any pair is zero. The system that is of particular interest to us in the analysis of time series is the system of trigonometric polynomials.

Assume that we have a function defined on a finite number of points,  $N$ . We shall investigate the properties of the set of functions  $\{\cos(2\pi mt/N), \sin(2\pi mt/N)\}$ :  $t = 0, 1, \dots, N - 1$  and  $m = 0, 1, \dots, L[N]$ , where  $L[N]$  is the largest integer less than or equal to  $N/2$ . The reader should observe that the points have been indexed by  $t$  running from zero to  $N - 1$ , while  $m$  runs from zero to  $L[N]$ . Note that for  $m = 0$  the cosine is identically equal to 1. The sine function is identically zero for  $m = 0$  and for  $m = N/2$  if  $N$  is even. Therefore, it is to be understood that these sine functions are not included in a discussion of the collection of functions. The collection of interest will always contain exactly  $N$  functions, none of which is identically zero. We shall demonstrate that these functions, defined on the integers  $0, 1, \dots, N - 1$ , are orthogonal, and we shall derive the sum of squares for each function. This constitutes a proof that the  $N$  functions defined on the integers furnish an orthogonal basis for the  $N$ -dimensional vector space.

**Theorem 3.1.1.** Given that  $m$  and  $r$  are contained in the set  $\{0, 1, 2, \dots, L[N]\}$ , then

$$\sum_{t=0}^{N-1} \cos \frac{2\pi m}{N} t \cos \frac{2\pi r}{N} t = \begin{cases} N, & m = r = 0 \text{ or } \frac{N}{2}, \\ \frac{N}{2}, & m = r \neq 0 \text{ or } \frac{N}{2}, \\ 0, & m \neq r; \end{cases}$$

$$\sum_{t=0}^{N-1} \sin \frac{2\pi m}{N} t \cos \frac{2\pi r}{N} t = 0, \quad \forall m, r;$$

$$\sum_{t=0}^{N-1} \sin \frac{2\pi m}{N} t \sin \frac{2\pi r}{N} t = \begin{cases} \frac{N}{2}, & m = r \neq 0 \text{ or } \frac{N}{2}, \\ 0, & m \neq r. \end{cases}$$

**Proof.** Consider first the sum of the products of two cosine functions, and let

$$\begin{aligned} S(m, r) &= \sum_{t=0}^{N-1} \cos \frac{2\pi m}{N} t \cos \frac{2\pi r}{N} t \\ &= \frac{1}{2} \sum_{t=0}^{N-1} \left[ \cos \frac{2\pi t}{N} (m+r) + \cos \frac{2\pi t}{N} (m-r) \right]. \end{aligned} \quad (3.1.1)$$

For  $m = r = 0$  or  $m = r = N/2$  ( $N$  even) the cosine terms on the right-hand side of (3.1.1) are always equal to one, and we have

$$S(m, r) = \frac{1}{2} \sum_{t=0}^{N-1} (1+1) = N.$$

For  $m = r$ , but not equal to zero and not equal to  $N/2$  if  $N$  is even, the sum (3.1.1) reduces to

$$S(m, r) = \frac{1}{2} \sum_{t=0}^{N-1} \cos \frac{2\pi t}{N} 2m + \frac{1}{2} N,$$

where the summation of cosines is given by

$$\sum_{t=0}^{N-1} \cos \frac{2\pi t}{N} 2m = \frac{1}{2} \sum_{t=0}^{N-1} [e^{\epsilon(2m)2\pi t/N} + e^{-\epsilon(2m)2\pi t/N}].$$

The two sums are geometric series whose rates are  $\exp\{\epsilon(2m)2\pi/N\}$  and  $\exp\{-\epsilon(2m)2\pi/N\}$ , respectively, and the first term is one in both cases. Now  $\epsilon(2m)2\pi/N$  is not an integer multiple of  $2\pi$ ; therefore the rates are not unity. Applying the well-known formula  $\sum_{t=0}^{N-1} \lambda^t = (1 - \lambda^N)/(1 - \lambda)$ , the partial sum is

$$\frac{1 - (e^{\epsilon(2m)2\pi/N})^N}{1 - e^{\epsilon(2m)2\pi/N}} + \frac{1 - (e^{-\epsilon(2m)2\pi/N})^N}{1 - e^{-\epsilon(2m)2\pi/N}}.$$

Since  $\exp\{\epsilon(2m)2\pi\} = \exp\{\epsilon 2\pi\} = 1$ , the numerators of the partial sums are 0,

and  $S(m, r)$  reduces to  $N/2$ . For  $m \neq r$ , we have

$$\begin{aligned} S(m, r) &= \frac{1}{2} \sum_{t=0}^{N-1} \left[ \cos \frac{2\pi t}{N} (m+r) + \cos \frac{2\pi t}{N} (m-r) \right] \\ &= \frac{1}{4} \sum_{t=0}^{N-1} \left[ e^{i(2\pi t/N)(m+r)} + e^{-i(2\pi t/N)(m+r)} \right. \\ &\quad \left. + e^{i(2\pi t/N)(m-r)} + e^{-i(2\pi t/N)(m-r)} \right] \\ &= 0. \end{aligned}$$

The sum of products of a sine and cosine function is given by

$$\begin{aligned} \sum_{t=0}^{N-1} \sin \frac{2\pi m}{N} t \cos \frac{2\pi r}{N} t \\ = \frac{1}{2} \sum_{t=0}^{N-1} \left[ \sin(m+r) \frac{2\pi t}{N} + \sin(m-r) \frac{2\pi t}{N} \right] = 0 \quad \forall m, r, \end{aligned}$$

where the proof follows the same pattern as that for the product of cosines with  $m = r \neq 0$ . We leave the details to the reader.  $\blacktriangle$

Having demonstrated that the  $N$  functions form an orthogonal basis, it follows that any function  $f(t)$  defined on  $N$  integers can be represented by

$$f(t) = \sum_{m=0}^{L[N]} (a_m \cos \omega_m t + b_m \sin \omega_m t), \quad t = 0, 1, \dots, N-1, \quad (3.1.2)$$

where

$$\omega_m = \frac{2\pi m}{N}, \quad m = 0, 1, 2, \dots, L[N];$$

$$a_m = \begin{cases} \frac{2}{N} \sum_{t=0}^{N-1} f(t) \cos \omega_m t, & m = 1, 2, \dots, L[N-1], \\ \frac{\sum_{t=0}^{N-1} f(t) \cos \omega_m t}{N}, & m = 0, \text{ and } m = \frac{N}{2} \text{ if } N \text{ is even}; \end{cases}$$

$$b_m = \frac{2}{N} \sum_{t=0}^{N-1} f(t) \sin \omega_m t, \quad m = 1, 2, \dots, L[N-1].$$

The  $a_m$  and  $b_m$  are called *Fourier coefficients*. One way to obtain the

representation (3.1.2) is to find the  $a_m$  and  $b_m$  such that

$$\sum_{t=0}^{N-1} \left\{ f(t) - \sum_{m=0}^{L(N)} (a_m \cos \omega_m t + b_m \sin \omega_m t) \right\}^2 \quad (3.1.3)$$

is a minimum. Differentiating with respect to the  $a_j$  and  $b_j$  and setting the derivatives equal to zero, we obtain

$$\sum_{t=0}^{N-1} \left\{ f(t) - \sum_{m=0}^{L(N)} (a_m \cos \omega_m t + b_m \sin \omega_m t) \right\} \cos \omega_j t = 0, \\ j = 0, 1, 2, \dots, L[N],$$

$$\sum_{t=0}^{N-1} \left\{ f(t) - \sum_{m=0}^{L(N)} (a_m \cos \omega_m t + b_m \sin \omega_m t) \right\} \sin \omega_j t = 0, \\ j = 1, 2, \dots, L[N-1].$$

By the results of Theorem 3.1.1, these equations reduce to

$$\sum_{t=0}^{N-1} f(t) \cos \omega_m t = a_m \sum_{t=0}^{N-1} \cos^2 \omega_m t, \quad m = 0, 1, \dots, L[N],$$

$$\sum_{t=0}^{N-1} f(t) \sin \omega_m t = b_m \sum_{t=0}^{N-1} \sin^2 \omega_m t, \quad m = 1, \dots, L[N-1],$$

and we have the coefficients of (3.1.2). Thus we see that the coefficients are the regression coefficients obtained by regressing the vector  $f(t)$  on the vectors  $\cos \omega_m t$  and  $\sin \omega_m t$ . By the orthogonality of the functions, the multiple regression coefficients are the simple regression coefficients.

Because of the different sum of squares for  $\cos 0t$  and  $\cos \pi t$ , the  $a$ 's are sometimes all defined with a common divisor,  $N/2$ , and the first and last terms of the series modified accordingly. Often  $N$  is restricted to be odd to simplify the discussion. Specifically, for  $N$  odd, we have

$$f(t) = \frac{a_0}{2} + \sum_{m=1}^{(N-1)/2} (a_m \cos \omega_m t + b_m \sin \omega_m t), \quad t = 0, 1, \dots, N-1, \quad (3.1.4)$$

where

$$a_m = \frac{2 \sum_{t=0}^{N-1} f(t) \cos \omega_m t}{N}, \quad m = 0, 1, \dots, \frac{N-1}{2},$$

$$b_m = \frac{2 \sum_{t=0}^{N-1} f(t) \sin \omega_m t}{N}, \quad m = 1, 2, \dots, \frac{N-1}{2}.$$

The reader will have no difficulty in identifying the definitions being used. If the leading term is given as  $a_0/2$ , the definition (3.1.4) is being used; if not, (3.1.2) is being used.

The preceding material demonstrates that any finite vector can be represented as a linear combination of vectors defined by the sine and cosine functions. We now turn to the investigation of similar representations for functions defined on the real line.

We first consider functions defined on a finite interval of the real line. Since the interval is finite, we may code the end points in any convenient manner. When dealing with trigonometric functions, it will be most convenient to treat intervals whose length is a multiple of  $2\pi$ . We shall most frequently take the interval to be  $[-\pi, \pi]$ .

**Definition 3.1.1.** An infinite system of square integrable functions  $\{\varphi_j\}_{j=0}^{\infty}$  is orthogonal on  $[a, b]$  if

$$\int_a^b \varphi_j(x) \varphi_m(x) dx = 0, \quad j \neq m, \quad j, m = 0, 1, \dots,$$

and

$$\int_a^b |\varphi_j(x)|^2 dx \neq 0, \quad j = 0, 1, \dots.$$

The following theorem states that the system of trigonometric functions is orthogonal on the interval  $[-\pi, \pi]$ .

**Theorem 3.1.2.** Given that  $m$  and  $j$  are nonnegative integers, then

$$\int_{-\pi}^{\pi} \sin mx \cos jx dx = 0 \quad \forall m, j;$$

$$\int_{-\pi}^{\pi} \sin mx \sin jx dx = \begin{cases} 0, & m \neq j, \\ \pi, & m = j \neq 0, \\ 0, & m = j = 0; \end{cases}$$

$$\int_{-\pi}^{\pi} \cos mx \cos jx dx = \begin{cases} 2\pi, & m = j = 0, \\ \pi, & m = j \neq 0, \\ 0, & m \neq j. \end{cases}$$

**Proof.** Reserved for the reader. ▲

The sum

$$S_n(x) = \sum_{k=0}^n (a_k \cos \lambda_k x + b_k \sin \lambda_k x), \quad (3.1.5)$$

$$\lambda_k = \frac{2\pi k}{T}, \quad k = 0, 1, \dots, n,$$

is called a *trigonometric polynomial* of order (or degree)  $n$  and period  $T$ . We shall take  $T = 2\pi$  (i.e.,  $\lambda_k = k$ ) in the sequel unless stated otherwise. If we let  $n$  increase without bound, we have the infinite trigonometric series

$$S_\infty(x) = \sum_{k=0}^{\infty} (a_k \cos \lambda_k x + b_k \sin \lambda_k x). \quad (3.1.6)$$

By Theorem 3.1.2 we know that the sine and cosine functions are orthogonal on the interval  $[-\pi, \pi]$ , or on any interval of length  $2\pi$ . We shall investigate the ability of a trigonometric polynomial to approximate a function defined on this interval.

In the finite theory of least squares we are given a vector  $y$  of dimension  $n$ , which we desire to approximate by a linear combination of the vectors  $x_k$ ,  $k = 1, 2, \dots, p$ . The coefficients of the predicting equation

$$\hat{Y}_j = \sum_{k=1}^p b_k x_{kj}$$

are obtained by minimizing

$$\sum_{j=1}^n \left( Y_j - \sum_{k=1}^p b_k x_{kj} \right)^2$$

with respect to the  $b_k$ .

The analogous procedure for a function  $f(x)$  defined on the interval  $[-\pi, \pi]$  is to minimize the integral

$$\int_{-\pi}^{\pi} \left[ f(x) - \sum_{k=0}^n (a_k \cos \lambda_k x + b_k \sin \lambda_k x) \right]^2 dx. \quad (3.1.7)$$

Of course,  $[f(x) - \sum_{k=0}^n (a_k \cos \lambda_k x + b_k \sin \lambda_k x)]^2$  must be integrable. In this section we shall say that a function  $g(x)$  is integrable over  $[a, b]$  if it is continuous, or if it has a finite number of discontinuities [at which  $g(x)$  can be either bounded or unbounded], provided the improper Riemann integral exists. More general definitions could be used. Our treatment in the sequel follows closely that of Tolstov (1962), and we use his definitions.

The reader may verify by expansion of the square and differentiation of the

resultant products that the coefficients that minimize (3.1.7) are

$$a_k = \frac{1}{\pi} \int_{-\pi}^{\pi} \cos \lambda_k x f(x) dx, \quad k = 0, 1, 2, \dots, \quad (3.1.8)$$

$$b_k = \frac{1}{\pi} \int_{-\pi}^{\pi} \sin \lambda_k x f(x) dx, \quad k = 1, 2, \dots, \quad (3.1.9)$$

where the first term is written as  $a_0/2$ . In this form the approximating sum is

$$S_n(x) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos \lambda_k x + b_k \sin \lambda_k x). \quad (3.1.10)$$

When  $T$ , the length of the interval, is not equal to  $2\pi$ ,  $\lambda_k$  is set equal to  $2\pi k/T$ , and we obtain the formulas

$$\begin{aligned} a_k &= \frac{2}{T} \int_{-T/2}^{T/2} \cos \lambda_k x f(x) dx, \quad k = 0, 1, 2, \dots, \\ b_k &= \frac{2}{T} \int_{-T/2}^{T/2} \sin \lambda_k x f(x) dx, \quad k = 1, 2, \dots, \\ S_n(x) &= \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos \lambda_k x + b_k \sin \lambda_k x). \end{aligned} \quad (3.1.11)$$

Observe that  $\pi[(a_0^2/2) + \sum_{k=1}^n (a_k^2 + b_k^2)]$  is completely analogous to the sum of squares due to regression on  $n$  orthogonal independent variables of finite least squares theory. We repeat below the theorem analogous to the statement in finite least squares that the multiple correlation coefficient is less than or equal to one.

**Theorem 3.1.3 (Bessel's inequality).** Let  $a_k$ ,  $b_k$ , and  $S_n(x)$  be defined by (3.1.8), (3.1.9), and (3.1.10), respectively. If  $f(x)$  defined on  $[-\pi, \pi]$  is square integrable, then

$$\frac{1}{\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx \geq \frac{a_0^2}{2} + \sum_{k=1}^n (a_k^2 + b_k^2). \quad (3.1.12)$$

**Proof.** Since the square is always nonnegative,

$$\begin{aligned} 0 &\leq \int_{-\pi}^{\pi} |f(x) - S_n(x)|^2 dx = \int_{-\pi}^{\pi} [f^2(x) - 2f(x)S_n(x) + S_n^2(x)] dx \\ &= \int_{-\pi}^{\pi} f^2(x) dx - 2 \int_{-\pi}^{\pi} f(x)S_n(x) dx + \int_{-\pi}^{\pi} S_n^2(x) dx. \end{aligned}$$

By the definitions of  $a_k$  and  $b_k$  and the orthogonality of the sine and cosine

functions, we have

$$\begin{aligned} \int_{-\pi}^{\pi} |f(x) - S_n(x)|^2 dx &= \int_{-\pi}^{\pi} f^2(x) dx - 2 \left[ \frac{a_0^2}{2} \pi + \pi \sum_{k=1}^n (a_k^2 + b_k^2) \right] \\ &\quad + \left[ \frac{a_0^2}{2} \pi + \pi \sum_{k=1}^n (a_k^2 + b_k^2) \right] \end{aligned}$$

and

$$\frac{1}{\pi} \int_{-\pi}^{\pi} f^2(x) dx \geq \frac{a_0^2}{2} + \sum_{k=1}^n (a_k^2 + b_k^2). \quad \blacktriangle$$

**Definition 3.1.2.** A system of functions  $\{\varphi_j(x)\}_{j=0}^{\infty}$  is *complete* if there does not exist a function  $f(x)$  such that

$$\int_a^b |f(x)| dx \neq 0$$

and

$$\int_a^b f(x) \varphi_j(x) dx = 0, \quad j = 0, 1, \dots$$

We shall give only a few of the theorems of Fourier analysis. The following two theorems constitute a proof of completeness of the trigonometric functions. Lebesgue's proofs are given in Rogosinski (1959).

**Theorem 3.1.4.** If  $f(x)$  is continuous on the interval  $[-\pi, \pi]$ , then all the Fourier coefficients are zero if and only if

$$f(x) = 0.$$

**Proof.** Omitted. ▲

**Theorem 3.1.5.** If  $f(x)$  defined on the interval  $[-\pi, \pi]$  is integrable and if all the Fourier coefficients  $\{a_k, b_k: k = 0, 1, 2, \dots\}$  are zero, then

$$\int_{-\pi}^{\pi} |f(x)| dx = 0.$$

**Proof.** Omitted. ▲

By Theorem 3.1.4 the Fourier coefficients of a continuous function are unique. Theorem 3.1.5 generalizes the result to absolutely integrable functions. Neither furnishes information on the nature of the convergence of the sequences  $\{a_k\}$  and  $\{b_k\}$ .

Theorem 3.1.6 (Parseval's theorem) answers the question of convergence for square integrable functions.

**Theorem 3.1.6.** If  $f(x)$  defined on  $[-\pi, \pi]$  is square integrable, then

$$\lim_{n \rightarrow \infty} \int_{-\pi}^{\pi} \left| f(x) - \frac{a_0}{2} - \sum_{k=1}^n (a_k \cos kx + b_k \sin kx) \right|^2 dx = 0$$

or, equivalently,

$$\frac{a_0^2}{2} + \sum_{k=1}^{\infty} (a_k^2 + b_k^2) = \frac{1}{\pi} \int_{-\pi}^{\pi} |f(x)|^2 dx .$$

**Proof.** By Bessel's inequality, the partial sum of squares of the Fourier coefficients converges for any square integrable function. (The sum of squares is monotone increasing and bounded.) Therefore, the function

$$g(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) ,$$

where  $a_k$  and  $b_k$  are the Fourier coefficients of a square integrable function  $f(x)$ , is square integrable on  $[-\pi, \pi]$ . Now  $D(x) = f(x) - g(x)$  is square integrable and all Fourier coefficients of  $D(x)$  are zero; hence, by Theorem 3.1.5,

$$\int_{-\pi}^{\pi} |f(x) - g(x)|^2 dx = \int_{-\pi}^{\pi} |f(x) - g(x)|^2 dx = 0 .$$

▲

Thus, any square integrable function defined on an interval can be approximated in mean square by a trigonometric polynomial.

As one might suspect, the convergence of the sequence of functions  $S_n(x)$  at a point  $x_0$  requires additional conditions. We shall closely follow Tolstov's (1962) text in presenting a proof of the pointwise convergence of  $S_n(x_0)$  to  $f(x_0)$ . The following definitions are needed.

**Definition 3.1.3.** The *left limit* of  $f(x)$  is given by

$$\lim_{\substack{x \rightarrow x_0 \\ x < x_0}} f(x) = f(x_0^-)$$

provided the limit exists and is finite. The limit of  $f(x)$  as  $x \rightarrow x_0$ ,  $x > x_0$ , is called the *right limit* of  $f(x)$  and is denoted by  $f(x_0^+)$ , provided it exists and is finite.

**Definition 3.1.4.** The *right derivative* of  $f(x)$  at  $x = x_0$  is defined by

$$f'(x_0^+) = \lim_{\substack{h \rightarrow 0 \\ h > 0}} \frac{f(x_0 + h) - f(x_0^+)}{h} ,$$

and the *left derivative* by

$$f'(x_0^-) = \lim_{\substack{h \rightarrow 0 \\ h > 0}} \frac{f(x_0 - h) - f(x_0^-)}{h},$$

provided the limits exist and are finite.

**Definition 3.1.5.** The function  $f(x)$  has a *jump discontinuity* at the point  $x_0$  if  $f(x_0^+) \neq f(x_0^-)$ .

From the definitions it is clear that the jump  $f(x_0^+) - f(x_0^-)$  is finite.

**Definition 3.1.6.** A function  $f(x)$  is *smooth* on the interval  $[c, d]$  if it has a continuous derivative on the interval  $[c, d]$ .

**Definition 3.1.7.** A function  $f(x)$  is *piecewise smooth* on the interval  $[c, d]$  if

- (i)  $f(x)$  and  $f'(x)$  are both continuous, or
- (ii)  $f(x)$  and  $f'(x)$  have a finite number of jump discontinuities.

Bessel's inequality guarantees that the Fourier coefficients of any square integrable function go to zero as the frequency increases; that is,

$$\lim_{m \rightarrow \infty} a_m = \lim_{m \rightarrow \infty} b_m = 0.$$

The following lemma is a proof of the same result for a different class of functions.

**Lemma 3.1.1.** If  $f(x)$  is a piecewise smooth function on  $[c, d]$ , then

$$\lim_{m \rightarrow \infty} \int_c^d f(x) \cos mx \, dx = \lim_{m \rightarrow \infty} \int_c^d f(x) \sin mx \, dx = 0.$$

**Proof.** Integrating by parts gives

$$\int_c^d f(x) \cos mx \, dx = \frac{1}{m} \left\{ [f(x) \sin mx]_c^d - \int_c^d f'(x) \sin mx \, dx \right\}.$$

Since both  $f(x)$  and  $f'(x)$  contain at most a finite number of jump discontinuities on the interval  $[c, d]$ , they are both bounded. Therefore, the quantity contained within the curly braces is bounded, and the result follows immediately. ▲

Using the fact that any absolutely integrable function can be approximated in the mean by a piecewise smooth function, it is possible to extend this result to any absolutely integrable function. We state this generalization without proof.

**Lemma 3.1.1A.** The Fourier coefficients  $a_m$  and  $b_m$  of an absolutely integrable function defined on a finite interval approach zero as  $m \rightarrow \infty$ .

Lemmas 3.1.2 and 3.1.3 are presented as preliminaries to the proof of Theorem 3.1.7.

**Lemma 3.1.2.** Define

$$A_n(u) = \frac{1}{2} + \sum_{j=1}^n \cos ju .$$

Then

$$A_n(u) = \frac{\sin(n + \frac{1}{2})u}{2 \sin(u/2)} \quad (3.1.13)$$

and

$$\frac{1}{\pi} \int_0^\pi A_n(u) du = \frac{1}{\pi} \int_{-\pi}^0 A_n(u) du = \frac{1}{2} . \quad (3.1.14)$$

**Proof.** Multiplying both sides of the definition of  $A_n(u)$  by  $2 \sin(u/2)$ , we have

$$\begin{aligned} 2A_n(u)\sin(u/2) &= \sin(u/2) + 2 \sum_{j=1}^n \cos ju \sin(u/2) \\ &= \sin(u/2) + \sum_{j=1}^n [\sin(j + \frac{1}{2})u - \sin(j - \frac{1}{2})u] \\ &= \sin(n + \frac{1}{2})u , \end{aligned}$$

and the result (3.1.13) follows. Since the integral of  $\cos ju$ ,  $j = 1, 2, \dots$ , is zero on an interval of length  $2\pi$ ,

$$\int_{-\pi}^\pi \left[ \frac{1}{2} + \sum_{j=1}^n \cos ju \right] du = \pi ,$$

and hence

$$\frac{1}{\pi} \int_{-\pi}^\pi \frac{\sin(n + \frac{1}{2})u}{2 \sin(u/2)} du = 1 . \quad (3.1.15)$$

The result (3.1.14) follows, since  $A_n(u)$  is an even function. ▲

**Lemma 3.1.3.** The partial sum

$$S_n(x) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx)$$

may be expressed in the form

$$S_n(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x+u) \frac{\sin(n+\frac{1}{2})u}{2 \sin(u/2)} du, \quad (3.1.16)$$

where  $a_k$  and  $b_k$  are the Fourier coefficients of a periodic function  $f(x)$  of period  $2\pi$  defined in (3.1.8) and (3.1.9).

**Proof.** Substituting (3.1.8) and (3.1.9) into (3.1.10), we have

$$\begin{aligned} S_n(x) &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(t) dt \\ &\quad + \frac{1}{\pi} \sum_{k=1}^n \left[ \left\{ \int_{-\pi}^{\pi} f(t) \cos kt dt \right\} \cos kx + \left\{ \int_{-\pi}^{\pi} f(t) \sin kt dt \right\} \sin kx \right] \\ &= \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \left[ \frac{1}{2} + \sum_{k=1}^n \cos k(t-x) \right] dt. \end{aligned}$$

Using (3.1.13) of Lemma 3.1.2, we have

$$S_n(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(t) \frac{\sin[(n+\frac{1}{2})(t-x)]}{2 \sin(\frac{1}{2}(t-x))} dt.$$

Setting  $u = t - x$  and noting that the integral of a periodic function of period  $2\pi$  over the interval  $[-\pi - x, \pi - x]$  is the same as the integral over  $[-\pi, \pi]$ , we have

$$S_n(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x+u) \frac{\sin(n+\frac{1}{2})u}{2 \sin(u/2)} du. \quad \blacktriangle$$

Equation (3.1.16) is called the *integral formula* for the partial sum of a Fourier series.

**Theorem 3.1.7.** Let  $f(x)$  be an absolutely integrable function of period  $2\pi$ . Then:

- (i) At a point of continuity where  $f(x)$  has a right derivative and a left derivative,

$$f(x) = \frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx).$$

(ii) At every point of discontinuity where  $f(x)$  has a right and a left derivative,

$$\frac{a_0}{2} + \sum_{k=1}^{\infty} (a_k \cos kx + b_k \sin kx) = \frac{f(x^+) + f(x^-)}{2}.$$

**Proof.** Consider first a point of continuity. We wish to show that the difference

$$\lim_{n \rightarrow \infty} \frac{1}{\pi} \int_{-\pi}^{\pi} f(x+u) \frac{\sin(n + \frac{1}{2})u}{2 \sin(u/2)} du - f(x) \quad (3.1.17)$$

equals zero. Multiplying (3.1.15) of Lemma 3.1.2 by  $f(x)$ , we have

$$f(x) = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \frac{\sin(n + \frac{1}{2})u}{2 \sin(u/2)} du. \quad (3.1.18)$$

Therefore, the difference (3.1.17) can be written as

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{1}{\pi} \int_{-\pi}^{\pi} [f(x+u) - f(x)] \frac{\sin(n + \frac{1}{2})u}{2 \sin(u/2)} du \\ &= \lim_{n \rightarrow \infty} \frac{1}{\pi} \int_{-\pi}^{\pi} \left( \frac{f(x+u) - f(x)}{u} \right) \left( \frac{u}{2 \sin(u/2)} \right) \sin(n + \frac{1}{2})u du. \end{aligned}$$

Now,  $u^{-1}[f(x+u) - f(x)]$  is absolutely integrable, since the existence of the left and right derivatives means that the ratio is bounded as  $u$  approaches zero. Also,  $[2 \sin(u/2)]^{-1}u$  is bounded. Therefore,

$$\frac{f(x+u) - f(x)}{u} \left( \frac{u}{2 \sin(u/2)} \right) \stackrel{\text{(say)}}{=} g(u)$$

is absolutely integrable. Hence, by Lemma 3.1.1A,

$$\lim_{n \rightarrow \infty} \frac{1}{\pi} \int_{-\pi}^{\pi} g(u) \sin(n + \frac{1}{2})u du = 0,$$

giving us the desired result for a point of continuity.

For a point of jump discontinuity we must prove

$$\lim_{n \rightarrow \infty} \frac{1}{\pi} \int_{-\pi}^{\pi} f(x+u) \frac{\sin(n + \frac{1}{2})u}{2 \sin(u/2)} du = \frac{f(x^+) + f(x^-)}{2}.$$

Using (3.1.14) of Lemma 3.1.2, we have

$$\frac{f(x^+)}{2} = \frac{1}{\pi} \int_0^{\pi} f(x^+) \frac{\sin(n + \frac{1}{2})u}{2 \sin(u/2)} du$$

and

$$\frac{f(x^-)}{2} = \frac{1}{\pi} \int_{-\pi}^0 f(x^-) \frac{\sin(n + \frac{1}{2})u}{2 \sin(u/2)} du.$$

The same arguments on boundedness and integrability that were used for a point of continuity can be used to show that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{\pi} \int_0^\pi \frac{f(x+u) - f(x^-)}{u} \frac{u}{2 \sin(u/2)} \sin(n + \frac{1}{2})u du \\ = \lim_{n \rightarrow \infty} \frac{1}{\pi} \int_{-\pi}^0 \frac{f(x+u) - f(x^-)}{u} \frac{u}{2 \sin(u/2)} \sin(n + \frac{1}{2})u du \\ = 0. \end{aligned}$$
▲

**Theorem 3.1.8.** Let  $f(x)$  be a continuous periodic function of period  $2\pi$  with derivative  $f'(x)$  that is square integrable. Then the Fourier series of  $f(x)$  converges to  $f(x)$  absolutely and uniformly.

**Proof.** Under the assumptions we may integrate by parts to obtain

$$\begin{aligned} a_h &= \frac{1}{\pi} \int_{-\pi}^\pi f(x) \cos hx dx \\ &= \frac{1}{\pi h} [f(x) \sin hx]_{-\pi}^\pi - \frac{1}{\pi h} \int_{-\pi}^\pi f'(x) \sin hx dx \\ &= -\frac{1}{\pi h} \int_{-\pi}^\pi f'(x) \sin hx dx. \end{aligned}$$

Therefore, the Fourier coefficients of the function are directly related to the Fourier coefficients of the derivative by  $a_h = -b'_h/h$ ,  $b_h = a'_h/h$ , where

$$a'_h = \frac{1}{\pi} \int_{-\pi}^\pi f'(x) \cos hx dx,$$

$$b'_h = \frac{1}{\pi} \int_{-\pi}^\pi f'(x) \sin hx dx.$$

The Fourier coefficients of the derivative are well defined by the assumption that  $f'(x)$  is square integrable. Furthermore, by Theorem 3.1.6 (Parseval's theorem), the series

$$\sum_{h=1}^{\infty} (|a'_h|^2 + |b'_h|^2)$$

converges. Since

$$\left(|a'_h| - \frac{1}{h}\right)^2 = |a'_h|^2 - \frac{2}{h}|a'_h| + \frac{1}{h^2} \geq 0$$

we have

$$\frac{1}{h}|a'_h| + \frac{1}{h}|b'_h| \leq \frac{1}{2}(|a'_h|^2 + |b'_h|^2) + \frac{1}{h^2}.$$

It follows that  $\sum_{h=1}^{\infty} (|a_h| + |b_h|)$  converges. Now,

$$\begin{aligned} |a_h \cos hx + b_h \sin hx| &\leq |a_h \cos hx| + |b_h \sin hx| \\ &\leq |a_h| + |b_h|, \end{aligned}$$

and therefore, by the Weierstrass  $M$ -test, the trigonometric series

$$\frac{a_0}{2} + \sum_{h=1}^{\infty} (a_h \cos hx + b_h \sin hx)$$

converges to  $f(x)$  absolutely and uniformly. ▲

As a simple consequence of the proof of Theorem 3.1.8, we have the following important result.

**Corollary 3.1.8.** If

$$\sum_{h=1}^{\infty} (|a_h| + |b_h|)$$

converges, then the associated trigonometric series

$$\frac{a_0}{2} + \sum_{h=1}^{\infty} (a_h \cos hx + b_h \sin hx)$$

converges absolutely and uniformly to a continuous periodic function of period  $2\pi$  of which it is the Fourier series.

We are now in a position to prove a portion of Theorem 1.4.4 of Chapter 1. In the proof we require a result that will be used at later points, and therefore we state it as a lemma.

**Lemma 3.1.4. (Kronecker's lemma)** If the sequence  $\{a_j\}$  is such that

$$\lim_{n \rightarrow \infty} \sum_{j=0}^n |a_j| = A < \infty,$$

then

$$\lim_{n \rightarrow \infty} \sum_{j=0}^n \frac{j}{n} |a_j| = 0.$$

**Proof.** By assumption, given  $\epsilon > 0$ , there exists an  $N$  such that

$$\sum_{j=N+1}^{\infty} |a_j| < \epsilon.$$

Therefore, for  $n > N$ , we have

$$\sum_{j=0}^n \frac{j}{n} |a_j| < \frac{1}{n} \sum_{j=0}^N j |a_j| + \epsilon.$$

Clearly, for fixed  $N$ ,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=0}^N j |a_j| = 0,$$

and since  $\epsilon$  was arbitrary, the result follows. ▲

**Theorem 3.1.9.** Let the correlation function  $\rho(h)$  of a stationary time series be absolutely summable. Then there exists a continuous function  $f(\omega)$  such that:

- (i)  $\rho(h) = \int_{-\pi}^{\pi} f(\omega) \cos \omega h \, d\omega$ .
- (ii)  $f(\omega) \geq 0$ .
- (iii)  $\int_{-\pi}^{\pi} f(\omega) \, d\omega = 1$ .
- (iv)  $f(\omega)$  is an even function.

**Proof.** By Corollary 3.1.8

$$g(\omega) = \frac{1}{2} + \sum_{h=1}^{\infty} \rho(h) \cos h\omega$$

is a well-defined continuous function. Now, by the positive semidefinite property of the correlation function,

$$\sum_{m=1}^n \sum_{q=1}^n \rho(m-q) \cos m\omega \cos q\omega \geq 0$$

and

$$\sum_{m=1}^n \sum_{q=1}^n \rho(m-q) \sin m\omega \sin q\omega \geq 0.$$

Hence,

$$\begin{aligned} \sum_{m=1}^n \sum_{q=1}^n \rho(m-q) [\cos m\omega \cos q\omega + \sin m\omega \sin q\omega] \\ = \sum_{m=1}^n \sum_{q=1}^n \rho(m-q) \cos(m-q)\omega \geq 0. \end{aligned}$$

Letting  $m - q = h$ , we have

$$\sum_{h=-n+1}^{n-1} \left( \frac{n-|h|}{n} \right) \rho(h) \cos h\omega \geq 0.$$

Now,  $\rho(h) \cos h\omega$  is absolutely summable, and hence, by Lemma 3.1.4,

$$\lim_{n \rightarrow \infty} \sum_{h=-n+1}^{n-1} \frac{|h|}{n} \rho(h) \cos h\omega = 0.$$

Therefore,

$$\lim_{n \rightarrow \infty} \sum_{h=-n+1}^{n-1} \frac{n-|h|}{n} \rho(h) \cos h\omega = \sum_{h=-\infty}^{\infty} \rho(h) \cos h\omega = 2g(\omega) \geq 0.$$

Having shown that  $g(\omega)$  satisfies conditions (i) and (ii), we need only multiply  $g(\omega)$  by a constant to meet condition (iii). Since

$$\frac{1}{\pi} \int_{-\pi}^{\pi} g(\omega) d\omega = 1,$$

the appropriate constant is  $\pi^{-1}$ , and we define  $f(\omega)$  by

$$\begin{aligned} f(\omega) &= \frac{1}{\pi} \left[ \frac{1}{2} + \sum_{h=1}^{\infty} \rho(h) \cos h\omega \right] \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \rho(h) \cos h\omega. \end{aligned}$$

The function  $f(\omega)$  is an even function, since it is the uniform limit of a sum of even functions (cosines).  $\blacktriangle$

In Theorem 3.1.8 the square integrability of the derivative was used to demonstrate the convergence of the Fourier series. In fact, the Fourier series of a continuous function not meeting such restrictions need not converge. However, Cesàro's method may be used to recover any continuous periodic function from its

Fourier series. Given the sequence  $\{S_j\}_{j=1}^{\infty}$ , the sequence  $\{C_n\}$  defined by

$$C_n = \frac{1}{n} \sum_{j=1}^n S_j$$

is called the sequence of arithmetic means of  $\{S_j\}$ . If the sequence  $\{C_n\}$  is convergent, we say the sequence  $\{S_j\}$  is *Cesàro summable*. If the original sequence was convergent, then  $\{C_n\}$  converges.

**Lemma 3.1.5.** If the sequence  $\{S_j\}$  converges to  $s$ , then the sequence  $\{C_n\}$  converges to  $s$ .

**Proof.** By hypothesis, given  $\epsilon > 0$ , we may choose an  $N$  such that  $|S_j - s| < \frac{1}{2}\epsilon$  for all  $j > N$ . For  $n > N$ , we have

$$\begin{aligned} \left| \frac{1}{n} \sum_{j=1}^n S_j - s \right| &\leq \frac{1}{n} \sum_{j=1}^N |S_j - s| + \frac{1}{n} \sum_{j=N+1}^n |S_j - s| \\ &\leq \frac{1}{n} \sum_{j=1}^N |S_j - s| + \frac{1}{2}\epsilon. \end{aligned}$$

Since we can choose an  $n$  large enough so that the first term is less than  $\frac{1}{2}\epsilon$ , the result follows.  $\blacktriangle$

**Theorem 3.1.10.** Let  $f(\omega)$  be a continuous function of period  $2\pi$ . Then the Fourier series of  $f(\omega)$  is uniformly summable to  $f(\omega)$  by the method of Cesàro.

**Proof.** The Cesàro sum is

$$\begin{aligned} C_n(\omega) &= \frac{1}{n} \sum_{j=0}^{n-1} \sum_{k=0}^j (a_k \cos k\omega + b_k \sin k\omega) \\ &= \frac{1}{n} \sum_{j=0}^{n-1} \frac{1}{\pi} \int_{-\pi}^{\pi} f(\omega + u) \frac{\sin(j + \frac{1}{2})u}{2 \sin(u/2)} du \\ &= \frac{1}{n\pi} \int_{-\pi}^{\pi} \frac{f(\omega + u)}{2 \sin^2(u/2)} \sum_{j=0}^{n-1} \sin(u/2) \sin(j + \frac{1}{2})u du, \end{aligned}$$

where we have used Lemma 3.1.3. Now,

$$\begin{aligned} \sum_{j=0}^{n-1} \sin(u/2) \sin(j + \frac{1}{2})u &= \frac{1}{2} \sum_{j=0}^{n-1} [\cos ju - \cos(j + 1)u] \\ &= \frac{1}{2}(1 - \cos nu) \\ &= \sin^2(nu/2). \end{aligned} \tag{3.1.19}$$

Therefore,

$$\begin{aligned} C_n(\omega) - f(\omega) &= \frac{1}{n\pi} \int_{-\pi}^0 [f(\omega + u) - f(\omega)] \frac{\sin^2(nu/2)}{2 \sin^2(u/2)} du \\ &\quad + \frac{1}{n\pi} \int_0^\pi [f(\omega + u) - f(\omega)] \frac{\sin^2(nu/2)}{2 \sin^2(u/2)} du, \quad (3.1.20) \end{aligned}$$

where we have used

$$\frac{1}{\pi n} \int_{-\pi}^\pi \frac{\sin^2(nu/2)}{2 \sin^2(u/2)} du = 1.$$

Since  $f(\omega)$  is uniformly continuous on  $[-\pi, \pi]$ , given any  $\epsilon > 0$ , there is a  $\delta > 0$  such that

$$|f(\omega + u) - f(\omega)| < \frac{\epsilon}{2}$$

for all  $|u| < \delta$  and all  $\omega \in [-\pi, \pi]$ . We write the second integral of (3.1.20) as

$$\begin{aligned} &\frac{1}{\pi n} \int_0^\delta [f(\omega + u) - f(\omega)] \frac{\sin^2(nu/2)}{2 \sin^2(u/2)} du \\ &\quad + \frac{1}{\pi n} \int_\delta^\pi [f(\omega + u) - f(\omega)] \frac{\sin^2(nu/2)}{2 \sin^2(u/2)} du \\ &\leq \frac{\epsilon}{2\pi n} \int_0^\delta \frac{\sin^2(nu/2)}{2 \sin^2(u/2)} du \\ &\quad + \frac{1}{2\pi n \sin^2(\delta/2)} \int_0^\pi |f(\omega + u) - f(\omega)| du \\ &\leq \frac{\epsilon}{4} + \frac{M}{n \sin^2(\delta/2)}. \end{aligned}$$

where  $M$  is the maximum of  $|f(\omega)|$  on  $[-\pi, \pi]$ . A similar argument holds for the first integral of (3.1.20). Therefore, there exists an  $N$  such that (3.1.20) is less than  $\epsilon$  for all  $n > N$  and all  $\omega \in [-\pi, \pi]$ .  $\blacktriangle$

### 3.2. COMPLEX REPRESENTATION OF TRIGONOMETRIC SERIES

We can represent the trigonometric series in a complex form that is somewhat more compact and that will prove useful in certain applications. Since

$$\cos \theta = \frac{\cos \theta + i \sin \theta + \cos \theta - i \sin \theta}{2} = \frac{e^{i\theta} + e^{-i\theta}}{2}$$

and

$$\sin \theta = \frac{\cos \theta + i \sin \theta - (\cos \theta - i \sin \theta)}{2i} = \frac{(-i)(e^{i\theta} - e^{-i\theta})}{2},$$

we have

$$\begin{aligned} a_k \cos kx + b_k \sin kx &= a_k \left( \frac{e^{ikx} + e^{-ikx}}{2} \right) - b_k i \left( \frac{e^{ikx} - e^{-ikx}}{2} \right) \\ &= \left( \frac{a_k - ib_k}{2} \right) e^{ikx} + \left( \frac{a_k + ib_k}{2} \right) e^{-ikx}. \end{aligned}$$

Thus we can write the approximating sum for a function  $f(x)$ , defined on  $[-\pi, \pi]$ , as

$$S_n(x) = \frac{a_0}{2} + \sum_{k=1}^n (a_k \cos kx + b_k \sin kx) = \sum_{k=-n}^n c_k e^{ikx},$$

where

$$c_k = \frac{a_k - ib_k}{2}, \quad c_{-k} = c_k^* = \frac{a_k + ib_k}{2}, \quad k = 0, 1, 2, \dots.$$

The coefficients  $c_k$  are given by the integrals

$$\begin{aligned} c_k &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) [\cos kx - i \sin kx] dx \\ &= \frac{1}{2}(a_k - ib_k). \end{aligned} \tag{3.2.1}$$

Consider now a function  $X_i$  defined on  $N$  integers. Let us identify the integers by  $\{-m, -(m-1), \dots, -1, 0, 1, \dots, m\}$  if  $N$  is odd and by  $\{-(m-1), -(m-2), \dots, -1, 0, 1, \dots, m\}$  if  $N$  is even. Taking  $N$  to be even, the  $c_k$  are given by

$$\begin{aligned} c_k &= \frac{1}{N} \sum_{t=-(m-1)}^m X_t e^{-ik2\pi t/N} \\ &= \frac{1}{N} \sum_{t=-(m-1)}^m X_t \left( \cos \frac{2\pi kt}{N} - i \sin \frac{2\pi kt}{N} \right), \\ k &= -(m-1), \dots, 0, \dots, m, \end{aligned} \tag{3.2.2}$$

and

$$X_t = \sum_{k=-(m-1)}^m c_k e^{i 2 \pi k t / N}.$$

Note that  $c_0$  and  $c_m$  do not require separate definitions. The complex form thus makes manipulation somewhat easier, since the correct divisor need not be specified by frequency.

If  $N$  is odd and  $X_t$  is an even function (i.e.,  $X_t = X_{-t}$ ), then the coefficients  $c_k$  are real. Conversely, if  $X_t$  is an odd function, the coefficients  $c_k$  of equation (3.2.1) are pure imaginary.

### 3.3. FOURIER TRANSFORM—FUNCTIONS DEFINED ON THE REAL LINE

The results of Theorems 3.1.8 and 3.1.9 represent a special case of a more general result known as the *Fourier integral theorem*. By Theorem 3.1.8 the sequence of Fourier coefficients for a continuous periodic function with square integrable derivative can be used to construct a sequence of functions that converges to the original function. This result can be stated in a very compact form by substituting the definition of  $a_k$  and  $b_k$  into the statement of Theorem 3.1.7 to obtain

$$\begin{aligned} f(x) &= \frac{a_0}{2} + \sum_{k=1}^{\infty} \left[ \frac{1}{\pi} \int_{-\pi}^{\pi} f(\omega) \cos k\omega \cos kx d\omega \right. \\ &\quad \left. + \frac{1}{\pi} \int_{-\pi}^{\pi} f(\omega) \sin k\omega \sin kx d\omega \right] \\ &= \frac{a_0}{2} + \sum_{k=1}^{\infty} \frac{1}{\pi} \int_{-\pi}^{\pi} f(\omega) \cos k(\omega - x) d\omega \\ &= \frac{1}{2\pi} \sum_{k=-\infty}^{\infty} e^{-ikx} \int_{-\pi}^{\pi} f(\omega) e^{i k \omega} d\omega. \end{aligned} \quad (3.3.1)$$

Since the reciprocal relationships are well defined, we could also write, using the notation of Theorem 3.1.9,

$$\rho(k) = \frac{1}{2\pi} \int_{-\pi}^{\pi} e^{i k \omega} \sum_{h=-\infty}^{\infty} \rho(h) e^{-i h \omega} d\omega.$$

We say that  $\rho(k)$  and  $f(x)$  form a *transform pair*. We shall associate the constant and the negative exponential with one transform and call this the *Fourier transform* or *spectral density*. The terms *spectrum* or *spectral function* are also

used. Thus, in Theorem 3.1.9, the function  $f(\omega)$  defined by

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \rho(h) e^{-\epsilon \omega h}$$

is the Fourier transform of  $\rho(h)$ , or the spectral density associated with  $\rho(h)$ .

The transform in (3.3.1) with positive exponent and no constant we call the *inverse transform* or the *characteristic function*. Thus the correlation function

$$\rho(h) = \int_{-\pi}^{\pi} f(\omega) e^{\epsilon h \omega} d\omega$$

is the inverse transform, or characteristic function, of  $f(\omega)$ . We have mentioned several times that this is the statistical characteristic function if  $f(\omega)$  is a probability density.

These definitions are merely to aid us in remembering the transform to which we have attached the constant  $1/2\pi$ . We trust that the reader will not be disturbed to find that our definitional placement of the constant may differ from that of authors in other fields.

Theorem 3.1.8 was for a periodic function. However, there exists a considerable body of theory applicable to integrable functions defined on the real line. For an integrable function  $f(x)$  defined on the real line we formally define the Fourier transform by

$$c(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-\epsilon ux} dx, \quad (3.3.2)$$

where  $u \in (-\infty, \infty)$ . The Fourier integral theorem states that if the function  $f(x)$  meets certain regularity conditions, the inverse transform of the Fourier transform is again the function. We state one version without proof. [See Tolstov (1962, p. 188) for a proof.]

**Theorem 3.3.1.** Let  $f(x)$  be an absolutely integrable continuous function defined on the real line with a right and a left derivative at every point. Then, for all  $x \in (-\infty, \infty)$ ,

$$f(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{\epsilon ux} \int_{-\infty}^{\infty} f(t) e^{-\epsilon ut} dt du. \quad (3.3.3)$$

Table 3.3.1 contains a summary of the Fourier transforms of the different types of functions we have considered. We have presented theorems for more general functions than the continuous functions described in the table. Likewise, the finite transform obviously holds for a function defined on an odd number of integers.

For the first, third, and fourth entries in the column headed "Inverse Transform" we have first listed the domain of the original function. However, the inverse transform, being a multiple of the transform of the transform, is defined for the values indicated in parentheses.

### Summary of Fourier transforms

		Domain of Function	Fourier Transform	Inverse Transform
discrete $X_i$	$-(n-1), \dots, 0, \dots, n$	$c_k = \frac{1}{2n} \sum_{t=-\lfloor n-1 \rfloor}^{\lfloor n \rfloor} X_t e^{-i\pi k t/n}$ $k = -(n-1), \dots, 0, \dots, n$	$X_t = \sum_{k=-\lfloor n-1 \rfloor}^n c_k e^{i\pi k t/n}$ $t = -(n-1), \dots, 0$ (can be extended to $t$ )	
continuous $y$	$(0, \pm 1, \pm 2, \dots)$	$f(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h}$ $\omega \in (-\pi, \pi)$ $f(\omega)$ periodic of period $2\pi$	$\gamma(h) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega h} d\omega$ $h = 0, \pm 1, \pm 2, \dots$	
piecewise continuous	$[-\pi, \pi]$	$c_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(x) e^{-ikx} dx$ $k = 0, \pm 1, \pm 2, \dots$	$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{ikx}$ $x \in [-\pi, \pi]$ (can be extended as a function on the real line)	
periodic smooth	$(-T/2, T/2)$ (can be extended to the real line)	$c_k = \frac{1}{T} \int_{-T/2}^{T/2} f(x) e^{-i2\pi k x/T} dx$ $k = 0, \pm 1, \pm 2, \dots$	$f(x) = \sum_{k=-\infty}^{\infty} c_k e^{i2\pi k x/T}$ $x \in (-T/2, T/2)$ (can be extended to real line)	
square	$(-\infty, \infty)$	$c(u) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-iux} dx$ $u \in (-\infty, \infty)$	$\tilde{f}(x) = \int_{-\infty}^{\infty} c(u) e^{iux} du$ $x \in (-\infty, \infty)$	

Given a continuous function defined on an interval, it may be convenient to record the function at a finite number of equally spaced points and find the Fourier transform of the  $N$  points. Time series are often created by reading “continuous” functions of time at fixed intervals. For example, river level and temperature have the appearance of continuous functions of time, but we may record for analysis the readings at only a few times during a day or season.

Let  $g(x)$  be a continuous function with square integrable derivative defined on the interval  $[-T, T]$ . We evaluate the function at the  $2m$  points

$$x_t = \left( \frac{t}{m} \right) T, \quad t = -(m-1), -(m-2), \dots, m-1, m.$$

The complex form of the Fourier coefficients for the vector  $g(x_i)$  is given by

$$c_k^{(m)} = \frac{1}{2m} \sum_{t=-(m-1)}^m g(x_t) e^{-i\pi k t/m}, \quad k = -(m-1), -(m-2), \dots, m,$$

and  $g(x_i)$  is expressible as

$$g(x_t) = \sum_{k=-(m-1)}^m c_k^{(m)} e^{i\pi k t/m}, \quad t = -(m-1), -(m-2), \dots, m. \quad (3.3.4)$$

The superscript  $(m)$  on the coefficients is to remind us that they are based on  $2m$  equally spaced values of  $g(x)$ .

If we compute the Fourier coefficients for the function using the integral form (3.2.1), we have

$$c_k = \frac{1}{2T} \int_{-T}^T e^{-i(k\pi/T)x} g(x) dx, \quad k = 0, \pm 1, \pm 2, \dots,$$

and the original function evaluated at the points  $x_t$ ,  $t = -(m-1), -(m-2), \dots, m$ , is given by

$$\begin{aligned} g(x_t) &= \sum_{k=-\infty}^{\infty} c_k e^{i(k\pi/T)Tt/m} \\ &= \sum_{k=-\infty}^{\infty} c_k e^{i\pi k t/m}. \end{aligned}$$

By the periodicity of the complex exponential we have

$$g(x_t) = \sum_{k=-(m-1)}^m e^{i\pi k t/m} [c_k + (c_{k-2m} + c_{k+2m}) + (c_{k-4m} + c_{k+4m}) + \dots]. \quad (3.3.5)$$

**Table 3.3.2. Computed Coefficients for Function (3.3.6) Observed at 10 Points**

Period	Frequency	Cosine Coefficient $c_k^{(5)} + c_{-k}^{(5)}$	Contributing Alias Frequencies
—	0	0.5	10/20
20	1/20	0.0	—
10	2/20	0.0	—
20/3	3/20	1.6	13/20, 23/20
5	4/20	0.0	—
4	5/20	0.3	15/20

Equating the coefficients of  $e^{i\pi k t/m}$  in (3.3.4) and (3.3.5), we have

$$c_k^{(m)} = c_k + \sum_{s=1}^{\infty} (c_{k-2ms} + c_{k+2ms}).$$

It follows that the coefficient for the  $k$ th frequency of the function defined at  $2m$  points is the sum of the coefficients of the continuous function at the  $k, k+2m, k-2m, k+4m, k-4m, \dots$  frequencies. The frequencies  $k \pm 2m, k \pm 4m, \dots$  are called the *aliases* of the  $k$ th frequency.

In our representation the distance between two points is  $T/m$ . The cosine of period  $2T/m$  or frequency  $m/2T$  is the function of highest frequency that will be used in the Fourier transform. The frequency  $m/2T$  is called the *Nyquist* frequency. The aliases of an observed frequency are frequencies obtained by adding or subtracting integer multiples of twice the Nyquist frequency.

To illustrate these ideas, let  $g(x)$  defined on  $[-10, 10]$  be given by

$$\begin{aligned} g(x) = & 1.0 \cos 2\pi \frac{3}{20}x + 0.5 \cos \pi x \\ & + 0.4 \cos 2\pi \frac{13}{20}x + 0.3 \cos 2\pi \frac{15}{20}x \\ & + 0.2 \cos 2\pi \frac{23}{20}x. \end{aligned} \quad (3.3.6)$$

If the function is observed at the 10 points  $-8, -6, \dots, 8, 10$  and the  $\{c_k^{(5)} : k = -4, -3, \dots, 4, 5\}$  are computed by

$$c_k^{(5)} = \frac{1}{10} \sum_{t=-4}^5 g\left(\frac{t}{5} 10\right) e^{-i\pi k t/5},$$

we will obtain the coefficients given in Table 3.3.2.

### 3.4. FOURIER TRANSFORM OF A CONVOLUTION

Fourier transform theory is particularly useful for certain function-of-function problems. We consider one such problem that occurs in statistics and statistical time series.

For absolutely integrable functions  $f(x)$  and  $g(x)$  defined on the real line, the function

$$\varphi(x) = \int_{-\infty}^{\infty} f(x-y)g(y) dy$$

is called the *convolution* of  $f(x)$  and  $g(x)$ . Since  $f(x)$  and  $g(x)$  are absolutely integrable,  $\varphi(x)$  is absolutely integrable.

We have the following theorem on the Fourier transform of a convolution.

**Theorem 3.4.1.** If the functions  $f(x)$  and  $g(x)$  are absolutely integrable, then the Fourier transform of the convolution of  $f(x)$  and  $g(x)$  is given by

$$\begin{aligned} c_{\varphi}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(x) e^{-i\omega x} dx \\ &= 2\pi c_f(\omega)c_g(\omega), \end{aligned}$$

where

$$c_f(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} f(x) e^{-i\omega x} dx$$

and

$$c_g(\omega) = \frac{1}{2\pi} \int_{-\infty}^{\infty} g(x) e^{-i\omega x} dx$$

are the Fourier transforms of  $f(x)$  and  $g(x)$ .

**Proof.** The integrals defining  $c_{\varphi}(\omega)$ ,  $c_f(\omega)$ , and  $c_g(\omega)$  are absolutely convergent. Hence,

$$\begin{aligned} c_{\varphi}(\omega) &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \varphi(x) e^{-i\omega x} dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x-y)g(y) e^{-i\omega x} dy dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x-y) e^{-i\omega(x-y)} g(y) e^{-i\omega y} dy dx \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} f(z) e^{-i\omega z} dz \int_{-\infty}^{\infty} g(y) e^{-i\omega y} dy \\ &= 2\pi c_f(\omega)c_g(\omega), \end{aligned} \tag{3.4.1}$$

where the absolute integrability has enabled us to interchange the order of integration. ▲

We discussed the convolution of two absolutely summable sequences in Section 2.2 and demonstrated that the convolution was absolutely summable.

**Corollary 3.4.1.1.** Given that  $\{a_j\}$  and  $\{b_j\}$  are absolutely summable, the Fourier transform of

$$d_m = \sum_{j=-\infty}^{\infty} a_{m-j} b_j$$

is

$$f_d(\omega) = 2\pi f_a(\omega) f_b(\omega),$$

where

$$f_a(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} a_j e^{-\epsilon\omega j}$$

and

$$f_b(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} b_j e^{-\epsilon\omega j}.$$

**Proof.** The Fourier transform of  $\{d_m\}$  is

$$\begin{aligned} f_d(\omega) &= \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} d_m e^{-\epsilon\omega m} = \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_{m-j} b_j e^{-\epsilon\omega m} \\ &= \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} a_{m-j} b_j e^{-\epsilon\omega(m-j)} e^{-\epsilon\omega j} \\ &= 2\pi f_a(\omega) f_b(\omega). \end{aligned}$$

▲

We may paraphrase these results as follows: the spectral density (Fourier transform) of a convolution is the product of the spectral densities (Fourier transforms) multiplied by  $2\pi$ . That the converse is also true is clear from the proofs. We give a direct statement and proof for the product of absolutely summable sequences.

**Corollary 3.4.1.2.** Let  $\{a_m\}$  and  $\{b_m\}$  be absolutely summable, and define  $d_m$  by

$$d_m = a_m b_m.$$

Then

$$\begin{aligned} f_d(\omega) &= \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} d_m e^{-im\omega} = \int_{-\pi}^{\pi} f_a(u) f_b(\omega - u) du \\ &= \int_{-\pi}^{\pi} f_a(\omega - u) f_b(u) du, \end{aligned}$$

where  $f_a(\omega)$  and  $f_b(\omega)$  are defined in Corollary 3.4.1.1.

**Proof.** We have

$$\begin{aligned} f_d(\omega) &= \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} d_m e^{-im\omega} \\ &= \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} a_m b_m e^{-im\omega} \\ &= \frac{1}{2\pi} \sum_{m=-\infty}^{\infty} \left( \int_{-\pi}^{\pi} f_a(x) e^{imx} dx \right) b_m e^{-im\omega} \\ &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f_a(x) \sum_{m=-\infty}^{\infty} b_m e^{-im(\omega-x)} dx \\ &= \int_{-\pi}^{\pi} f_a(x) f_b(\omega - x) dx. \end{aligned}$$
▲

## REFERENCES

Jenkins and Watts (1968), Lighthill (1970), Nerlove (1964), Rogosinski (1959), Tolstov (1962), Zygmund (1959).

## EXERCISES

- Give the 12 orthogonal sine and cosine functions that furnish an orthogonal basis for the 12-dimensional vector space. Find the linear combinations of these vectors that yield the vectors  $(1, 0, \dots, 0)$ ,  $(0, 1, 0, \dots, 0), \dots, (0, \dots, 0, 1)$ .
- Expand the following functions defined on  $[-\pi, \pi]$  in Fourier series:
  - $f(\omega) = \cos a\omega$ , where  $a$  is not an integer.
  - $f(\omega) = \sin a\omega$ , where  $a$  is not an integer.
  - $f(\omega) = e^{a|\omega|}$ , where  $a \neq 0$ .
  - $f(\omega) = \begin{cases} 0, & -\pi \leq \omega \leq 0 \\ \sin \omega, & 0 < \omega < \pi. \end{cases}$

3. Define  $g(x)$  by

$$g(x) = \begin{cases} 0, & -\pi \leq x \leq 0, \\ 1, & 0 < x < \pi. \end{cases}$$

(a) Find the Fourier coefficients for  $g(x)$ .

(b) What is the maximum value for

$$S_3(x) = \frac{1}{2} + \sum_{k=1}^3 (a_k \cos kx + b_k \sin kx)?$$

Where does this maximum occur? What is the maximum value of  $S_4(x)$ ?  $S_5(x)$ ? The fact that the approximating function always overestimates the true function near the point of discontinuity is called *Gibbs' phenomenon*.

4. Prove Theorem 3.1.2.

5. Let  $f(x)$  be the periodic function defined on the real line by

$$f(x) = \begin{cases} b^{-1}, & 2\pi j - b < x < 2\pi j + b, \\ 0 & \text{otherwise,} \end{cases}$$

where  $j = 0, \pm 1, \pm 2, \dots$  and  $0 < b < \pi$ . Find the Fourier transform of  $f(x)$ .

6. Let

$$f(x) = \begin{cases} 1, & -b \leq x < b, \\ 0 & \text{otherwise,} \end{cases}$$

where  $b$  is a positive number and  $f(x)$  is defined on the real line. Find the Fourier transform of  $f(x)$ . Show that the limit of this transform at zero is infinity as  $b \rightarrow \infty$ . Show that as  $b \rightarrow \infty$  the transform is bounded except at zero.

7. Let

$$\delta_n(x) = \begin{cases} 2\pi n, & -\frac{1}{2n} \leq x < \frac{1}{2n}, \\ 0 & \text{otherwise,} \end{cases}$$

where  $n$  is a positive integer and  $\delta_n(x)$  is defined on the real line. Find the Fourier transform of  $\delta_n(x)$ . Show that, as  $n \rightarrow \infty$ , the transform tends to the constant function of unit height.

8. Let the *generalized function*  $\delta(x)$  represent a sequence of functions  $\{\delta_n(x)\}_{n=1}^\infty$

such that

$$\lim_{n \rightarrow \infty} \int_{-\infty}^{\infty} \delta_n(x) f(x) dx = f(0),$$

where  $f(x)$  is a continuous absolutely integrable function defined on the real line. Then  $\delta(x)$  is called *Dirac's delta function*. Show that the sequence of functions  $\{\delta_n(x)\}$  of Exercise 7 defines a generalized function.

9. Let  $g_n(x) = (n/\pi)^{1/2} e^{-nx^2}$  for  $x \in (-\infty, \infty)$ . Show that the sequence  $\{g_n(x)\}_{n=1}^{\infty}$  yields a Dirac delta function as defined in Exercise 8.
10. Let  $f(x)$  defined for  $x \in (-\infty, \infty)$  have the Fourier transform  $c(u)$ . Show that the Fourier transform of  $f(ax + b)$  is  $|a|^{-1} e^{ibu/a} c(u/a)$ ,  $a \neq 0$ .
11. Assume that price of a commodity is recorded on the last day of each month for a period of 144 months. The finite Fourier coefficients for the data are computed using the formulas following (3.1.2). Which coefficients will be affected if there is a weekly periodicity in prices that is perfectly represented by a sine wave of period 7 days? Assume that there are 30.437 days in a month. Which coefficients will be affected if the weekly periodicity in prices can be represented by the sum of two sine waves, one of period 7 days and one of period  $3\frac{1}{2}$  days? See Granger and Hatanaka (1964).
12. Let  $f(x)$  and  $g(x)$  be absolutely integrable functions, and define

$$\varphi(x) = \int_{-\infty}^{\infty} f(x-y) g(y) dy.$$

Show that  $\varphi(x)$  satisfies

$$\int_{-\infty}^{\infty} |\varphi(x)| dx \leq \left( \int_{-\infty}^{\infty} |f(x)| dx \right) \left( \int_{-\infty}^{\infty} |g(y)| dy \right).$$

13. Let  $f(x)$  and  $g(x)$  be continuous absolutely integrable functions defined on the real line. State and prove the result analogous to Corollary 3.4.1.2 for  $\psi(x) = f(x)g(x)$ .
14. Give a direct proof of Corollary 3.4.1.2 for finite transforms. That is, for the two functions  $\gamma(h)$  and  $w(h)$  defined on the  $2n - 1$  integers  $h = 0, \pm 1,$

$\pm 2, \dots, \pm(n-1)$ , show that

$$\begin{aligned} g(\omega_s) &= \frac{1}{2n-1} \sum_{h=-\lfloor n-1 \rfloor}^{n-1} w(h) \gamma(h) e^{-s\omega_s h} \\ &= \sum_{k=-\lfloor n-1 \rfloor}^{n-1} W(\omega_k) f(\omega_{s-k}), \end{aligned}$$

where

$$\begin{aligned} \omega_k &= \frac{2\pi k}{2n-1}, \quad k = 0, \pm 1, \pm 2, \dots, \pm(n-1), \\ W(\omega_k) &= \frac{1}{2n-1} \sum_{h=-\lfloor n-1 \rfloor}^{n-1} w(h) e^{-s\omega_k h}, \\ f(\omega_k) &= \frac{1}{2n-1} \sum_{h=-\lfloor n-1 \rfloor}^{n-1} \gamma(h) e^{-s\omega_k h}. \end{aligned}$$

15. Let  $f(\omega)$  be a nonnegative even continuous periodic function of period  $2\pi$ . Show that

$$c(h) = \int_{-\pi}^{\pi} f(\omega) e^{-s\omega h} d\omega, \quad h = 0, \pm 1, \pm 2, \dots,$$

is an even positive semidefinite function.

## CHAPTER 4

# Spectral Theory and Filtering

In Chapter 1 we discussed the correlation function as a way of characterizing a time series. In Chapter 2 we investigated representations of time series in terms of more elementary time series. These two descriptions are sometimes called descriptions in the *time domain* because of the obvious importance of the index set in the representations.

In Chapter 3 we introduced the Fourier transform of the correlation function. For certain correlation functions we demonstrated the uniqueness of the transform and showed that the correlation function is expressible as the inverse transform of the Fourier transform. The Fourier transform of the absolutely summable correlation function was called the spectral density. The spectral density furnishes another important representation of the time series. Because of the periodic nature of the trigonometric functions, the Fourier transform is often called the representation in the *frequency domain*.

### 4.1. THE SPECTRUM

In Chapter 1 we stated the result that the correlation function of a time series is analogous to a statistical characteristic function and may be expressed in the form

$$\rho(h) = \int_{-\pi}^{\pi} e^{i\omega h} dG(\omega), \quad (4.1.1)$$

where the integral is a Lebesgue–Stieltjes integral and  $G(\omega)$  is a statistical distribution function. In Theorem 3.1.9 we proved this result for time series with absolutely summable covariance functions.

Since the covariance function of a stationary time series is the correlation function multiplied by the variance of the process, we have

$$\gamma(h) = \int_{-\pi}^{\pi} e^{i\omega h} dF(\omega), \quad (4.1.1a)$$

where

$$dF(\omega) = \gamma(0) dG(\omega).$$

Both of the functions  $G(\omega)$  and  $F(\omega)$  have been called the *spectral distribution function* in time series analysis. The spectral distribution function is a non-decreasing function that, for our purposes, can be assumed to be composed of the sum of two parts: an absolutely continuous portion and a step function.<sup>1</sup> We take (4.1.1a) as the definitional relationship between the spectral distribution function and the covariance function. The spectral distribution function is also sometimes called the *integrated spectrum*.

Let us assume  $\gamma(h)$  is absolutely summable. Then, by Theorem 3.1.9,  $f(\omega)$  defined by

$$\begin{aligned} f(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i\omega h} \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \cos \omega h \end{aligned} \quad (4.1.2)$$

is a continuous nonnegative even function, and

$$\gamma(h) = \int_{-\pi}^{\pi} f(\omega) e^{i\omega h} d\omega.$$

Thus for time series with absolutely summable covariance function,  $dF(\omega) = f(\omega) d\omega$ , where  $f(\omega)$  was introduced as the spectral density function in Section 3.3.

Recall that we have taken  $\{e_i\}$  to be a time series of uncorrelated  $(0, \sigma^2)$  random variables. The spectral density of  $\{e_i\}$  is

$$\begin{aligned} f_e(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_e(h) \cos \omega h \\ &= \frac{1}{2\pi} \sigma^2, \end{aligned}$$

<sup>1</sup> Any statistical distribution function can be decomposed into three components: (1) a step function containing at most a countable number of finite jumps; (2) an absolutely continuous function; and (3) a "continuous singular" function. The third portion will be ignored in our treatment. See Tucker (1967, p. 15 ff.). While not formally correct, one may think of  $F(\omega)$  as the sum of two parts, a step function with jumps at the points  $\omega_j$ ,  $j = -M, -(M-1), \dots, M-1, M$ , and a function with continuous first derivative. Then the Lebesgue-Stieltjes integral  $\int g(\omega) dF(\omega)$  is the sum of  $\sum_{j=-M}^M g(\omega_j)l(\omega_j)$  and  $\int g(\omega)f(\omega) d\omega$ , where  $l(\omega_j)$  is the height of the jump in  $F(\omega)$  at the point  $\omega_j$ ,  $f(\omega)$  is the derivative of the continuous portion of  $F(\omega)$ , and  $\int g(\omega)f(\omega) d\omega$  is a Riemann integral.

which is positive, continuous, and trivially periodic. Similarly,

$$\begin{aligned}\gamma_e(h) &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sigma^2 e^{i\omega h} d\omega \\ &= \int_{-\pi}^{\pi} \frac{1}{2\pi} \sigma^2 \cos \omega h d\omega \\ &= \begin{cases} \sigma^2, & h = 0, \\ 0 & \text{otherwise.} \end{cases}\end{aligned}$$

As noted previously, the time series  $e$ , is often called white noise. The reason for this description is now more apparent. The spectrum is a constant multiple of the variance and, in analogy to white light, one might say that all frequencies contribute equally to the variance.

If the spectral distribution function contains finite jumps, we can visualize the spectrum containing discrete "spikes" at the jump points in the same way that we view discrete probability distributions. We now investigate a time series whose spectral distribution function is a step function.

Let a time series be composed of a finite sum of simple processes of the form (1.6.3). That is, define the time series  $Y_t$  by

$$Y_t = \sum_{j=0}^M (A_j \cos \omega_j t + B_j \sin \omega_j t), \quad (4.1.3)$$

where the  $A_j$  and  $B_j$  are random variables with zero mean and

$$\begin{aligned}E\{A_j^2\} &= E\{B_j^2\} = \sigma_j^2, & j = 0, 1, 2, \dots, M, \\ E\{B_j B_k\} &= E\{A_j A_k\} = 0, & j \neq k, \\ E\{A_j B_k\} &= 0 & \forall j, k,\end{aligned}$$

and  $\omega_j$ ,  $j = 0, 1, \dots, M$ , are distinct frequencies contained in the interval  $[-\pi, \pi]$ . By (1.6.4), we have

$$\begin{aligned}\gamma_Y(h) &= E\{Y_t Y_{t+h}\} = \sum_{j=0}^M \sigma_j^2 [\cos \omega_j t \cos \omega_j (t+h)] \\ &\quad + \sum_{j=0}^M \sigma_j^2 [\sin \omega_j t \sin \omega_j (t+h)] \\ &= \sum_{j=0}^M \sigma_j^2 \cos \omega_j h.\end{aligned} \quad (4.1.4)$$

Since the function  $\gamma_Y(h)$  is composed of a finite sum of cosine functions, the graph of  $\sigma_j^2$  against  $\omega_j$  (or  $j$ ) will give us a picture of the relative contribution of the variance associated with frequency  $\omega_j$  to the variance of the time series. This is

true because the variance of  $Y_t$  is given by

$$\gamma_Y(0) = \sum_{j=0}^M \sigma_j^2.$$

While we permitted our original frequencies in (4.1.3) to lie anywhere in the interval  $[-\pi, \pi]$ , it is clear that with no loss of generality we could have restricted the frequencies to  $[0, \pi]$ . That is, the covariance function for

$$X_t = A_j \cos(-\omega_j t) + B_j \sin(-\omega_j t)$$

is the same as that of

$$X_t = A_j \cos \omega_j t + B_j \sin \omega_j t.$$

This suggests that for a covariance function of the form  $\sigma_j^2 \cos \omega_j h$  we associate one half of the variance with the frequency  $-\omega_j$  and one half with the frequency  $\omega_j$ . To this end, we set

$$l(\omega_j) = l(-\omega_j) = \begin{cases} \frac{1}{2} \sigma_j^2, & \omega_j \neq 0, \\ \sigma_j^2, & \omega_j = 0. \end{cases}$$

To facilitate our representation, we assume that  $\omega_0 = 0$  and then write the sum (4.1.4) as

$$\gamma_Y(h) = \sum_{j=-M}^M l(\omega_j) \cos \omega_j h, \quad (4.1.5)$$

where  $\omega_{-j} = -\omega_j$ .

We say that a time series with a covariance function of the form (4.1.5) has a *discrete spectrum* or a *line spectrum*. Equation (4.1.5) may be written as the Lebesgue-Stieltjes integral

$$\begin{aligned} \gamma_Y(h) &= \int_{-\pi}^{\pi} \cos \omega h dF(\omega) \\ &= \int_{-\pi}^{\pi} e^{i\omega h} dF(\omega), \end{aligned} \quad (4.1.5a)$$

where  $F(\omega)$  is a step function with jumps of height  $\frac{1}{2} \sigma_j^2$  at the points  $\omega_j$  and  $-\omega_j$ ,  $\omega_j \neq 0$ , and a jump of height  $\sigma_0^2$  at  $\omega_0 = 0$ . By construction, the jumps  $l(\omega_j)$  are symmetric about zero, and we have expressed  $\gamma_Y(h)$  in the general form (4.1.1a). We used this method of construction because the covariance function  $\gamma_Y(h) = \sum_{j=0}^M \sigma_j^2 \cos \omega_j h$  is not absolutely summable and we could not directly apply (4.1.2).

For our purposes it is sufficient for us to be able to recognize the two types of autocovariance functions and associated spectra: (1) the autocovariance function

that is the sum of a finite number of cosine functions and is associated with a spectral distribution function, which is a step function; and (2) the absolutely summable autocovariance function that is associated with a continuous spectral density.

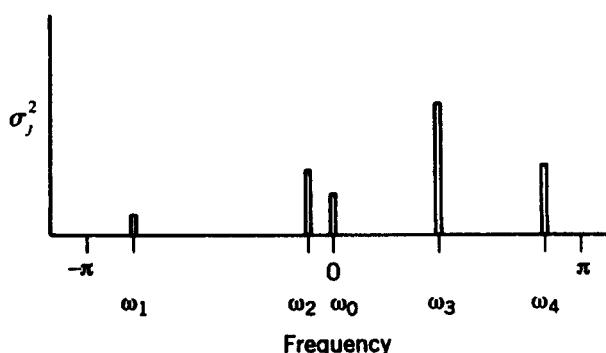
**Example 4.1.1.** Let us consider an example. Assume that the covariance function is given by (4.1.4) with the variances and frequencies specified by Table 4.1.1. Defining  $l(\omega_j)$  as in (4.1.5), we have

$$\begin{aligned} l(0) &= \frac{1}{2}, \\ l(-\frac{3}{4}\pi) &= l(\frac{3}{4}\pi) = \frac{1}{8}, \\ l(-\frac{1}{10}\pi) &= l(\frac{1}{10}\pi) = \frac{3}{8}, \\ l(-\frac{3}{8}\pi) &= l(\frac{3}{8}\pi) = \frac{5}{8}, \\ l(-\frac{7}{8}\pi) &= l(\frac{7}{8}\pi) = \frac{3}{8}. \end{aligned}$$

The original variances are plotted against frequency in Figure 4.1.1, and the line spectrum is plotted in Figure 4.1.2. The associated spectral distribution function is given in Figure 4.1.3.  $\blacktriangle\blacktriangle$

**Table 4.1.1. Examples of Variances for Time Series of Form (4.1.3)**

$j$	$\sigma_j^2$	$\omega_j$
0	$\frac{1}{2}$	0
1	$\frac{1}{4}$	$-\frac{3}{4}\pi$
2	$\frac{3}{4}$	$-\frac{1}{10}\pi$
3	$\frac{5}{4}$	$\frac{3}{8}\pi$
4	$\frac{3}{4}$	$\frac{7}{8}\pi$



**Figure 4.1.1. Graph of  $\sigma_j^2$  for the time series of Table 4.1.1.**

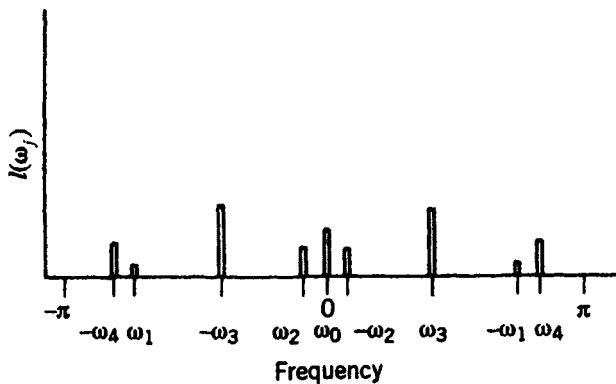


Figure 4.1.2. Graph of the line spectrum for the time series of Table 4.1.1.

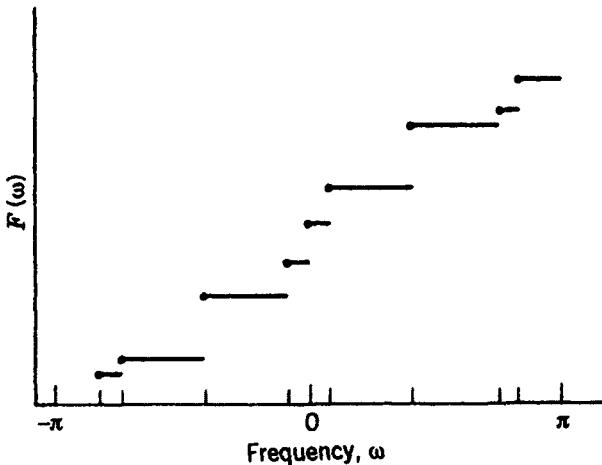


Figure 4.1.3. Spectral distribution function associated with the spectrum of Figure 4.1.2.

**Example 4.1.2.** As a second example, let  $X_t$  be defined by

$$X_t = \epsilon_1 \cos \frac{\pi}{2} t + \epsilon_2 \sin \frac{\pi}{2} t + e_t, \quad (4.1.6)$$

where the  $e_t$ ,  $t = 0, \pm 1, \pm 2, \dots$ , are independent  $(0, 0.2\pi)$  random variables independent of the  $\epsilon_j$ ,  $j = 1, 2$ , which are independent  $(0, 0.8\pi)$  random variables. Letting

$$Y_t = \epsilon_1 \cos \frac{\pi}{2} t + \epsilon_2 \sin \frac{\pi}{2} t,$$

it follows that

$$\begin{aligned} F_X(\omega) &= F_Y(\omega) + F_e(\omega), \\ F_e(\omega) &= 0.1\pi + 0.1\omega, \quad -\pi \leq \omega \leq \pi, \end{aligned}$$

and

$$F_Y(\omega) = \begin{cases} 0, & -\pi \leq \omega < -\pi/2, \\ 0.4\pi, & -\pi/2 \leq \omega < \pi/2, \\ 0.8\pi, & \pi/2 \leq \omega \leq \pi. \end{cases}$$

Therefore,

$$F_X(\omega) = \begin{cases} 0.1\pi + 0.1\omega, & -\pi \leq \omega < -\pi/2, \\ 0.5\pi + 0.1\omega, & -\pi/2 \leq \omega < \pi/2, \\ 0.9\pi + 0.1\omega, & \pi/2 \leq \omega \leq \pi. \end{cases}$$

The autocovariance function of  $X_i$  is

$$\begin{aligned} \gamma_X(h) &= \int_{-\pi}^{\pi} e^{i\omega h} dF_X(\omega) \\ &= \sum_{j=-1}^1 l(\omega_j) \cos \omega_j h + \int_{-\pi}^{\pi} 0.1 \cos \omega h d\omega \\ &= \begin{cases} \pi, & h = 0, \\ 0.8\pi \cos \frac{\pi}{2}h & \text{otherwise,} \end{cases} \end{aligned}$$

where  $\omega_{-1} = -\pi/2$  and  $\omega_1 = \pi/2$ . The reader may verify this expression by evaluating  $E\{X_i X_{i+h}\}$  using the definition of  $X_i$  given in (4.1.6).  $\blacktriangle \blacktriangle$

## 4.2. CIRCULANTS—DIAGONALIZATION OF THE COVARIANCE MATRIX OF STATIONARY PROCESSES

In this section we investigate some properties of matrices encountered in time series analysis and use these properties to obtain the matrix that will approximately diagonalize the  $n \times n$  covariance matrix of  $n$  observations from a stationary time series with absolutely summable covariance function. Let the  $n \times n$  covariance matrix be denoted by  $\Gamma$ . Then

$$\Gamma = \begin{pmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(n-3) \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma(n-1) & \gamma(n-2) & \gamma(n-3) & \cdots & \gamma(0) \end{pmatrix}.$$

It is well known<sup>2</sup> that for any  $n \times n$  positive semidefinite covariance matrix  $\Gamma$ , there exists an  $M$  satisfying  $M'M = I$  such that

$$M'\Gamma M = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_n),$$

where  $\lambda_i$ ,  $i = 1, 2, \dots, n$ , are the characteristic roots of  $\Gamma$ . Our investigation will demonstrate that for large  $n$  the  $\lambda_i$  are approximately equal to  $2\pi f(\omega_j)$ , where  $f(\omega)$  is the spectral density of  $X$ , and  $\omega_j = 2\pi j/n$ ,  $j = 0, 1, 2, \dots, n - 1$ . This permits us to interpret the spectral density as a multiple of the variance of the orthogonal random variables defined by the transformation  $M$ . We initiate our study by introducing a matrix whose roots have a particularly simple representation.

A matrix of the form

$$\Gamma_c = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-2) & \gamma(n-1) \\ \gamma(n-1) & \gamma(0) & \cdots & \gamma(n-3) & \gamma(n-2) \\ \gamma(n-2) & \gamma(n-1) & \cdots & \gamma(n-4) & \gamma(n-3) \\ \vdots & \vdots & & \vdots & \vdots \\ \gamma(2) & \gamma(3) & \cdots & \gamma(0) & \gamma(1) \\ \gamma(1) & \gamma(2) & \cdots & \gamma(n-1) & \gamma(0) \end{bmatrix} \quad (4.2.1)$$

is called a *circular matrix* or *circulant*. Of course, in this definition  $\gamma(j)$  may be any number, but we use the covariance notation, since our immediate application will be to a matrix whose elements are covariances.

The characteristic roots  $\lambda_j$  and vectors  $x_j$  of the matrix (4.2.1) satisfy the equation

$$\Gamma_c x_j = \lambda_j x_j, \quad j = 1, 2, \dots, n, \quad (4.2.2)$$

and we have

$$\begin{aligned} \gamma(0)x_{1j} + \gamma(1)x_{2j} + \cdots + \gamma(n-2)x_{n-1,j} + \gamma(n-1)x_{nj} &= \lambda_j x_{1j}, \\ \gamma(n-1)x_{1j} + \gamma(0)x_{2j} + \cdots + \gamma(n-3)x_{n-1,j} + \gamma(n-2)x_{nj} &= \lambda_j x_{2j}, \\ &\vdots \\ \gamma(1)x_{1j} + \gamma(2)x_{2j} + \cdots + \gamma(n-1)x_{n-1,j} + \gamma(0)x_{nj} &= \lambda_j x_{nj}, \end{aligned} \quad (4.2.3)$$

where  $x_{kj}$  is the  $k$ th element of the  $j$ th characteristic vector. Let  $r_j$  be a root of the scalar equation  $r^n = 1$  and set  $x_{kj} = r_j^k$ . The system of equations (4.2.3) becomes

$$\begin{aligned} \gamma(0)r_j + \gamma(1)r_j^2 + \cdots + \gamma(n-2)r_j^{n-1} + \gamma(n-1)r_j^n &= \lambda_j r_j, \\ \gamma(n-1)r_j + \gamma(0)r_j^2 + \cdots + \gamma(n-3)r_j^{n-1} + \gamma(n-2)r_j^n &= \lambda_j r_j^2, \\ &\vdots \\ \gamma(1)r_j + \gamma(2)r_j^2 + \cdots + \gamma(n-1)r_j^{n-1} + \gamma(0)r_j^n &= \lambda_j r_j^n. \end{aligned} \quad (4.2.4)$$

<sup>2</sup> See, for example, Rao (1973, p. 39).

If we multiply the first equation of (4.2.4) by  $r_j^{n-1}$ , the second by  $r_j^{n-2}$ , and so forth, using  $r_j^{n+k} = r_j^k$ , we see that we shall obtain equality for each equation if

$$\lambda_j = \sum_{h=0}^{n-1} \gamma(h) r_j^h. \quad (4.2.5)$$

The equation  $r^n = 1$  has  $n$  distinct roots,  $e^{2\pi j/n}$ ,  $j = 1, 2, \dots, n$ , which may also be expressed as  $r_j = e^{-s2\pi j/n}$ ,  $j = 0, 1, \dots, n-1$ . Therefore, the  $n$  characteristic vectors associated with the  $n$  characteristic roots are given by

$$\mathbf{x}_j = \mathbf{g}_j^* = n^{-1/2} [1, e^{-s2\pi j/n}, e^{-s2\pi 2j/n}, \dots, e^{-s2\pi(n-1)j/n}]', \quad j = 0, 1, \dots, n-1, \quad (4.2.6)$$

where  $\mathbf{g}_j^*$  is the conjugate transpose of the row vector

$$\mathbf{g}_j = n^{-1/2} [1, e^{s2\pi j/n}, e^{s2\pi 2j/n}, \dots, e^{s2\pi(n-1)j/n}].$$

If we define the matrix  $\mathbf{G}$  by

$$\mathbf{G} = (\mathbf{g}_0', \mathbf{g}_1', \dots, \mathbf{g}_{n-1}')',$$

then

$$\mathbf{G}\mathbf{G}^* = \mathbf{I},$$

$$\mathbf{G}\Gamma_c\mathbf{G}^* = \text{diag}(\lambda_0, \lambda_1, \dots, \lambda_{n-1}),$$

where  $\mathbf{G}^*$  is the conjugate transpose of  $\mathbf{G}$ .

Setting  $\gamma(1) = \gamma(n-1)$ ,  $\gamma(2) = \gamma(n-2)$ ,  $\dots$ ,  $\gamma(h) = \gamma(n-h)$ ,  $\dots$  in  $\Gamma_c$ , we obtain the circular symmetric matrix

$$\Gamma_c = \begin{bmatrix} \gamma(0) & \gamma(1) & \gamma(2) & \cdots & \gamma(2) & \gamma(1) \\ \gamma(1) & \gamma(0) & \gamma(1) & \cdots & \gamma(3) & \gamma(2) \\ \gamma(2) & \gamma(1) & \gamma(0) & \cdots & \gamma(4) & \gamma(3) \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ \gamma(1) & \gamma(2) & \gamma(3) & \cdots & \gamma(1) & \gamma(0) \end{bmatrix}. \quad (4.2.7)$$

Substituting into (4.2.5), we obtain for the roots of (4.2.7)

$$\lambda_j = \begin{cases} \sum_{h=-(n-1)/2}^{(n-1)/2} \gamma(h) e^{-s2\pi h j/n}, & n \text{ odd}, \\ \sum_{h=-(n/2)+1}^{n/2} \gamma(h) e^{-s2\pi h j/n}, & n \text{ even}, \end{cases} \quad (4.2.8)$$

where we have used the periodic property  $e^{-s2\pi j(n-h)/n} = e^{s2\pi jh/n}$ . Note that these roots are real, as they must be for a real symmetric matrix.

In some applications it is preferable to work with a real matrix rather than the complex matrix  $\mathbf{G}$ . Consider first the case of  $n$  odd. Equation (4.2.8) may also be written

$$\lambda_j = \sum_{h=-\frac{(n-1)}{2}}^{\frac{(n-1)}{2}} \gamma(h) \cos \frac{2\pi}{n} h j, \quad j = 0, 1, \dots, n-1. \quad (4.2.9)$$

Since for  $0 \leq m \leq 2\pi$  we have  $\cos m = \cos(2\pi - m)$ , we see that there is a root for  $j = 0$  (or  $j = n$ ) and  $(n-1)/2$  roots of multiplicity two associated with  $j = 1, 2, \dots, (n-1)/2$ . For each of these repeated roots we can find two real orthogonal vectors. These are chosen to be

$$2^{1/2} n^{-1/2} \left[ 1, \cos 2\pi \frac{j}{n}, \cos 2\pi \frac{2j}{n}, \cos 2\pi \frac{3j}{n}, \dots, \cos 2\pi \frac{(n-1)j}{n} \right]$$

and

$$2^{1/2} n^{-1/2} \left[ 0, \sin 2\pi \frac{j}{n}, \sin 2\pi \frac{2j}{n}, \sin 2\pi \frac{3j}{n}, \dots, \sin 2\pi \frac{(n-1)j}{n} \right]. \quad (4.2.10)$$

If we choose the vectors in (4.2.6) associated with the  $j$ th and  $n-j$ th roots of one, the vectors in (4.2.10) are given by

$$2^{-1/2} (\mathbf{g}_j + \mathbf{g}_{n-j}) \quad \text{and} \quad 2^{-1/2} i(\mathbf{g}_j - \mathbf{g}_{n-j}).$$

Much the same pattern holds for the roots of a circular symmetric matrix of dimension  $n \times n$  where  $n$  is even. There is a characteristic vector  $n^{-1/2}(1, 1, \dots, 1)$  associated with  $j = 0$ , and a vector  $n^{-1/2}(1, -1, 1, \dots, -1)$  associated with  $j = n/2$ . The remaining  $(n/2) - 1$  roots have multiplicity two, and the roots are given by (4.2.8).

Define the orthogonal matrix  $\mathbf{Q}$  by setting  $n^{1/2} 2^{-1/2} \mathbf{Q}'$  equal to

$$\begin{bmatrix} 2^{-1/2} & 2^{-1/2} & 2^{-1/2} & \cdots & 2^{-1/2} \\ 1 & \cos 2\pi \frac{1}{n} & \cos 2\pi \frac{2}{n} & \cdots & \cos 2\pi \frac{n-1}{n} \\ 0 & \sin 2\pi \frac{1}{n} & \sin 2\pi \frac{2}{n} & \cdots & \sin 2\pi \frac{n-1}{n} \\ 1 & \cos 4\pi \frac{1}{n} & \cos 4\pi \frac{2}{n} & \cdots & \cos 4\pi \frac{n-1}{n} \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & \sin \frac{n-1}{2} 2\pi \frac{1}{n} & \sin \frac{n-1}{2} 2\pi \frac{2}{n} & \cdots & \sin \frac{n-1}{2} 2\pi \frac{n-1}{n} \end{bmatrix}. \quad (4.2.11)$$

Note that  $\mathbf{Q}$  is the matrix composed of the  $n$  characteristic vectors defined by

(4.2.10) and our illustration (4.2.11) is for odd  $n$ . Let  $X_t$  be a stationary time series with absolutely summable covariance function  $\gamma(h)$ . For odd  $n$ , define the  $n \times n$  diagonal matrix  $\mathbf{D}$  by

$$\mathbf{D} = \text{diag}\{d_1, d_2, \dots, d_n\}, \quad (4.2.12)$$

where

$$\begin{aligned} d_1 &= f(0) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h), \\ d_{2j} &= d_{2j+1} = f\left(2\pi \frac{j}{n}\right) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i2\pi h j/n}, \end{aligned} \quad (4.2.13)$$

$j = 1, 2, \dots, 0.5(n - 1)$ . For  $\Gamma_s$  defined by (4.2.7) the matrix  $\mathbf{Q}'\Gamma_s\mathbf{Q}$  is a diagonal matrix whose elements converge to  $2\pi\mathbf{D}$  as  $n$  increases. This also holds for even  $n$  if the definition of  $\mathbf{Q}$  is slightly modified. An additional row,

$$n^{-1/2}[1, -1, 1, \dots, 1, -1],$$

which is the characteristic vector associated with  $j = n/2$ , is added to the  $\mathbf{Q}'$  of (4.2.11) when  $n$  is even. The last entry in  $\mathbf{D}$  for even  $n$  is

$$d_n = f(\pi) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma(h) \cos \pi h.$$

The covariance matrix for  $n$  observations on  $X_t$  is given by

$$\Gamma = \begin{bmatrix} \gamma(0) & \gamma(1) & \cdots & \gamma(n-1) \\ \gamma(1) & \gamma(0) & \cdots & \gamma(n-2) \\ \vdots & \vdots & & \vdots \\ \gamma(n-1) & \gamma(n-2) & \cdots & \gamma(0) \end{bmatrix}. \quad (4.2.14)$$

For  $\gamma(h)$  absolutely summable, we now demonstrate that  $\mathbf{Q}'\Gamma\mathbf{Q}$  also converges to  $2\pi\mathbf{D}$ .

Let  $\mathbf{q}_i = [q_{1i}, q_{2i}, \dots, q_{ni}]'$  be the  $i$ th column of  $\mathbf{Q}$ . We have, for  $\Gamma$  defined in (4.2.14) and  $\Gamma_s$  defined in (4.2.7),

$$\begin{aligned} |\mathbf{q}_i' \Gamma_s \mathbf{q}_j - \mathbf{q}_i' \Gamma \mathbf{q}_j| &= \left| \sum_{m=1}^M [\gamma(m) - \gamma(n-m)] \right. \\ &\quad \times \left. \sum_{k=1}^m [q_{ki} q_{n-m+k,j} + q_{n-m+k,i} q_{kj}] \right|, \end{aligned} \quad (4.2.15)$$

where  $M = (n - 1)/2$  if  $n$  is odd and  $M = (n/2) - 1$  if  $n$  is even. Now (4.2.15) is less than

$$\begin{aligned} \frac{4}{n} \left[ \sum_{m=1}^M m|\gamma(m)| + \sum_{m=1}^M m|\gamma(n-m)| \right] \\ \leq \frac{4}{n} \left[ \sum_{m=1}^M m|\gamma(m)| + \sum_{h=M+1}^n M|\gamma(h)| \right], \quad (4.2.16) \end{aligned}$$

since  $q_k q_r \leq 2/n$  for all  $k, i, r, j \in \{1, 2, \dots, n\}$ . As  $n$  increases, the limit of the first term is zero by Lemma 3.1.4 and the limit of the second term is zero by the absolute summability of  $\gamma(h)$ . Therefore, the elements of  $\mathbf{Q}'\Gamma\mathbf{Q}$  converge to  $2\pi\mathbf{D}$ . We state the result as a theorem.

**Theorem 4.2.1.** Let  $\Gamma$  be the covariance matrix of  $n$  observations from a stationary time series with absolutely summable covariance function. Let  $\mathbf{Q}$  be defined by (4.2.11), and take  $\mathbf{D}$  to be the  $n \times n$  diagonal matrix defined in (4.2.12). Then, given  $\epsilon > 0$ , there exists an  $n_\epsilon$  such that for  $n > n_\epsilon$  every element of the matrix

$$\mathbf{Q}'\Gamma\mathbf{Q} - 2\pi\mathbf{D}$$

is less than  $\epsilon$  in magnitude.

**Corollary 4.2.1.** Let  $\Gamma$  be the covariance matrix of  $n$  observations from a stationary time series with covariance function that satisfies

$$\sum_{h=-\infty}^{\infty} |h| |\gamma(h)| = L < \infty.$$

Let  $\mathbf{Q}$  and  $\mathbf{D}$  be as defined in Theorem 4.2.1. Then every element of the matrix  $\mathbf{Q}'\Gamma\mathbf{Q} - 2\pi\mathbf{D}$  is less than  $4L/n$  in magnitude.

**Proof.** By (4.2.16) the difference between  $\mathbf{q}_i'\Gamma\mathbf{q}_j$  and  $\mathbf{q}_i'\Gamma\mathbf{q}_j$  is less than

$$\frac{4}{n} \left[ \sum_{m=1}^M m|\gamma(m)| + \sum_{h=M+1}^n M|\gamma(h)| \right] \leq \frac{4}{n} \sum_{m=1}^n m|\gamma(m)| < \frac{2L}{n}.$$

Now  $\mathbf{Q}'\Gamma\mathbf{Q}$  is a diagonal matrix, and the difference between an element of  $\mathbf{Q}'\Gamma\mathbf{Q}$  and an element of  $2\pi\mathbf{D}$  is

$$\begin{aligned} & \left| \sum_{h=-(n-1)/2}^{(n-1)/2} \gamma(h) e^{-i2\pi h j/n} - \sum_{h=-\infty}^{\infty} \gamma(h) e^{-i2\pi h j/n} \right| \\ & \leq 2 \sum_{h=(n-1)/2+1}^{\infty} |\gamma(h)| \leq \frac{4}{n} \sum_{h=(n+1)/2}^{\infty} h |\gamma(h)| < \frac{2L}{n}. \quad \blacktriangle \end{aligned}$$

We have demonstrated that, asymptotically, the  $\mathbf{Q}$  defined by (4.2.11) or the  $\mathbf{G}$  defined by (4.2.6) will diagonalize *all*  $\Gamma$ -matrices associated with stationary time series with absolutely summable covariance functions. That is, the transformation  $\mathbf{Q}$  applied to the  $n$  observations defines  $n$  new random variables that are “nearly” uncorrelated. Each of the new random variables is a linear combination of the  $n$  original observations with weights given in (4.2.10). The variances of both the  $2j$ th and  $(2j+1)$ th random variables created by the transformation are approximately  $2\pi f(2\pi j/n)$ .

For time series with absolutely summable covariance function the variance result holds for the random variable defined for any  $\omega$ , not just for integer multiples of  $2\pi/n$ . That is, the complex random variable

$$Y_{\omega,n} = n^{-1/2} \sum_{t=1}^n X_t e^{i\omega t}$$

has variance given by

$$\begin{aligned} E\{|Y_{\omega,n}|^2\} &= \frac{1}{n} E\left\{\sum_{t=1}^n X_t e^{-i\omega t} \sum_{t=1}^n X_t e^{i\omega t}\right\} \\ &= \frac{1}{n} E\left\{\sum_{t=1}^n \sum_{j=1}^n X_t X_j e^{-i\omega(t-j)}\right\} \\ &= \frac{1}{n} \sum_{t=1}^n \sum_{j=1}^n \gamma(t-j) e^{-i\omega(t-j)} \\ &= \sum_{h=-(n-1)}^{n-1} n^{-1} (n - |h|) \gamma(h) e^{-i\omega h}. \end{aligned}$$

By Lemma 3.1.4,

$$\lim_{n \rightarrow \infty} E\{|Y_{\omega,n}|^2\} = 2\pi f_X(\omega).$$

We shall return to this point in Chapter 7 and investigate functions of the random variables created by applying the transformation (4.2.10) or (4.2.6) to a finite number of observations from a realization.

### 4.3. THE SPECTRAL DENSITY OF MOVING AVERAGE AND AUTOREGRESSIVE TIME SERIES

By the Fourier integral theorem we know that the Fourier transform of an absolutely summable sequence is unique and that the sequence is given once again by the inverse transform of the transform. Therefore, we expect the spectral density of a particular kind of time series, say a finite moving average, to have a

distinctive form. We shall see that this is the case: the spectral densities of autoregressive and moving average processes are easily recognizable.

Stationary moving average, stationary autoregressive, and stationary autoregressive moving average time series are contained within the representation

$$X_t = \sum_{j=-\infty}^{\infty} b_j e_{t-j}, \quad (4.3.0)$$

where

$$\sum_{j=-\infty}^{\infty} |b_j| < \infty$$

and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables. Therefore, if we obtain the spectral density of a time series of the form (4.3.0), we shall have the spectral density for such types of time series. In fact, it is convenient to treat the case of an infinite weighted sum of a more general time series.

**Theorem 4.3.1.** Let  $X_t$  be a stationary time series with an absolutely summable covariance function and let  $\{a_j\}_{j=-\infty}^{\infty}$  be absolutely summable. Then the spectral density of

$$Y_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}$$

is

$$f_Y(\omega) = (2\pi)^2 f_X(\omega) f_a(\omega) f_a^*(\omega), \quad (4.3.1)$$

where  $f_X(\omega)$  is the spectral density of  $X_t$ ,

$$f_a(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} a_j e^{-j\omega j}$$

is the Fourier transform of  $a_j$ , and

$$f_a^*(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} a_j e^{j\omega j}$$

is the complex conjugate of the Fourier transform of  $a_j$ .

**Proof.** Given the absolute summability, we may interchange integration and summation and calculate the expectation term by term. Therefore, letting

$E\{X_t\} = 0$ , we have

$$\begin{aligned} E\{Y_t Y_{t+h}\} &= E\left\{\left(\sum_{j=-\infty}^{\infty} a_j X_{t-j}\right)\left(\sum_{k=-\infty}^{\infty} a_k X_{t-k+h}\right)\right\} \\ &= E\left\{\sum_{j=-\infty}^{\infty} a_j \sum_{k=-\infty}^{\infty} a_k X_{t-j} X_{t-k+h}\right\} \\ &= \sum_{j=-\infty}^{\infty} a_j \sum_{k=-\infty}^{\infty} a_k \gamma_X(h-k+j). \end{aligned}$$

If we set  $p = h - k + j$ , then

$$\gamma_Y(h) = \sum_{j=-\infty}^{\infty} a_j \sum_{p=-\infty}^{\infty} a_{j+h-p} \gamma_X(p). \quad (4.3.2)$$

Since  $\gamma_Y(h)$  is absolutely summable, we may write

$$\begin{aligned} f_Y(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-\epsilon\omega h} \gamma_Y(h) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-\epsilon\omega h} \sum_{j=-\infty}^{\infty} a_j \sum_{p=-\infty}^{\infty} a_{j+h-p} \gamma_X(p) \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} a_j a_{j+h-p} \gamma_X(p) e^{\epsilon\omega j} e^{-\epsilon\omega p} e^{-\epsilon\omega(j+h-p)} \\ &= \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} a_j e^{\epsilon\omega j} \sum_{p=-\infty}^{\infty} \gamma_X(p) e^{-\epsilon\omega p} \sum_{s=-\infty}^{\infty} a_s e^{-\epsilon\omega s}, \end{aligned}$$

where we have made the transformation  $s = j + h - p$ . ▲

The theorem is particularly useful, and we shall make repeated application of the result. If the  $X_t$  of the theorem are assumed to be uncorrelated random variables, we obtain the spectral density of a moving average time series.

**Corollary 4.3.1.1.** The spectral density of the moving average process

$$X_t = \sum_{j=-\infty}^{\infty} a_j e_{t-j},$$

where  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables and the sequence  $\{a_j\}$  is absolutely summable, is

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \left( \sum_{j=-\infty}^{\infty} a_j e^{-\epsilon j \omega} \right) \left( \sum_{j=-\infty}^{\infty} a_j e^{\epsilon j \omega} \right). \quad (4.3.3)$$

**Proof.** The result follows immediately from Theorem 4.3.1. ▲

Let us use (4.3.3) to calculate the spectral density of the finite moving average

$$X_t = \sum_{j=0}^p a_j e_{t-j}, \quad (4.3.4)$$

We have

$$\begin{aligned} f_X(\omega) &= \frac{\sigma^2}{2\pi} \left( \sum_{j=0}^p a_j e^{-\epsilon j \omega} \right) \left( \sum_{j=0}^p a_j e^{\epsilon j \omega} \right) \\ &= \frac{\sigma^2}{2\pi} (a_0 + a_1 e^{-\epsilon \omega} + a_2 e^{-\epsilon 2\omega} + \cdots + a_p e^{-\epsilon p \omega}) \\ &\quad \times (a_0 + a_1 e^{\epsilon \omega} + a_2 e^{\epsilon 2\omega} + \cdots + a_p e^{\epsilon p \omega}) \\ &= \frac{\sigma^2}{2\pi} [(a_0^2 + a_1^2 + a_2^2 + \cdots + a_p^2) \\ &\quad + (a_0 a_1 + a_1 a_2 + a_2 a_3 + \cdots + a_{p-1} a_p) e^{-\epsilon \omega} \\ &\quad + (a_0 a_1 + a_1 a_2 + a_2 a_3 + \cdots + a_{p-1} a_p) e^{\epsilon \omega} \\ &\quad + (a_0 a_2 + a_1 a_3 + \cdots + a_{p-2} a_p) e^{-\epsilon 2\omega} \\ &\quad + \cdots + a_0 a_p e^{-\epsilon p \omega} + a_0 a_p e^{\epsilon p \omega}] \\ &= \frac{1}{2\pi} [\gamma_X(0) + \gamma_X(1) e^{-\epsilon \omega} + \gamma_X(1) e^{\epsilon \omega} + \gamma_X(2) e^{-\epsilon 2\omega} \\ &\quad + \cdots + \gamma_X(p) e^{-\epsilon p \omega} + \gamma_X(p) e^{\epsilon p \omega}] \\ &= \frac{1}{2\pi} \left[ \sum_{h=-p}^p \gamma_X(h) e^{-\epsilon \omega h} \right]. \end{aligned} \quad (4.3.5)$$

Equation (4.3.5) is a proof of the corollary for finite  $p$  and serves to reinforce our confidence in the general result. The spectral density of a finite moving average may, alternatively, be expressed in terms of the roots of the auxiliary equation. The quantity

$$\sum_{j=0}^p a_j e^{-\epsilon \omega j}$$

is seen to be a polynomial in  $e^{-\epsilon \omega}$ . Therefore, it may be written as

$$\sum_{j=0}^p a_j e^{-\epsilon \omega j} = e^{-\epsilon \omega p} \sum_{j=0}^p a_j e^{\epsilon \omega(p-j)} = e^{-\epsilon \omega p} \prod_{j=1}^p (e^{\epsilon \omega} - m_j),$$

where  $a_0 = 1$  and the  $m_j$  are the roots of

$$m^p + a_1 m^{p-1} + a_2 m^{p-2} + \cdots + a_p = 0.$$

It follows that the spectral density of the  $\{X_t\}$  defined by (4.3.4) is

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \left[ \prod_{j=1}^p (e^{-\epsilon_j \omega} - m_j) \right] \left[ \prod_{j=1}^p (e^{\epsilon_j \omega} - m_j) \right]. \quad (4.3.6)$$

We can also use Corollary 4.3.1.1 to compute the spectral density of the first order autoregressive process

$$X_t = \sum_{j=0}^{\infty} \rho^j e_{t-j}, \quad |\rho| < 1.$$

The transform of the weights  $a_j = \rho^j$  is given by

$$\begin{aligned} g_a(\omega) &= \frac{1}{2\pi} \sum_{j=0}^{\infty} \rho^j e^{-\epsilon_j \omega} \\ &= \frac{1}{2\pi} \sum_{j=0}^{\infty} (\rho e^{-\epsilon \omega})^j = \frac{1}{2\pi} \frac{1}{1 - \rho e^{-\epsilon \omega}} \end{aligned}$$

and the complex conjugate by

$$g_a^*(\omega) = \frac{1}{2\pi} \sum_{j=0}^{\infty} \rho^j e^{\epsilon_j \omega} = \frac{1}{2\pi} \frac{1}{1 - \rho e^{\epsilon \omega}}.$$

Therefore,

$$\begin{aligned} f_X(\omega) &= (2\pi)^2 f_e(\omega) g_a(\omega) g_a^*(\omega) \\ &= \frac{\sigma^2}{2\pi} \left( \frac{1}{1 - \rho e^{-\epsilon \omega}} \right) \left( \frac{1}{1 - \rho e^{\epsilon \omega}} \right) \\ &= \frac{\sigma^2}{2\pi} \frac{1}{1 - 2\rho \cos \omega + \rho^2}, \end{aligned} \quad (4.3.7)$$

since  $f_e(\omega)$ , the spectral density of the uncorrelated sequence  $\{e_t\}$ , is  $\sigma^2/2\pi$  for all  $\omega$ .

We note that we could also find the spectral density of  $X_t$  by setting  $Y_t = e_t$  in (4.3.1) and using the known spectral density of  $e_t$ . Applying this approach to the  $p$ th order autoregressive time series, we have the following corollary.

**Corollary 4.3.1.2.** The spectral density of the  $p$ th order autoregressive time series  $\{X_t\}$  defined by

$$\sum_{j=0}^p \alpha_j X_{t-j} = e_t,$$

where  $\alpha_0 = 1$ ,  $\alpha_p \neq 0$ , the  $e_j$  are uncorrelated  $(0, \sigma^2)$  random variables, and the roots  $\{m_j : j = 1, 2, \dots, p\}$  of

$$R(m) = \sum_{j=0}^p \alpha_j m^{p-j} = 0$$

are less than one in absolute value, is given by

$$\begin{aligned} f_X(\omega) &= \frac{\sigma^2}{2\pi} \left[ \left( \sum_{j=0}^p \alpha_j e^{-\epsilon\omega j} \right) \left( \sum_{j=0}^p \alpha_j e^{\epsilon\omega j} \right) \right]^{-1} \\ &= \frac{\sigma^2}{2\pi} \prod_{j=1}^p \frac{1}{(1 - 2m_j \cos \omega + m_j^2)}. \end{aligned} \quad (4.3.8)$$

**Proof.** Apply Theorem 4.3.1 to the finite moving average time series

$$e_t = \sum_{j=0}^p \alpha_j X_{t-j}$$

to obtain

$$(2\pi)^{-1} \sigma^2 = f_e(\omega) = T(\omega) f_X(\omega),$$

where

$$\begin{aligned} T(\omega) &= \left( \sum_{j=0}^p \alpha_j e^{-\epsilon\omega j} \right) \left( \sum_{j=0}^p \alpha_j e^{\epsilon\omega j} \right) \\ &= \left( e^{\epsilon\omega p} \sum_{j=0}^p \alpha_j e^{-\epsilon\omega j} \right) \left( e^{-\epsilon\omega p} \sum_{j=0}^p \alpha_j e^{\epsilon\omega j} \right) \\ &= R(e^{\epsilon\omega}) R(e^{-\epsilon\omega}). \end{aligned}$$

Neither  $R(e^{\epsilon\omega})$  nor  $R(e^{-\epsilon\omega})$  can be zero, since the roots of  $R(m)$  cannot have an absolute value of one. Therefore,  $T(\omega)$  is not equal to zero for any  $\omega$ , and we are able to write

$$f_X(\omega) = f_e(\omega) T^{-1}(\omega).$$

This is the first representation given in (4.3.8). The alternative form for  $f_X(\omega)$  follows by writing  $T(\omega)$  in the factored form

$$\begin{aligned} T(\omega) &= \left| \prod_{j=1}^p (1 - e^{-\epsilon\omega} m_j) \right|^2 \\ &= \prod_{j=1}^p (1 - 2m_j \cos \omega + m_j^2). \end{aligned}$$
▲

We can also write (4.3.8) in the factored form

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \left[ \prod_{j=1}^p (e^{-\omega} - m_j) \prod_{j=1}^p (e^{\omega} - m_j) \right]^{-1}. \quad (4.3.9)$$

The spectral density of the autoregressive moving average process is sometimes called a rational spectral density because it is the ratio of polynomials in  $e^{\omega}$ .

**Corollary 4.3.1.3.** Let an autoregressive moving average process of order  $(p, q)$  be defined by

$$\sum_{j=0}^p \alpha_j X_{t-j} = \sum_{k=0}^q \beta_k e_{t-k},$$

where  $\alpha_0 = \beta_0 = 1$ ,  $\alpha_p \neq 0$ ,  $\beta_q \neq 0$ , the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables, and the roots  $\{m_j: j = 1, 2, \dots, p\}$  of

$$R(m) = \sum_{j=0}^p \alpha_j m^{p-j} = 0$$

are less than one in absolute value. Then the spectral density of  $\{X_t\}$  is given by

$$\begin{aligned} f_X(\omega) &= \frac{\sigma^2}{2\pi} \frac{\left( \sum_{k=0}^q \beta_k e^{\omega k} \right) \left( \sum_{k=0}^q \beta_k e^{-\omega k} \right)}{\left( \sum_{j=0}^p \alpha_j e^{\omega j} \right) \left( \sum_{j=0}^p \alpha_j e^{-\omega j} \right)} \\ &= \frac{\sigma^2}{2\pi} \frac{\prod_{k=1}^q (e^{\omega} - r_k) \prod_{k=1}^q (e^{-\omega} - r_k)}{\prod_{j=1}^p (e^{\omega} - m_j) \prod_{j=1}^p (e^{-\omega} - m_j)}, \end{aligned} \quad (4.3.10)$$

where  $\{r_k: k = 1, 2, \dots, q\}$  are the roots of

$$\sum_{k=0}^q \beta_k r^{q-k} = 0.$$

**Proof.** Reserved for the reader. ▲

In Theorem 2.6.4 we demonstrated that a finite moving average process with no roots of unit absolute value always has a representation with characteristic equation whose roots are less than one in absolute value. In Theorem 2.6.3 we obtained a moving average representation for a time series with  $|\rho(1)| < 0.5$  and  $\rho(h) = 0$ ,  $h = \pm 2, \pm 3, \dots$ . We can now extend Theorem 2.6.3 to time series with a finite number of nonzero covariances.

**Theorem 4.3.2.** Let the stationary time series  $X_t$  have zero mean and spectral density

$$f(\omega) = \frac{1}{2\pi} \sum_{h=-q}^q \gamma(h) e^{-i\omega h},$$

where  $f(\omega)$  is strictly positive,  $\gamma(h)$  is the covariance function of  $X_t$ , and  $\gamma(q) \neq 0$ . Then  $X_t$  has the representation

$$X_t = \sum_{j=1}^q \beta_j e_{t-j} + e_t = \sum_{j=0}^q \beta_j e_{t-j},$$

where  $\beta_0 = 1$ , the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables with  $\sigma^2 = \gamma(0)[\sum_{j=0}^q \beta_j^2]^{-1}$ , and the roots of

$$m^q + \sum_{j=1}^q \beta_j m^{q-j} = 0$$

are less than one in absolute value.

**Proof.** The function

$$A(z) = z^q [\gamma(q)]^{-1} \sum_{h=-q}^q \gamma(h) z^h = [\gamma(q)]^{-1} \sum_{h=0}^{2q} \gamma(h-q) z^h$$

is a polynomial of degree  $2q$  in  $z$  with unit coefficient for  $z^{2q}$ . Therefore, it can be written in the factored form

$$A(z) = \prod_{r=1}^{2q} (z - m_r),$$

where the  $m_r$  are the roots of  $A(z) = 0$ . If  $A(m_r) = 0$ , then

$$\sum_{h=-q}^q \gamma(h) m_r^h = 0,$$

and by symmetry

$$\sum_{h=-q}^q \gamma(h) m_r^{-h} = 0.$$

Since  $f(\omega)$  is strictly positive, none of the roots are of unit absolute value. Also, the coefficients are real, so that any complex roots occur as conjugate pairs. Therefore, we can arrange the roots in  $q$  pairs,  $(m_r, m_{-r}) = (m_r, m_r^{-1})$ , where multiple roots are repeated the proper number of times. We let  $m_r$ ,  $r = 1, 2, \dots, q$ , denote the roots less than one in absolute value. Using these roots, we define  $\beta_j$ , for

$j = 0, 1, 2, \dots, q$ , by

$$\prod_{r=1}^q (z - m_r) = \sum_{j=0}^q \beta_j z^{q-j}.$$

Therefore,

$$\begin{aligned} z^{-q} \gamma(q) A(z) &= z^{-q} \gamma(q) \prod_{r=1}^q (z - m_r) \prod_{r=1}^q (z - m_r^{-1}) \\ &= \gamma(q) \prod_{r=1}^q (z - m_r) \prod_{r=1}^q (1 - z^{-1} m_r^{-1}) \\ &= \gamma(q) \left( \sum_{j=0}^q \beta_j z^j \right) \left( \sum_{j=0}^q \beta_j z^{-j} \right) \prod_{r=1}^q (-m_r)^{-1}. \end{aligned}$$

It follows that the expression in  $e^{-\epsilon\omega}$  defining  $f(\omega)$  can be written as

$$f(\omega) = 2\pi\sigma^2 |f_\beta(\omega)|^2,$$

where  $f_\beta(\omega)$  is the Fourier transform of  $\{\beta_j\}$  and  $\sigma^2 = \gamma(0)(\sum_{j=0}^q \beta_j^2)^{-1}$ . Now define the sequence of random variables  $\{e_i\}$  by  $e_i = \sum_{j=0}^\infty c_j X_{i-j}$ , where the sequence  $\{c_j\}$  is defined in Theorem 2.6.2. These  $c_j$  are such that  $X_i = \sum_{j=0}^q \beta_j e_{i-j}$ , which means that  $|f_c(\omega)f_\beta(\omega)|^2 = (2\pi)^{-4}$ . Therefore, the spectral density of  $e_i$  is

$$\begin{aligned} f_e(\omega) &= (2\pi)^2 |f_c(\omega)|^2 f(\omega) \\ &= (2\pi)^2 |f_c(\omega)|^2 \left| \frac{\sigma^2 |2\pi f_\beta(\omega)|^2}{2\pi} \right| \\ &= \frac{\sigma^2}{2\pi}, \end{aligned}$$

and the  $e_i$  are uncorrelated with variance  $\sigma^2$ . ▲

By our earlier results in Fourier series we know that a continuous periodic function can be approximated by a trigonometric polynomial. This means that a time series with a continuous spectral density is "very nearly" a moving average process and also "very nearly" an autoregressive process.

**Theorem 4.3.3.** Let  $g(\omega)$  be a nonnegative even continuous periodic function of period  $2\pi$ . Then, given an  $\epsilon > 0$ , there is a time series with the representation

$$X_t = \sum_{j=0}^q \beta_j e_{t-j}$$

such that  $|f_X(\omega) - g(\omega)| < \epsilon$  for all  $\omega \in [-\pi, \pi]$  where  $\beta_0 = 1$ ,  $q$  is a finite

integer, the  $e_i$  are uncorrelated  $(0, \sigma^2)$  random variables, and

$$\sigma^2 = \left( \sum_{j=0}^q \beta_j^2 \right)^{-1} \int_{-\pi}^{\pi} g(\omega) d\omega.$$

**Proof.** The result is trivial if  $g(\omega) = 0$ ; hence, we assume  $g(\omega) > 0$  for some  $\omega$ . Given  $\epsilon > 0$ , let  $\delta = 2\epsilon G(4\pi M + 3G)^{-1}$  and set

$$C(\omega) = \begin{cases} g(\omega) & \text{if } g(\omega) > \delta, \\ \delta & \text{otherwise,} \end{cases}$$

where

$$M = \max_{\omega} g(\omega)$$

$$G = \int_{-\pi}^{\pi} g(\omega) d\omega.$$

Then by Theorem 3.1.10, there is a  $q$  such that  $|f_Y(\omega) - C(\omega)| < \frac{1}{2}\delta$  for all  $\omega \in [-\pi, \pi]$ , where

$$f_Y(\omega) = \frac{1}{2\pi} \sum_{h=-q}^q \gamma(h) e^{-\epsilon \omega h},$$

$$\gamma(h) = q^{-1}(q - |h|)a_h,$$

and

$$a_h = \int_{-\pi}^{\pi} e^{\epsilon \omega h} C(\omega) d\omega.$$

By Exercise 4.14,  $\gamma(h)$  is positive semidefinite, and  $f_Y(\omega)$  is of the same form as the spectral density in Theorem 4.3.2. Therefore, we may write

$$f_Y(\omega) = (2\pi)^{-1} \tau^2 \left| \sum_{j=0}^q \beta_j e^{-\epsilon \omega j} \right|^2,$$

where  $\beta_0 = 1$ , the roots of  $\sum_{j=0}^q \beta_j m^{q-j}$  are less than one in absolute value, and  $\tau^2 = [\sum_{j=0}^q \beta_j^2]^{-1} \int_{-\pi}^{\pi} C(\omega) d\omega$ . Setting

$$f_X(\omega) = (2\pi)^{-1} \sigma^2 \left| \sum_{j=0}^q \beta_j e^{-\epsilon \omega j} \right|^2,$$

where

$$\sigma^2 = \tau^2 \left[ \int_{-\pi}^{\pi} f_Y(\omega) d\omega \right]^{-1} \int_{-\pi}^{\pi} g(\omega) d\omega,$$

we obtain the desired result. ▲

**Theorem 4.3.4.** Let  $g(\omega)$  be a nonnegative even continuous periodic function of period  $2\pi$ . Then, given an  $\epsilon > 0$ , there is a time series with the representation

$$\sum_{j=0}^p \alpha_j Y_{t-j} = e_t$$

such that  $|f_y(\omega) - g(\omega)| < \epsilon$  for all  $\omega \in [-\pi, \pi]$ , where  $\alpha_0 = 1$ ,  $p$  is a finite integer, the roots of

$$\sum_{j=0}^p \alpha_j m^{p-j} = 0$$

are less than one in absolute value, and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables.

**Proof.** The result is trivial if  $g(\omega) = 0$ . Assuming  $g(\omega) > 0$  for some  $\omega$  define

$$d(\omega) = \begin{cases} g^{-1}(\omega) & \text{if } g(\omega) > \frac{1}{2}\epsilon, \\ 2\epsilon^{-1} & \text{otherwise.} \end{cases}$$

Also define  $G = \max g(\omega)$ , and let  $0 < \delta < \epsilon[G(2G + \epsilon)]^{-1}$ . Then, by Theorem 3.1.10, there is a finite  $p$  such that

$$\left| \sum_{h=-p}^p C_h e^{-\epsilon \omega h} - d(\omega) \right| < \delta$$

for all  $\omega \in [-\pi, \pi]$ . By Theorem 4.3.2,

$$\sum_{h=-p}^p C_h e^{-\epsilon \omega h} = M \sum_{j=0}^p \alpha_j e^{-\epsilon \omega j} \sum_{j=0}^p \alpha_j e^{\epsilon \omega j},$$

where  $\alpha_0 = 1$ ,  $M$  is a constant, and the roots of  $\sum_{j=0}^p \alpha_j m^{p-j} = 0$  are less than one in absolute value. Hence, by defining

$$f_y(\omega) = M^{-1} \left[ \left( \sum_{j=0}^p \alpha_j e^{-\epsilon \omega j} \right) \left( \sum_{j=0}^p \alpha_j e^{\epsilon \omega j} \right) \right]^{-1}$$

and  $\sigma^2 = 2\pi M^{-1}$ , we have the conclusion. ▲

Operations on a time series are sometimes called *filtering*. In some areas, particularly in electronics, physical devices are constructed to filter time series. Examples include the modification of radio waves and radar signals. The creation of a linearly weighted sum is the application of a *linear filter*, the weights being identified as the filter. If, as in (4.3.0), the weights are constant over time, the filter

is called *time invariant*.<sup>3</sup> Obviously such operations change the behavior of the time series, and in practice this is often the objective. By investigating the properties of the filter, one is able to make statements about the change in behavior.

It is clear that moving average and autoregressive processes are obtained from a white noise process by the application of a linear time invariant filter. Thus, using this terminology, we have been studying the effects of a linear time invariant filter on the spectral density of a time series.

To further illustrate the effects of the application of a linear filter, consider the simple time series

$$\begin{aligned} X_t &= 2\alpha \cos(\omega t + \beta) \\ &= \alpha[e^{i(\omega t + \beta)} + e^{-i(\omega t + \beta)}] \end{aligned}$$

and apply the absolutely summable linear filter  $a_j$  to obtain

$$\begin{aligned} Y_t &= \alpha \sum_{j=-\infty}^{\infty} a_j [e^{i(\omega(t-j) + \beta)} + e^{-i(\omega(t-j) + \beta)}] \\ &= \alpha e^{i(\omega t + \beta)} \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j} + \alpha e^{-i(\omega t + \beta)} \sum_{j=-\infty}^{\infty} a_j e^{i\omega j}. \end{aligned}$$

We may express the Fourier transform of  $a_j$  in a general complex form

$$\begin{aligned} 2\pi f_a(\omega) &= \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j} \stackrel{\text{(say)}}{=} u(\omega) + i v(\omega) \\ &= \psi(\omega)[\cos \varphi(\omega) + i \sin \varphi(\omega)] = \psi(\omega)e^{i\varphi(\omega)}, \end{aligned} \quad (4.3.11)$$

where

$$\begin{aligned} \psi(\omega) &= \sqrt{[u(\omega)]^2 + [v(\omega)]^2}, \\ \varphi(\omega) &= \tan^{-1} \left[ \frac{v(\omega)}{u(\omega)} \right]. \end{aligned}$$

The reader should note the convention used in defining the sign of  $v(\omega)/u(\omega)$ :  $v(\omega)$  is the coefficient of  $i$  (not  $-i$ ) in the original transform.

In this notation, the filtered cosine wave becomes

$$\begin{aligned} Y_t &= \alpha \psi(\omega) \{e^{i[\omega t + \beta + \varphi(\omega)]} + e^{-i[\omega t + \beta + \varphi(\omega)]}\} \\ &= 2\alpha \psi(\omega) \cos\{\omega t + \beta + \varphi(\omega)\}. \end{aligned}$$

Thus, the process  $Y_t$  is also a cosine function with the original period, but with amplitude  $2\alpha \psi(\omega)$  and the phase angle  $\beta + \varphi(\omega)$ . The function  $\psi(\omega)$  is called the

<sup>3</sup> We may define the time invariant property as follows. If the output of a filter  $g$  applied to  $X_t$  is  $Y_t$ , then the filter is time invariant if  $g(X_{t+h}) = Y_{t+h}$  for all integer  $h$ .

gain of the filter, and the function  $\varphi(\omega)$  is the *phase angle* (or simply *phase*) of the filter. The rationale for these terms is clear when one observes the effect of the filter on the cosine time series. The transform  $2\pi f_a(\omega)$  is called the *frequency response function* or the *transfer function* of the filter. From (4.3.1) we know that the spectrum of a time series  $Y$ , created by linearly filtering a time series  $X$ , is the spectrum of the original series,  $f_X(\omega)$ , multiplied by the squared gain of the filter. The squared gain is sometimes called the *power transfer function*.

A particularly simple filter is the perfect *delay*. Let  $\{X_t: t \in (0, \pm 1, \pm 2, \dots)\}$  be a stationary time series. The time series delayed or lagged by  $\tau$  periods is defined by

$$Y_t = X_{t-\tau},$$

where the filter is defined by  $a_\tau = 1$  and  $a_j = 0$ ,  $j \neq \tau$ . Hence the frequency response function of the filter is  $e^{-\epsilon\omega\tau} = \cos \omega\tau - i \sin \omega\tau$ , the gain is  $(\cos^2 \omega\tau + \sin^2 \omega\tau)^{1/2} = 1$ , and the phase angle is  $\tan^{-1}(-\sin \omega\tau / \cos \omega\tau) = -\omega\tau$ . A cosine wave of frequency  $\omega$  and corresponding period  $2\pi\omega^{-1}$  completes  $(2\pi)^{-1}\omega\tau$  cycles during a "time period" of length  $\tau$ . Thus the phase of such a cosine wave lagged  $\tau$  periods is shifted by the quantity  $(2\pi)^{-1}\omega\tau$ .

Theorem 4.3.1 is a very useful theorem for absolutely summable filters. In some situations, a mean square result applicable to a broader range of filters and time series is appropriate.

**Theorem 4.3.5.** Assume  $X$  is a zero mean stationary time series with spectral density  $f_X(\omega)$ . Assume  $\{c_j\}_{j=-\infty}^\infty$  is a square summable sequence and that

$$\sum_{j=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} c_r c_{j+r} \gamma_X(j+h-r) \quad (4.3.12)$$

converges absolutely. Define

$$Y_t = \sum_{j=-\infty}^{\infty} c_j X_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots$$

Then

$$\gamma_Y(h) = \int_{-\pi}^{\pi} e^{i\omega h} \left| \sum_{j=-\infty}^{\infty} c_j e^{-i\omega j} \right|^2 f_X(\omega) d\omega$$

and

$$f_Y(\omega) = (2\pi)^2 f_X(\omega) f_c(\omega) f_c^*(\omega), \quad (4.3.13)$$

where

$$[f_c(\omega), f_c^*(\omega)] = (2\pi)^{-1} \left[ \sum_{j=-\infty}^{\infty} c_j e^{-i\omega j}, \sum_{j=-\infty}^{\infty} c_j e^{i\omega j} \right].$$

**Proof.** Because (4.3.12) converges, we can interchange orders of integration and summation and we have

$$\begin{aligned} E\{Y_t Y_{t+h}\} &= E\left\{\sum_{j=-\infty}^{\infty} c_j X_{t-j} \sum_{r=-\infty}^{\infty} c_r X_{t+h-r}\right\} \\ &= \sum_{j=-\infty}^{\infty} \sum_{r=-\infty}^{\infty} c_j c_r \int_{-\pi}^{\pi} e^{i\omega(j+h-r)} f_X(\omega) d\omega \\ &= \int_{-\pi}^{\pi} \sum_{j=-\infty}^{\infty} c_j e^{i\omega j} \sum_{r=-\infty}^{\infty} c_r e^{-i\omega r} e^{i\omega h} f_X(\omega) d\omega \\ &= \int_{-\pi}^{\pi} e^{i\omega h} \left| \sum_{j=-\infty}^{\infty} c_j e^{i\omega j} \right|^2 f_X(\omega) d\omega. \end{aligned}$$

Because of the one-to-one correspondence in mean square between the Fourier coefficients of a square integrable function and the function,  $f_Y(\omega)$  is the function specified in (4.3.13).  $\Delta$

In Section 2.11, we introduced the long memory processes that satisfy

$$(1 - \mathcal{B})^d Y_t = Z_t \quad (4.3.14)$$

or

$$Y_t = (1 - \mathcal{B})^{-d} Z_t, \quad (4.3.15)$$

where  $Z_t$  is a finite autoregressive moving average. By Theorem 3.1.6, there is a function  $f_Y(\omega)$  defined as a limit in mean square,

$$f_Y(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma_Y(h) e^{-i\omega h},$$

because  $\gamma_Y(h)$  is square summable. Also,

$$\gamma_Y(h) = \int_{-\pi}^{\pi} f_Y(\omega) e^{i\omega h} d\omega.$$

Consider the simple case in which the  $Z_t$  of (4.3.14) is  $e_t$ , a time series of uncorrelated  $(0, \sigma^2)$  random variables. Then, by Theorem 4.3.5,

$$\begin{aligned} f_Y(\omega) &= (1 - e^{-i\omega})^{-d} (1 - e^{i\omega})^{-d} f_e(\omega) \\ &= (2\pi)^{-1} |1 - e^{i\omega}|^{-2d} \sigma^2 \\ &= (2\pi)^{-1} [4 \sin^2(0.5\omega)]^{-d} \sigma^2. \end{aligned}$$

Observe that  $f_Y(\omega)$  is unbounded at  $\omega = 0$  when  $d > 0$ . Because  $x^{-1} \sin x \rightarrow 1$  as  $x \rightarrow 0$ ,

$$f_Y(\omega) \rightarrow (2\pi)^{-1} \omega^{-2d} \sigma^2$$

as  $\omega \rightarrow 0$ . Sometimes a long memory process is defined to be a process whose spectral density is approximately equal to a multiple of  $\omega^{-2d}$  for  $-0.5 < d < 0.5$  and  $\omega$  near zero.

It follows from the spectral density that

$$\begin{aligned}\gamma_r(h) &= \int_{-\pi}^{\pi} e^{i\omega h} f_r(\omega) d\omega \\ &= \frac{(-1)^h \Gamma(1-2d)}{\Gamma(h-d+1)\Gamma(1-h-d)} \sigma^2, \quad h = 0, 1, 2, \dots.\end{aligned}$$

#### 4.4. VECTOR PROCESSES

The spectral representation of vector time series follows in a straightforward manner from that of scalar time series. We denote the cross covariance of two zero mean stationary time series  $X_{jt}$  and  $X_{mt}$  by  $\gamma_{jm}(h) = \{\Gamma(h)\}_{jm} = E\{X_{jt} X_{m,t+h}\}$  and assume that  $\{\gamma_{jm}(h)\}$  is absolutely summable. Then

$$\frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{jm}(h) e^{-i\omega h} = f_{jm}(\omega) \quad (4.4.1)$$

is a continuous periodic function of  $\omega$ , which we call the *cross spectral function* of  $X_{jt}$  and  $X_{mt}$ . Since  $\gamma_{jm}(h)$  may not be symmetric about 0,  $f_{jm}(\omega)$  is, in general, a complex valued function. As such, it can be written as

$$f_{jm}(\omega) = c_{jm}(\omega) - i q_{jm}(\omega),$$

where  $c_{jm}(\omega)$  and  $q_{jm}(\omega)$  are real valued functions of  $\omega$ . The function  $c_{jm}(\omega)$  is called the *coincident spectral density* or simply the *cospectrum*. The function  $q_{jm}(\omega)$  is called the *quadrature spectral density*. The function  $c_{jm}(\omega)$  is the cosine portion of the transform and is an even function of  $\omega$ , and  $q_{jm}(\omega)$  is the sine portion and is an odd function of  $\omega$ . Thus we may define these quantities as the transforms

$$\begin{aligned}c_{jm}(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \frac{1}{2} [\gamma_{jm}(h) + \gamma_{jm}(-h)] e^{-i\omega h}, \\ -i q_{jm}(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \frac{1}{2} [\gamma_{jm}(h) - \gamma_{jm}(-h)] e^{-i\omega h},\end{aligned}$$

or, in real terms,

$$\begin{aligned}c_{jm}(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \frac{1}{2} [\gamma_{jm}(h) + \gamma_{jm}(-h)] \cos \omega h, \\ q_{jm}(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \frac{1}{2} [\gamma_{jm}(h) - \gamma_{jm}(-h)] \sin \omega h.\end{aligned} \quad (4.4.2)$$

By the Fourier integral theorem,

$$\gamma_{jm}(h) = \int_{-\pi}^{\pi} e^{i\omega h} f_{jm}(\omega) d\omega .$$

If we let  $\mathbf{f}(\omega)$  denote the matrix with typical element  $f_{jm}(\omega)$ , we have the matrix representations

$$\Gamma(h) = \int_{-\pi}^{\pi} e^{i\omega h} \mathbf{f}(\omega) d\omega , \quad (4.4.3)$$

$$\mathbf{f}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} e^{-i\omega h} \Gamma(h) . \quad (4.4.4)$$

For a general stationary time series we can write

$$\Gamma(h) = \int_{-\pi}^{\pi} e^{i\omega h} d\mathbf{F}(\omega) \quad (4.4.5)$$

in complete analogy to (4.1.1a).

Let us investigate some of the properties of the matrix  $\mathbf{f}(\omega)$ .

**Definition 4.4.1.** A square complex valued matrix  $\mathbf{B}$  is called a *Hermitian matrix* if it is equal to its conjugate transpose, that is,

$$\mathbf{B} = \mathbf{B}^* ,$$

where the  $jm$ th element of  $\mathbf{B}^*$  is the complex conjugate of the  $mj$ th element of  $\mathbf{B}$ .

**Definition 4.4.2.** A Hermitian matrix  $\mathbf{B}$  is *positive definite* if for any complex vector  $\mathbf{w}$  such that  $\mathbf{w}^* \mathbf{w} > 0$ ,

$$\mathbf{w}^* \mathbf{B} \mathbf{w} > 0 ,$$

and it is *positive semidefinite* if

$$\mathbf{w}^* \mathbf{B} \mathbf{w} \geq 0 .$$

**Lemma 4.4.1.** For stationary vector time series of dimension  $k$  satisfying

$$\sum_{h=-\infty}^{\infty} |\gamma_{jm}(h)| < \infty$$

for  $j, m = 1, 2, \dots, k$ , the matrix  $\mathbf{f}(\omega)$  is a positive semidefinite Hermitian matrix for all  $\omega$  in  $[-\pi, \pi]$ .

**Proof.** The matrix  $\mathbf{f}(\omega)$  is Hermitian, since, for all  $\omega$  in  $[-\pi, \pi]$  and for

$j, m = 1, 2, \dots, k$ ,

$$\begin{aligned} f_{jm}^*(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{jm}^*(h) e^{-\epsilon\omega h} \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{mj}(-h) e^{\epsilon\omega h} \\ &= \frac{1}{2\pi} \sum_{r=-\infty}^{\infty} \gamma_{mj}(r) e^{-\epsilon\omega r} \\ &= f_{mj}(\omega). \end{aligned}$$

Consider the complex valued time series

$$Z_t = \alpha' X_t = \sum_{j=1}^k \alpha_j X_{jt},$$

where  $\alpha' = (\alpha_1, \alpha_2, \dots, \alpha_k)$  is a vector of arbitrary complex numbers. The autocovariance function of  $Z_t$  is given by

$$\gamma_Z(h) = \sum_{j=1}^k \sum_{r=1}^k \alpha_j \alpha_r^* \gamma_{jr}(h).$$

Now  $\gamma_Z(h)$  is positive semidefinite, and hence, for any  $n$ ,

$$\sum_{m=1}^n \sum_{q=1}^n \gamma_Z(m-q) e^{-\epsilon m \omega} e^{\epsilon q \omega} \geq 0$$

and

$$\sum_{h=-(n-1)}^{n-1} \frac{n-|h|}{n} \left[ \sum_{j=1}^k \sum_{r=1}^k \alpha_j \alpha_r^* \gamma_{jr}(h) \right] e^{-\epsilon \omega h} \geq 0.$$

Taking the limit as  $n \rightarrow \infty$ , we have

$$\sum_{j=1}^k \sum_{r=1}^k \alpha_j \alpha_r^* f_{jr}(\omega) \geq 0,$$

which establishes that  $f(\omega)$  is positive semidefinite. ▲

It follows immediately from Lemma 4.4.1 that the determinant of any two by two matrix of the form

$$\begin{pmatrix} f_{jj}(\omega) & f_{jm}(\omega) \\ f_{mj}(\omega) & f_{mm}(\omega) \end{pmatrix}$$

is nonnegative. Hence,

$$\begin{aligned} |f_{jm}(\omega)|^2 &= f_{jm}(\omega)f_{jm}^*(\omega) = f_{jm}(\omega)f_{mj}(\omega) \\ &\leq f_{jj}(\omega)f_{mm}(\omega). \end{aligned} \quad (4.4.6)$$

The quantity

$$\mathcal{K}_{jm}^2(\omega) = \frac{|f_{jm}(\omega)|^2}{f_{jj}(\omega)f_{mm}(\omega)} = \frac{c_{jm}^2(\omega) + q_{jm}^2(\omega)}{f_{jj}(\omega)f_{mm}(\omega)} \quad (4.4.7)$$

is called the *squared coherency function*. The spectral density may be zero at certain frequencies, in which case  $\mathcal{K}_{jm}^2(\omega)$  is of the form 0/0. We adopt the convention of assigning zero to the coherency in such situations. The inequality (4.4.6), sometimes written as

$$\mathcal{K}_{jm}^2(\omega) \leq 1, \quad (4.4.8)$$

is called the *coherency inequality*.

To further appreciate the properties of  $f(\omega)$ , consider the time series

$$X_1 = A_1 \cos rt + B_1 \sin rt, \quad (4.4.9)$$

$$X_2 = A_2 \cos rt + B_2 \sin rt,$$

where  $r \in (0, \pi)$  and  $(A_1, B_1, A_2, B_2)'$  is distributed as a multivariate normal with zero mean and covariance matrix

$$\begin{pmatrix} \sigma_{11} & 0 & \sigma_{13} & \sigma_{14} \\ 0 & \sigma_{11} & -\sigma_{14} & \sigma_{13} \\ \sigma_{13} & -\sigma_{14} & \sigma_{33} & 0 \\ \sigma_{14} & \sigma_{13} & 0 & \sigma_{33} \end{pmatrix}.$$

Then

$$\gamma_{11}(h) = \sigma_{11} \cos rh,$$

$$\gamma_{22}(h) = \sigma_{33} \cos rh,$$

$$\begin{aligned} \gamma_{12}(h) &= E\{A_1 A_2 [\cos rt] \cos r(t+h) + A_1 B_2 [\cos rt] \sin r(t+h) \\ &\quad + B_1 A_2 [\sin rt] \cos r(t+h) + B_1 B_2 [\sin rt] \sin r(t+h)\} \\ &= \sigma_{13} \cos rh + \sigma_{14} \sin rh, \end{aligned}$$

and

$$\gamma_{21}(h) = \gamma_{12}(-h).$$

The matrix analog of the spectral distribution function introduced in Section 4.1

is

$$\mathbf{F}(\omega) = \begin{pmatrix} F_{11}(\omega) & F_{12}(\omega) \\ F_{21}(\omega) & F_{22}(\omega) \end{pmatrix},$$

where  $F_{12}(\omega)$  is the cross spectral distribution function. For the example (4.4.9),  $F_{11}(\omega)$  is a step function with jumps of height  $\frac{1}{2}\sigma_{11}$  at  $\pm r$ ,  $F_{22}(\omega)$  is a step function with jumps of height  $\frac{1}{2}\sigma_{33}$  at  $\pm r$ ,  $F_{12}(\omega)$  is a complex valued function where  $\operatorname{Re} F_{12}(\omega)$  is a step function with a jump of height  $\frac{1}{2}\sigma_{13}$  at  $\pm r$ , and  $\operatorname{Im} F_{12}(\omega)$  is a step function with a jump of  $\frac{1}{2}\sigma_{14}$  at  $-r$  and a jump of  $-\frac{1}{2}\sigma_{14}$  at  $r$ . Since the elements of  $\mathbf{F}(\omega)$  are pure jump functions, we have a pure line spectrum. The real portion of the cross line spectrum is one-half the covariance between the coefficients of the cosine portions of the two time series, which is also one-half the covariance between the coefficients of the sine portions. The absolute value of the imaginary portion of the cross line spectrum is one-half the covariance between the coefficient of the cosine portion of the first time series and the coefficient of the sine portion of the second time series. This is one-half the negative of the covariance between the coefficient of the sine of the first time series and the coefficient of the cosine of the second. To consider the point further, we write the time series  $X_{2t}$  as a sine wave,

$$X_{2t} = \psi \sin(rt + \varphi),$$

where

$$\psi = (A_2^2 + B_2^2)^{1/2},$$

$$\varphi = \tan^{-1} \frac{A_2}{B_2}.$$

Let  $X_{3t}$  be the cosine with the same amplitude and phase,

$$\begin{aligned} X_{3t} &= \psi \cos(rt + \varphi) \\ &= B_2 \cos rt - A_2 \sin rt. \end{aligned}$$

It follows that  $X_{2t}$  is uncorrelated with  $X_{3t}$ . The covariance between  $X_{1t}$  and  $X_{2t}$  is

$$\begin{aligned} E\{X_{1t}X_{2t}\} &= E\{(A_1 \cos rt + B_1 \sin rt)(A_2 \cos rt + B_2 \sin rt)\} \\ &= \sigma_{13}(\cos^2 rt + \sin^2 rt) = \sigma_{13}, \end{aligned}$$

and the covariance between  $X_{1t}$  and  $X_{3t}$  is

$$\begin{aligned} E\{X_{1t}X_{3t}\} &= E\{(A_1 \cos rt + B_1 \sin rt)(B_2 \cos rt - A_2 \sin rt)\} \\ &= \sigma_{14}. \end{aligned}$$

The covariance between  $X_{1t}$  and  $X_{2t}$  is proportional to the real part of the cross spectrum at  $r$ . The covariance between  $X_{1t}$  and  $X_{3t}$  is proportional to the imaginary

portion of the cross spectrum at  $r$ . The fact that  $X_3$ , is  $X_2$ , "shifted" by a phase angle of  $\pi/2$  explains why the real portion of the cross spectrum is called the cospectrum and the imaginary part is called the quadrature spectrum. The squared coherency is the multiple correlation between  $X_{1t}$  and the pair  $X_{2t}, X_{3t}$ , that is,

$$\mathcal{H}_{12}^2(r) = \frac{\sigma_{13}^2 + \sigma_{14}^2}{\sigma_{33}\sigma_{11}}.$$

We now introduce some cross spectral quantities useful in the analysis of input-output systems. Let the bivariate time series  $(X_t, Y_t)'$  have absolutely summable covariance function. We may then write the cross spectral function as

$$f_{XY}(\omega) = A_{XY}(\omega)e^{i\varphi_{XY}(\omega)}, \quad (4.4.10)$$

where

$$\begin{aligned} \varphi_{XY}(\omega) &= \tan^{-1} \left[ \frac{-q_{XY}(\omega)}{c_{XY}(\omega)} \right], \\ A_{XY}(\omega) &= [c_{XY}^2(\omega) + q_{XY}^2(\omega)]^{1/2}. \end{aligned}$$

We use the convention of setting  $\varphi_{XY}(\omega) = 0$  when both  $c_{XY}(\omega)$  and  $q_{XY}(\omega)$  are zero. The quantity  $\varphi_{XY}(\omega)$  is called the *phase spectrum*, and  $A_{XY}(\omega)$  is called the *cross amplitude spectrum*. The *gain* of  $Y_t$  over  $X_t$  is defined by

$$\psi_{XY}(\omega) = \frac{A_{XY}(\omega)}{f_{XX}(\omega)} \quad (4.4.11)$$

for those  $\omega$  where  $f_{XX}(\omega) > 0$ .

Let us assume that an absolutely summable linear filter is applied to an input time series  $X_t$  with zero mean, absolutely summable covariance function, and everywhere positive spectral density to yield an output time series

$$Y_t = \sum_{j=-\infty}^{\infty} a_j X_{t-j}. \quad (4.4.12)$$

The cross covariance function is

$$\gamma_{XY}(h) = E\{X_t Y_{t+h}\} = \sum_{j=-\infty}^{\infty} a_j \gamma_{XX}(h-j), \quad h = 0, \pm 1, \pm 2, \dots,$$

and, by Corollary 3.4.1.1, the cross spectral function is

$$f_{XY}(\omega) = 2\pi f_a(\omega) f_{XX}(\omega), \quad (4.4.13)$$

where

$$f_a(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j},$$

$$f_{XX}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{XX}(h) e^{-i\omega h}.$$

It follows that  $f_{XY}(\omega)/f_{XX}(\omega)$  is the transfer function of the filter.

Recall that the phase  $\varphi(\omega)$  and gain  $\psi(\omega)$  of the filter  $\{a_j\}$  were defined by

$$2\pi f_a(\omega) = \psi(\omega) e^{i\varphi(\omega)}.$$

Since  $f_{XY}(\omega)$  is the product of  $2\pi f_a(\omega)$  and  $f_{XX}(\omega)$ , where  $f_{XX}(\omega)$  is a real function of  $\omega$ , it follows that the phase spectrum is the same as the phase of the filter. The cross amplitude spectrum is the product of the spectrum of the input time series and the gain of the filter. That is,

$$A_{XY}(\omega) = |f_{XY}(\omega)| = f_{XX}(\omega) |2\pi f_a(\omega)| \quad (4.4.14)$$

and the gain of  $Y_t$  over  $X_t$  is simply the gain of the filter.

By Theorem 4.3.1, the spectral density of  $Y_t$  is

$$f_{YY}(\omega) = (2\pi)^2 f_a(\omega) f_a^*(\omega) f_{XX}(\omega),$$

and it follows that the squared coherency, for  $|f_a(\omega)| > 0$ , is

$$\mathcal{H}_{XY}^2(\omega) = \frac{|f_{XY}(\omega)|^2}{f_{YY}(\omega) f_{XX}(\omega)} = \frac{|2\pi f_a(\omega) f_{XX}(\omega)|^2}{(2\pi)^2 |f_a(\omega)|^2 |f_{XX}(\omega)|^2} = 1. \quad (4.4.15)$$

This is an interesting result, because it shows that the squared coherency between an output time series created by the application of a linear filter and the original time series is one at all frequencies. The addition of an error (noise) time series to the output will produce a coherency less than one in a linear system. For example, consider the bivariate time series

$$X_{1t} = \beta X_{1,t-1} + e_{1t},$$

$$X_{2t} = \alpha_1 X_{1t} + \alpha_2 X_{1,t-1} + e_{2t}, \quad (4.4.16)$$

where  $|\beta| < 1$  and  $\{(e_{1t}, e_{2t})'\}$  is a sequence of uncorrelated vector random variables with  $E\{e_{1t}^2\} = \sigma_{11}$ ,  $E\{e_{2t}^2\} = \sigma_{22}$ , and  $E\{e_{1t} e_{2,t+h}\} = 0$  for all  $t$  and  $h$ .

The input time series  $X_{1t}$  is a first order autoregressive time series, and the output  $X_{2t}$  is a linear function of  $X_{1t}$ ,  $X_{1,t-1}$ , and  $e_{2t}$ . The autocovariance and cross

covariance functions are therefore easily computed:

$$\begin{aligned}\gamma_{11}(h) &= E\{X_{1,t} X_{1,t+h}\} = \frac{\beta^{|h|}}{1 - \beta^2} \sigma_{11}, \\ \gamma_{22}(h) &= \begin{cases} (\alpha_1^2 + \alpha_2^2)\gamma_{11}(0) + 2\alpha_1\alpha_2\gamma_{11}(1) + \sigma_{22}, & h = 0, \\ (\alpha_1^2 + \alpha_2^2)\gamma_{11}(h) + \alpha_1\alpha_2[\gamma_{11}(h+1) + \gamma_{11}(h-1)], & h \neq 0, \end{cases} \\ \gamma_{12}(h) &= E\{X_{1,t} X_{2,t+h}\} = \alpha_1\gamma_{11}(h) + \alpha_2\gamma_{11}(h-1).\end{aligned}$$

The cross covariance and cross correlation functions for parameters  $\alpha_1 = 0.5$ ,  $\alpha_2 = 1.0$ ,  $\gamma_{11}(0) = 1.0$ ,  $\sigma_{22} = 0.5$ , and  $\beta = 0.8$  are displayed in Table 4.4.1.

The special matrix has elements

$$\begin{aligned}f_{11}(\omega) &= g_{11}(\omega)(1 - \beta e^{-\beta\omega})^{-1}(1 - \beta e^{\beta\omega})^{-1}, \\ f_{22}(\omega) &= (\alpha_1 + \alpha_2 e^{-\beta\omega})(\alpha_1 + \alpha_2 e^{\beta\omega})f_{11}(\omega) + g_{22}(\omega), \\ f_{12}(\omega) &= (\alpha_1 + \alpha_2 e^{-\beta\omega})f_{11}(\omega), \\ f_{21}(\omega) &= f_{12}^*(\omega) = (\alpha_1 + \alpha_2 e^{\beta\omega})f_{11}(\omega),\end{aligned}$$

where  $g_{11}(\omega)$  is the spectral density of  $e_{1t}$ , and  $g_{22}(\omega)$  is the spectral density of  $e_{2t}$ ; that is,

$$\begin{aligned}g_{11}(\omega) &= (2\pi)^{-1}\sigma_{11}, \\ g_{22}(\omega) &= (2\pi)^{-1}\sigma_{22}.\end{aligned}$$

**Table 4.4.1. Autocovariance and Cross Covariance Functions of the Time Series Defined in (4.4.16)**

$h$	$\gamma_{11}(h)$	$\gamma_{22}(h)$	$\gamma_{12}(h)$	$\rho_{12}(h)$
-6	0.262	0.596	0.341	0.213
-5	0.328	0.745	0.426	0.267
-4	0.410	0.932	0.532	0.333
-3	0.512	1.165	0.666	0.417
-2	0.640	1.456	0.832	0.521
-1	0.800	1.820	1.040	0.651
0	1.000	2.550	1.300	0.814
1	0.800	1.820	1.400	0.877
2	0.640	1.456	1.120	0.701
3	0.512	1.165	0.896	0.561
4	0.410	0.932	0.717	0.449
5	0.328	0.745	0.573	0.359
6	0.262	0.596	0.459	0.287

The cospectrum and quadrature spectral density are

$$\begin{aligned} c_{12}(\omega) &= (\alpha_1 + \alpha_2 \cos \omega)f_{11}(\omega), \\ q_{12}(\omega) &= \alpha_2 [\sin \omega]f_{11}(\omega), \end{aligned}$$

the phase spectrum and cross amplitude spectrum are

$$\begin{aligned} \varphi_{12}(\omega) &= \tan^{-1}\left(\frac{-\alpha_2 \sin \omega}{\alpha_1 + \alpha_2 \cos \omega}\right), \\ A_{12}(\omega) &= f_{11}(\omega)[\alpha_1^2 + 2\alpha_1\alpha_2 \cos \omega + \alpha_2^2]^{1/2}, \end{aligned}$$

and the squared coherency is

$$\begin{aligned} \mathcal{H}_{12}^2(\omega) &= \frac{|2\pi f_a(\omega)|^2 |f_{11}(\omega)|^2}{f_{11}(\omega)[|2\pi f_a(\omega)|^2 f_{11}(\omega) + g_{22}(\omega)]} \\ &= \frac{1}{1 + \eta(\omega)}, \end{aligned}$$

where

$$\eta(\omega) = \frac{g_{22}(\omega)}{|2\pi f_a(\omega)|^2 f_{11}(\omega)}.$$

Since  $g_{22}(\omega)$  is positive at all frequencies, the squared coherency is strictly less than one at all frequencies. The quantity  $\eta(\omega)$  is sometimes called the *noise to signal ratio* in physical applications. Although the presence of noise reduces the squared coherency, the ratio of the cross spectrum to the spectrum of the input series still gives the transfer function of the filter. This is because the time series  $e_{2t}$  is uncorrelated with the input  $X_{1t}$ . The quantity

$$f_{22}(\omega)[1 - \mathcal{H}_{12}^2(\omega)] = f_{22}(\omega) - f_{21}(\omega)f_{11}^{-1}(\omega)f_{12}(\omega)$$

is sometimes called the *error spectral density* or *error spectrum*. We see that for a model such as (4.4.16),

$$g_{22}(\omega) = f_{22}(\omega) - f_{21}(\omega)f_{11}^{-1}(\omega)f_{12}(\omega).$$

For the example of Table 4.4.1 the elements of the spectral matrix are

$$\begin{aligned} f_{11}(\omega) &= \frac{0.36}{2\pi(1 - 0.8e^{-i\omega})(1 - 0.8e^{i\omega})}, \\ f_{22}(\omega) &= \frac{(0.5 + 1.0e^{-i\omega})(0.5 + 1.0e^{i\omega})(0.36)}{2\pi(1 - 0.8e^{-i\omega})(1 - 0.8e^{i\omega})} + \frac{0.5}{2\pi} \\ &= \frac{1.23(1 - 0.17e^{-i\omega})(1 - 0.17e^{i\omega})}{2\pi(1 - 0.8e^{-i\omega})(1 - 0.8e^{i\omega})}, \end{aligned}$$

$$f_{12}(\omega) = \frac{(0.5 + 1.0e^{-\omega})(0.36)}{2\pi(1 - 0.8e^{-\omega})(1 - 0.8e^{\omega})}.$$

It is interesting that the spectral density of  $X_{2,t}$  is that of an autoregressive moving average (1, 1) process. Note that if  $X_{2,t}$  had been defined as the simple sum of  $X_{1,t}$  and  $e_{2,t}$ , the spectral density would also have been that of an autoregressive moving average (1, 1) process but with different parameters.

The noise to signal ratio is

$$\eta(\omega) = \frac{g_{22}(\omega)}{|2\pi f_a(\omega)|^2 f_{11}(\omega)} = \frac{(1 - 0.8e^{-\omega})(1 - 0.8e^{\omega})}{(0.72)(0.5 + e^{-\omega})(0.5 + e^{\omega})},$$

and the squared coherency is

$$\mathcal{H}_{12}^2(\omega) = \frac{0.45 + 0.36 \cos \omega}{1.27 - 0.44 \cos \omega}.$$

Let us consider the example a bit further. The input time series  $X_{1,t}$  is autocorrelated. Let us filter both the input and output with the same filter, choosing the filter so that the input time series becomes a sequence of uncorrelated random variables. Thus we define

$$\begin{aligned} X_{3,t} &= X_{1,t} - \beta X_{1,t-1} = e_{1,t}, \\ X_{4,t} &= X_{2,t} - \beta X_{2,t-1}, \end{aligned}$$

and it follows that

$$\begin{aligned} X_{4,t} &= \alpha_1(X_{1,t} - \beta X_{1,t-1}) + \alpha_2(X_{1,t-1} - \beta X_{1,t-2}) + e_{2,t} - \beta e_{2,t-1} \\ &= \alpha_1 e_{1,t} + \alpha_2 e_{1,t-1} + e_{2,t} - \beta e_{2,t-1}. \end{aligned}$$

The cross covariance function of  $X_{3,t}$  and  $X_{4,t}$  then has a particularly simple form:

$$\gamma_{34}(h) = \begin{cases} \alpha_1 \sigma_{11}, & h = 0, \\ \alpha_2 \sigma_{11}, & h = 1, \\ 0 & \text{otherwise.} \end{cases}$$

By transforming the input series to white noise, the cross covariance function is transformed into the coefficients of the function (or linear filter) that defines  $X_{4,t}$  as a function of  $X_{3,t}$ . The spectral matrix of  $X_{3,t}$  and  $X_{4,t}$  has elements

$$\begin{aligned} f_{33}(\omega) &= \frac{\sigma_{11}}{2\pi}, \\ f_{44}(\omega) &= (\alpha_1 + \alpha_2 e^{-\omega})(\alpha_1 + \alpha_2 e^{\omega}) \frac{\sigma_{11}}{2\pi} \\ &\quad + (1 - \beta e^{-\omega})(1 - \beta e^{\omega}) \frac{\sigma_{22}}{2\pi}, \\ f_{34}(\omega) &= (\alpha_1 + \alpha_2 e^{-\omega}) \frac{\sigma_{11}}{2\pi}. \end{aligned}$$

The reader may verify that

$$\varphi_{34}(\omega) = \tan^{-1} \left( \frac{-\alpha_2 \sin \omega}{\alpha_1 + \alpha_2 \cos \omega} \right)$$

and

$$A_{34}(\omega) = \frac{\sigma_{11}}{2\pi} (\alpha_1^2 + 2\alpha_1\alpha_2 \cos \omega + \alpha_2^2)^{1/2}.$$

As we expected from (4.4.11) and (4.4.13), the phase spectrum is unchanged by the transformation, since we transformed both input and output with the same filter. The filter changed the input spectrum, and as a result, that portion of the cross amplitude spectrum associated with the input was altered. Also, the error spectral density

$$f_{44}(\omega) - f_{43}(\omega)f_{33}^{-1}(\omega)f_{34}(\omega) = (1 - \beta e^{-\epsilon\omega})(1 - \beta e^{\epsilon\omega}) \frac{\sigma_{22}}{2\pi}$$

is that of a moving average with parameter  $-\beta$ . The reader may verify that  $\mathcal{H}_{34}^2(\omega)$  is the same as  $\mathcal{H}_{12}^2(\omega)$ .

We have introduced and illustrated the ideas of phase spectrum and amplitude spectrum using the input-output model. Naturally, these quantities can be computed, in much the same manner as we compute the correlation for a bivariate normal distribution, without recourse to this model. The same is true of squared coherency and error spectrum, which have immediate generalizations to higher dimensions.

The effect of the application of a matrix filter to a vector time series is summarized in Theorem 4.4.1.

**Theorem 4.4.1.** Let  $X_t$  be a real  $k$ -dimensional stationary time series with absolutely summable covariance matrix and let  $\{\mathbf{A}_j\}_{j=-\infty}^{\infty}$  be a sequence of real  $k \times k$  matrices that is absolutely summable. Then the spectral density of

$$\mathbf{Y}_t = \sum_{j=-\infty}^{\infty} \mathbf{A}_j' \mathbf{X}_{t-j}$$

is

$$\mathbf{f}_Y(\omega) = (2\pi)^2 \mathbf{f}_A^*(\omega) \mathbf{f}_X(\omega) \mathbf{f}_A(\omega),$$

where

$$\mathbf{f}_X(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma_X(h) e^{-\epsilon\omega h},$$

$$\mathbf{f}_A(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mathbf{A}_j e^{-\epsilon\omega j},$$

and  $\mathbf{f}_A^*(\omega)$  is the conjugate transpose of  $\mathbf{f}_A(\omega)$ ,

$$\mathbf{f}_A^*(\omega) = \frac{1}{2\pi} \sum_{j=-\infty}^{\infty} \mathbf{A}'_j e^{i\omega j}.$$

**Proof.** The proof parallels that of Theorem 4.3.1. We have

$$\begin{aligned} \mathbf{f}_Y(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \mathbf{A}'_j \Gamma_X(h-s+j) \mathbf{A}_s e^{-i\omega h} \\ &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} \sum_{s=-\infty}^{\infty} \mathbf{A}'_j e^{i\omega j} \Gamma_X(h-s+j) e^{-i\omega(h-s+j)} \mathbf{A}_s e^{-i\omega s} \end{aligned}$$

and the result follows.  $\blacktriangle$

As in the scalar case, the spectral matrix of autoregressive and moving average time series follows immediately from Theorem 4.4.1.

**Corollary 4.4.1.1.** The spectral density of the moving average process

$$\mathbf{X}_t = \sum_{j=-\infty}^{\infty} \mathbf{B}_j \mathbf{e}_{t-j},$$

where  $\{\mathbf{e}_t\}$  is a sequence of uncorrelated  $(\mathbf{0}, \Sigma)$  vector random variables and the sequence  $\{\mathbf{B}_j\}$  is absolutely summable, is

$$\begin{aligned} \mathbf{f}_X(\omega) &= \frac{1}{2\pi} \left( \sum_{j=-\infty}^{\infty} \mathbf{B}_j e^{i\omega j} \right) \Sigma \left( \sum_{j=-\infty}^{\infty} \mathbf{B}'_j e^{-i\omega j} \right) \\ &= (2\pi)^2 \mathbf{f}_B^*(\omega) \Sigma \mathbf{f}_B(\omega). \end{aligned}$$

**Corollary 4.4.1.2.** Define the vector autoregressive process  $\mathbf{X}_t$  by

$$\sum_{j=0}^p \mathbf{A}_j \mathbf{X}_{t-j} = \mathbf{e}_t,$$

where  $\mathbf{A}_0 = \mathbf{I}$ ,  $\mathbf{A}_p \neq \mathbf{0}$ , the  $\mathbf{e}_t$  are uncorrelated  $(\mathbf{0}, \Sigma)$  vector random variables, and the roots of

$$\left| \sum_{j=0}^p \mathbf{A}_j m^{p-j} \right| = 0$$

are less than one in absolute value. Then the spectral density of  $\mathbf{X}_t$  is

$$\mathbf{f}_X(\omega) = 2\pi [\mathbf{f}_A^*(\omega)]^{-1} \Sigma [\mathbf{f}_A(\omega)]^{-1},$$

where  $\mathbf{f}_A(\omega)$  is defined in Theorem 4.4.1.

#### 4.5. MEASUREMENT ERROR – SIGNAL DETECTION

In any statistical model the manner in which the “errors” enter the model is very important. In models of the “input–output” or “independent-variable–dependent-variable” form, measurement error in the output or dependent variable is relatively easy to handle. On the other hand, measurement error in the input or independent variable typically introduces additional complexity into the analysis. In the simple regression model with normal errors the presence of normal measurement error in the independent variable requires additional information, such as knowledge of the variance of the measurement error, before one can estimate the slope of the regression line.

In time series analysis the distinction between independent variable and dependent variable may be blurred. For example, in predicting a future observation in the realization, the past observations play the role of independent variables, while the future observation plays the role of (unknown) dependent variable. As a result, considerable care is required in specifying and treating measurement errors in such analyses. One important problem where errors of observation play a central role is the estimation of the values of an underlying time series that is observed with measurement error.

To introduce the problem, let  $\{X_t; t \in (0, \pm 1, \pm 2, \dots)\}$  be a stationary time series with zero mean. Because of measurement error we do not observe  $X_t$  directly. Instead, we observe

$$Y_t = X_t + u_t, \quad (4.5.1)$$

where  $\{u_t; t \in (0, \pm 1, \pm 2, \dots)\}$  is a time series with  $u_t$  independent of  $X_j$  for all  $t, j$ . The  $u_t$  are the measurement errors or the noise in the system. A problem of *signal measurement* or signal detection is the construction of an estimator of  $X_t$ , given a realization on  $Y_t$  (or a portion of the realization). We assume the covariance functions of  $X_t$  and  $u_t$  are known.

We first consider the problem in the time domain and restrict ourselves to finding a linear filter that minimizes the mean square error of our estimator of  $X_t$ . Thus we desire weights  $\{a_j; j = -L, -(L-1), \dots, M-1, M\}$ , for  $L \geq 0$  and  $M \geq 0$  fixed, such that

$$E\left\{\left(X_t - \sum_{j=-L}^M a_j Y_{t-j}\right)^2\right\} = E\left\{\left[X_t - \sum_{j=-L}^M a_j (X_{t-j} + u_{t-j})\right]^2\right\} \quad (4.5.2)$$

is a minimum. We set the derivatives of (4.5.2) with respect to the  $a_j$  equal to zero and obtain the system of equations

$$\sum_{j=-L}^M a_j [\gamma_{XX}(j-r) + \gamma_{uu}(j-r)] = \gamma_{XX}(r), \\ r = -L, -(L-1), \dots, M-1, M. \quad (4.5.3)$$

For modest choices of  $L$  and  $M$  this system of liner equations is easily solved for the  $a_j$ . The mean square error of the estimator can then be obtained from (4.5.2).

To obtain the set of weights one would use if one had available the entire realization, we investigate the problem in the frequency domain. This permits us to establish the bound on the mean square error and to use this bound to evaluate the performance of a filter with a finite number of weights. We assume that  $u$ , and  $X$ , have continuous spectral densities with bounded derivatives. The spectral density of  $Y$ , and the cross spectral density follow immediately from the independence:

$$\begin{aligned} f_{YY}(\omega) &= f_{XX}(\omega) + f_{uu}(\omega), \\ f_{XY}(\omega) &= f_{XX}(\omega). \end{aligned}$$

We assume that  $f_{YY}(\omega)$  is strictly positive and that  $f_{YY}^{-1}(\omega)$  has bounded first derivative.

We shall search in the class of absolutely summable filters  $\{a_j\}$  for that filter such that our estimator of  $X_t$ ,

$$\hat{X}_t = \sum_{j=-\infty}^{\infty} a_j Y_{t-j},$$

will have minimum mean square error. We write

$$\begin{aligned} f_{\hat{X}\hat{X}}(\omega) &= (2\pi)^2 f_a(\omega) f_a^*(\omega) f_{YY}(\omega), \\ f_{X\hat{X}}(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \gamma_{X\hat{X}}(h) e^{-i h \omega} \\ &= 2\pi f_a(\omega) f_{XY}(\omega). \end{aligned}$$

Now the mean square error is

$$E\{(X_t - \hat{X}_t)^2\} = E\left\{\left(X_t - \sum_{j=-\infty}^{\infty} a_j Y_{t-j}\right)^2\right\}, \quad (4.5.4)$$

and  $W_t = X_t - \sum_{j=-\infty}^{\infty} a_j Y_{t-j}$  is a stationary time series with a spectral density, say  $f_{WW}(\omega)$ . Hence, the variance of  $W_t$  is

$$\begin{aligned} \gamma_{WW}(0) &= E\{W_t^2\} = \int_{-\pi}^{\pi} f_{WW}(\omega) d\omega \\ &= \int_{-\pi}^{\pi} \{f_{XX}(\omega) - 2\pi[f_a(\omega) f_{XY}(\omega) + f_a^*(\omega) f_{YX}(\omega)] \\ &\quad + (2\pi)^2 f_a(\omega) f_a^*(\omega) f_{YY}(\omega)\} d\omega. \quad (4.5.5) \end{aligned}$$

We have converted the problem of finding the  $a_j$  that minimize (4.5.4) to the problem of finding the  $f_a(\omega)$  that minimizes (4.5.5). This problem retains the same form and appearance as the classical regression problem with  $f_a(\omega)$  playing the role

of the vector of coefficients. Therefore, we take as our candidate solution

$$f_a(\omega) = \frac{1}{2\pi} \frac{f_{YX}(\omega)}{f_{YY}(\omega)}, \quad (4.5.6)$$

which gives

$$\begin{aligned} \gamma_{WW}(0) &= \int_{-\pi}^{\pi} \left[ f_{XX}(\omega) - \frac{f_{XY}(\omega)f_{YX}^*(\omega)}{f_{YY}(\omega)} \right] d\omega \\ &= \int_{-\pi}^{\pi} f_{XX}(\omega)[1 - \mathcal{H}_{XY}^2(\omega)] d\omega. \end{aligned} \quad (4.5.7)$$

The weights  $a_j$  are given by the inverse transform of  $f_a(\omega)$ ,

$$a_j = \int_{-\pi}^{\pi} f_a(\omega) e^{i\omega j} d\omega, \quad j = 0, \pm 1, \pm 2, \dots$$

To demonstrate that these weights yield the minimum value for the mean square error, we consider an alternative to (4.5.6):

$$f_b(\omega) = \frac{f_{YX}(\omega)}{2\pi f_{YY}(\omega)} + \frac{f_d(\omega)}{2\pi f_{YY}(\omega)},$$

where  $f_b(\omega)$  must be in the class of functions such that the integral defining the mean square error exists. The mean square error is then

$$\begin{aligned} &\int_{-\pi}^{\pi} \{f_{XX}(\omega) - 2\pi[f_b(\omega)f_{YX}^*(\omega) + f_b^*(\omega)f_{YX}(\omega)] \\ &\quad + (2\pi)^2 f_b(\omega)f_b^*(\omega)f_{YY}(\omega)\} d\omega \\ &= \int_{-\pi}^{\pi} \left[ f_{XX}(\omega) - \frac{|f_{YX}(\omega)|^2}{f_{YY}(\omega)} + \frac{|f_d(\omega)|^2}{f_{YY}(\omega)} \right] d\omega. \end{aligned} \quad (4.5.8)$$

Since  $|f_d(\omega)|^2/f_{YY}(\omega)$  is nonnegative, we conclude that the  $f_a(\omega)$  of (4.5.6) yields the minimum value for (4.5.5).

Note that  $2\pi f_a(\omega) = f_{YY}^{-1}(\omega)f_{YX}(\omega)$  is a real valued symmetric function of  $\omega$  because  $f_{YX}(\omega) = f_{XX}(\omega)$  is a real valued symmetric function of  $\omega$ . Therefore, the weights  $a_j$  are also symmetric about zero.

We summarize in Theorem 4.5.1.

**Theorem 4.5.1.** Let  $\{X_t: t \in (0, \pm 1, \pm 2, \dots)\}$  and  $\{u_t: t \in (0, \pm 1, \pm 2, \dots)\}$  be independent zero mean stationary time series, and define  $Y_t = X_t + u_t$ . Let  $f_{YY}(\omega)$ ,  $f_{YY}^{-1}(\omega)$ , and  $f_{XX}(\omega)$  be continuous with bounded first derivatives. Then the best linear filter for extracting  $X_t$  from a realization of  $Y_t$  is given by

$$a_j = \int_{-\pi}^{\pi} f_a(\omega) e^{i\omega j} d\omega,$$

where

$$f_a(\omega) = \frac{f_{YX}(\omega)}{2\pi f_{YY}(\omega)}.$$

Furthermore, the mean square error of  $\sum_{j=-\infty}^{\infty} a_j Y_{t-j}$  as an estimator for  $X_t$  is

$$\int_{-\pi}^{\pi} \{f_{XX}(\omega) - [f_{YY}(\omega)]^{-1}|f_{XY}(\omega)|^2\} d\omega = \int_{-\pi}^{\pi} f_{XX}(\omega)[1 - K_{XY}^2(\omega)] d\omega.$$

**Proof.** Since the derivatives of  $f_{YY}^{-1}(\omega)$  and  $f_{YX}(\omega)$  are bounded, the derivative of  $f_{YY}^{-1}(\omega)f_{YX}(\omega)$  is square integrable. Therefore, by Theorem 3.1.8, the Fourier coefficients of  $f_{YY}^{-1}(\omega)f_{YX}(\omega)$ , the  $a_j$ , are absolutely summable, and by Theorem 2.2.1,  $\sum_{j=-\infty}^{\infty} a_j Y_{t-j}$  is a well-defined random variable. That  $\{a_j\}$  is the best linear filter follows from (4.5.8), and the mean square error of the filter follows from (4.5.7). ▲

**Example 4.5.1.** To illustrate the ideas of this section, we use some data on the sediment suspended in the water of the Des Moines River at Boone, Iowa. A portion of the data obtained by daily sampling of the water during 1973 are displayed in Table 4.5.1. The data are the logarithm of the parts per million of suspended sediment. Since the laboratory determinations are made on a small sample of water collected from the river, the readings can be represented as

$$Y_t = X_t + u_t,$$

where  $Y_t$  is the recorded value,  $X_t$  is the "true" average sediment in the river water, and  $u_t$  is the measurement error introduced by sampling and laboratory determination. Assume that  $X_t$  can be represented as a first order autoregressive process

$$X_t - 5.28 = 0.81(X_{t-1} - 5.28) + e_t,$$

where the  $e_t$  are independent  $(0, 0.172)$  random variables. Assume further that  $u_t$  is a sequence of independent  $(0, 0.053)$  random variables independent of  $X_t$  for all  $t, j$ .

To construct a filter  $\{a_{-2}, a_{-1}, a_0, a_1, a_2\}$  that will best estimate  $X_t$  using the observations  $\{Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2}\}$ , we solve the system of equations

$$\begin{pmatrix} 0.553 & 0.405 & 0.328 & 0.266 & 0.215 \\ 0.405 & 0.553 & 0.405 & 0.328 & 0.266 \\ 0.328 & 0.405 & 0.553 & 0.405 & 0.328 \\ 0.266 & 0.328 & 0.405 & 0.553 & 0.405 \\ 0.215 & 0.266 & 0.328 & 0.405 & 0.553 \end{pmatrix} \begin{pmatrix} a_{-2} \\ a_{-1} \\ a_0 \\ a_1 \\ a_2 \end{pmatrix} = \begin{pmatrix} 0.328 \\ 0.405 \\ 0.500 \\ 0.405 \\ 0.328 \end{pmatrix}$$

to obtain

$$\mathbf{a} = \{a_{-2}, a_{-1}, a_0, a_1, a_2\} = \{0.023, 0.120, 0.702, 0.120, 0.023\}.$$

**Table 4.5.1. Logarithm of Sediment Suspended in Des Moines River, Boone, Iowa, 1973**

Daily Observations	Smoothed Observations	
	Two-Sided Filter	One-Sided Filter
5.44	—	—
5.38	—	—
5.43	5.40	—
5.22	5.26	5.26
5.28	5.27	5.28
5.21	5.22	5.22
5.23	5.25	5.23
5.33	5.37	5.31
5.58	5.63	5.53
6.18	6.08	6.04
6.16	6.14	6.10
6.07	6.13	6.04
6.56	6.38	6.42
5.93	5.96	5.99
5.70	5.69	5.74
5.36	5.39	5.42
5.17	5.24	5.22
5.35	5.36	5.33
5.51	5.51	5.47
5.80	5.67	5.72
5.29	5.36	5.37
5.28	5.29	5.30
5.27	—	5.27
5.17	—	5.19

SOURCE: U.S. Department of Interior Geological Survey—Water Resources Division, Sediment Concentration Notes, Des Moines River, Boone, Iowa.

The mean square error of the filtered time series  $5.28 + \sum_{j=-2}^2 a_j(Y_{t-j} - 5.28)$  as an estimator of  $X_t$  is  $0.500 - \mathbf{a}[0.328, 0.405, 0.500, 0.405, 0.328]' = 0.0372$ . The data transformed by this filter are displayed in the second column of Table 4.5.1. Note that the filtered data are “smoother” in that the variance of changes from one period to the next is smaller for the filtered data than for the original data.

We now obtain a one-sided filter that can be used to estimate the most recent value of  $X_t$  using only the most recent and the four preceding values of  $Y_t$ . The estimator of  $X_t$  is given by

$$5.28 + b_0(Y_t - 5.28) + b_1(Y_{t-1} - 5.28) \\ + b_2(Y_{t-2} - 5.28) + b_3(Y_{t-3} - 5.28) + b_4(Y_{t-4} - 5.28),$$

where

$$\begin{pmatrix} b_0 \\ b_1 \\ b_2 \\ b_3 \\ b_4 \end{pmatrix} = \begin{pmatrix} 0.553 & 0.405 & 0.328 & 0.266 & 0.215 \\ 0.405 & 0.553 & 0.405 & 0.328 & 0.266 \\ 0.328 & 0.405 & 0.553 & 0.405 & 0.328 \\ 0.266 & 0.328 & 0.405 & 0.553 & 0.405 \\ 0.215 & 0.266 & 0.328 & 0.405 & 0.553 \end{pmatrix}^{-1} \begin{pmatrix} 0.500 \\ 0.405 \\ 0.328 \\ 0.266 \\ 0.215 \end{pmatrix} = \begin{pmatrix} 0.790 \\ 0.134 \\ 0.023 \\ 0.005 \\ 0.000 \end{pmatrix}.$$

This one-sided estimator of  $X_t$  has a mean square error of 0.0419. The filtered data using this filter are given in the last column of Table 4.5.1.

To obtain the minimum value for the mean square error of the two-sided filter, we evaluated (4.5.7), where

$$\begin{aligned} f_{XX}(\omega) &= \frac{0.172}{2\pi(1 - 0.81e^{-i\omega})(1 - 0.81e^{i\omega})}, \\ f_{YY}(\omega) &= \frac{0.172}{2\pi(1 - 0.81e^{-i\omega})(1 - 0.81e^{i\omega})} + \frac{0.053}{2\pi} \\ &= \frac{0.2524(1 - 0.170e^{i\omega})(1 - 0.170e^{-i\omega})}{2\pi(1 - 0.81e^{-i\omega})(1 - 0.81e^{i\omega})}, \end{aligned}$$

obtaining

$$\begin{aligned} \gamma_{WW}(0) &= \int_{-\pi}^{\pi} f_{XX}(\omega) d\omega - \int_{-\pi}^{\pi} \frac{(0.172)^2 d\omega}{2\pi(0.2524)|1 - 0.170e^{-i\omega}|^2|1 - 0.81e^{-i\omega}|^2} \\ &= 0.500 - 0.463 = 0.037. \end{aligned}$$

The infinite sequence of weights is given by the inverse transform of

$$\frac{0.172}{2\pi(0.2524)(1 - 0.17e^{-i\omega})(1 - 0.17e^{i\omega})}.$$

which yields

$$(\dots, 0.1193, 0.7017, 0.1193, 0.0203, 0.0034, \dots).$$

While it is possible to use spectral methods to evaluate the minimum mean square error of a one-sided filter [see, for example, Yaglom (1962, p. 97)], we examine the problem in a slightly different manner. Since  $Y_t$  is an autoregressive moving average (1, 1) process, the methods of Section 2.9 can be used to obtain a one-period-ahead predictor  $\hat{Y}_t(Y_{t-1}, Y_{t-2}, \dots)$  based on an infinite past. As  $Y_t = X_t + u_t$ , where  $u_t$  is a sequence of independent random variables, the best predictor of  $X_t$  based on  $Y_{t-1}, Y_{t-2}, \dots$  must be the same as the best predictor of  $Y_t$  based on  $Y_{t-1}, Y_{t-2}, \dots$ . Furthermore, given the predictor, the partial correlation between any  $Y_{t-j}$ ,  $j > 0$ , and  $X_t$  is zero. Therefore, to obtain the best filter for  $X_t$  using  $Y_t, Y_{t-1}, \dots$ , we find the optimal linear combination of  $\hat{Y}_t(Y_{t-1}, Y_{t-2}, \dots)$  and  $Y_t$ .

Denote the linear combination by  $c_0 Y_t + c_1 \hat{Y}_t(Y_{t-1}, Y_{t-2}, \dots)$ , where

$$\begin{pmatrix} c_0 \\ c_1 \end{pmatrix} = \begin{pmatrix} 0.5530 & 0.3006 \\ 0.3006 & 0.3006 \end{pmatrix}^{-1} \begin{pmatrix} 0.5000 \\ 0.3006 \end{pmatrix} = \begin{pmatrix} 0.7900 \\ 0.2100 \end{pmatrix},$$

the matrix

$$\begin{pmatrix} 0.5530 & 0.3006 \\ 0.3006 & 0.3006 \end{pmatrix}$$

is the covariance matrix of  $[Y_t, \hat{Y}_t(Y_{t-1}, Y_{t-2}, \dots)]$ , and  $(0.5000, 0.3006)$  is the vector of covariances between  $[Y_t, \hat{Y}_t(Y_{t-1}, Y_{t-2}, \dots)]$  and  $X_t$ . It follows that the minimum mean square error for a one-sided predictor is 0.0419.  $\blacktriangle\blacktriangle$

## 4.6. STATE SPACE MODELS AND KALMAN FILTERING

We begin our discussion under the assumption that the univariate time series  $X_t$  is a stationary, zero mean, first order autoregressive process. We write

$$X_t = \alpha X_{t-1} + e_t, \quad t = 1, 2, \dots, \quad (4.6.1)$$

where  $e_t$  are independent identically distributed  $(0, \sigma_e^2)$  random variables, denoted by  $e_t \sim \text{II}(0, \sigma_e^2)$ . Assume we are unable to observe  $X_t$  directly. Instead, we observe  $Y_t$ , where

$$Y_t = X_t + u_t, \quad t = 1, 2, \dots, \quad (4.6.2)$$

and  $u_t$  is the measurement error. We assume  $u_t \sim \text{II}(0, \sigma_u^2)$  and that  $u_t$  is independent of  $e_j$  for all  $t$  and  $j$ .

The model (4.6.1)–(4.6.2) is a special case of the *state space* representation of time series. Equation (4.6.1) is called the *state equation* or the *transition equation*, and  $X_t$  is called the *state* of the system at time  $t$ . Equation (4.6.2) is called the *measurement equation* or the *observation equation*. The model (4.6.1)–(4.6.2) was introduced in the Des Moines River example of Section 4.5. In that example it is very natural to think of the unknown level of the river as the true “state” of nature.

In Section 4.5, we constructed linear filters to estimate the values of the time series  $X_t$  that is observed subject to measurement error. The filters were designed to minimize the mean square error of the estimation error for a particular set of observations. In many applications, the observation set is composed of all previous observations plus the current observation. In some engineering applications, it is important to have an efficient method of computing the current estimated value requiring as little storage of information as possible. Such computational methods have been developed by Kalman (1960, 1963) and others. The methods are often

called *Kalman filters*. In this section, we study the state space model and the Kalman filter.

Let the model (4.6.1) and (4.6.2) hold, and assume that an initial estimator for  $X_0$ , denoted by  $\hat{X}_0$ , is available. Let

$$\hat{X}_0 = X_0 + v_0, \quad (4.6.3)$$

where  $v_0$  is a  $(0, \sigma_{v_0}^2)$  random variable, independent of  $(u_t, e_t)$ ,  $t = 1, 2, \dots$ . The initial estimator  $\hat{X}_0$  and parameters  $\sigma_u^2$ ,  $\sigma_e^2$ ,  $\sigma_{v_0}^2$ , and  $\alpha$  are assumed known. Equation (4.6.3) is called the *initial condition equation* (or *starting equation*). A possible choice for  $\hat{X}_0$  for the model (4.6.1) is  $\hat{X}_0 = 0$ . With  $\hat{X}_0 = 0$ , we have  $\sigma_{v_0}^2 = \gamma_X(0) = (1 - \alpha^2)^{-1} \sigma_e^2$ .

At time  $t = 1$ , we have the observation  $(Y_1, \hat{X}_0)$  and knowledge of  $\sigma_u^2$ ,  $\sigma_e^2$ , and  $\alpha$  to use in constructing an estimator (predictor) of  $X_1$ . On the basis of the specification, we have

$$Y_1 = X_1 + u_1, \quad (4.6.4)$$

$$\alpha\hat{X}_0 = X_1 + w_1,$$

where  $w_1 = \alpha v_0 - e_1$ . The system of equations (4.6.4) can be written in matrix form as

$$\mathbf{Z}_1 = \mathbf{J}X_1 + \boldsymbol{\epsilon}_1,$$

where  $\mathbf{Z}_1 = (Y_1, \alpha\hat{X}_0)'$ ,  $\mathbf{J} = (1, 1)'$ , and  $\boldsymbol{\epsilon}_1 = (u_1, w_1)'$ . Because  $u_1$ ,  $e_1$ , and  $v_0$  are mutually uncorrelated, the covariance matrix of  $\boldsymbol{\epsilon}_1$  is  $\text{diag}(\sigma_u^2, \sigma_{w_1}^2)$ , where  $\sigma_{w_1}^2 = \sigma_e^2 + \alpha^2 \sigma_{v_0}^2$ . Therefore, it is natural to construct an estimator of  $X_1$  as the weighted average of  $Y_1$  and  $\alpha\hat{X}_0$ , where the weights are proportional to the inverses of the variances of  $u_1$  and  $w_1$ . Thus,

$$\hat{X}_1 = (\sigma_u^{-2} + \sigma_{w_1}^{-2})^{-1} (\sigma_u^{-2} Y_1 + \sigma_{w_1}^{-2} \alpha\hat{X}_0). \quad (4.6.5)$$

The estimator (4.6.5) is constructed by analogy to linear regression theory. In the problem formulation (4.6.4), the information about the unknown random value,  $X_1$ , is contained in the second equation of system (4.6.4). See Exercise 4.19. The same approach can be used to construct succeeding estimators.

Let the error in  $\hat{X}_1$  as an estimator of  $X_1$  be  $v_1$ , where

$$v_1 = \hat{X}_1 - X_1 = (\sigma_u^{-2} + \sigma_{w_1}^{-2})^{-1} (\sigma_u^{-2} u_1 + \sigma_{w_1}^{-2} w_1). \quad (4.6.6)$$

The variance of  $v_1$  is

$$\sigma_{v_1}^2 = (\sigma_u^{-2} + \sigma_{w_1}^{-2})^{-1}. \quad (4.6.7)$$

At time  $t = 2$ , it is desired to estimate  $X_2$  using  $Y_2$ ,  $Y_1$ , and  $\hat{X}_0$ . Now  $\alpha\hat{X}_1$  is the best

predictor of  $X_2$  constructed from the data  $(Y_1, \hat{X}_0)$ . Therefore, we need only combine the information in  $\alpha\hat{X}_1$  with that in  $Y_2$  to obtain the best estimator of  $X_2$ . Using (4.6.1)–(4.6.2) and the identity

$$\alpha\hat{X}_1 = \alpha X_1 + \alpha(\hat{X}_1 - X_1) = \alpha X_1 + \alpha v_1,$$

we have the system of two equations containing  $X_2$ ,

$$\begin{aligned} Y_2 &= X_2 + u_2, \\ \alpha\hat{X}_1 &= X_2 + w_2, \end{aligned} \tag{4.6.8}$$

where  $w_2 = \alpha v_1 - e_2$ . Because the vectors  $(u_t, e_t)$  are uncorrelated, we have

$$E\{(u_2, e_2, v_1)'(u_2, e_2, v_1)\} = \text{diag}(\sigma_u^2, \sigma_e^2, \sigma_{v1}^2)$$

and

$$\mathbf{V}\{(u_2, w_2)\} = \text{diag}(\sigma_u^2, \sigma_e^2 + \alpha^2 \sigma_{v1}^2).$$

It follows that the best estimator of  $X_2$ , given  $(\hat{X}_0, Y_1, Y_2)$ , is

$$\hat{X}_2 = (\sigma_u^{-2} + \sigma_{w2}^{-2})^{-1}(\sigma_u^{-2}Y_2 + \sigma_{w2}^{-2}\alpha\hat{X}_1), \tag{4.6.9}$$

where  $\sigma_{w2}^2 = \sigma_e^2 + \alpha^2 \sigma_{v1}^2$ . Letting the error  $\hat{X}_2 - X_2$  be

$$v_2 = (\sigma_u^{-2} + \sigma_{w2}^{-2})^{-1}(\sigma_u^{-2}u_2 + \sigma_{w2}^{-2}w_2),$$

the variance of  $v_2$  is  $\sigma_{v2}^2 = (\sigma_u^{-2} + \sigma_{w2}^{-2})^{-1}$ .

Because  $\hat{X}_{t-1}$  contains all of the information about  $X_t$  available from the previous observations, the estimator of  $X_t$  for general  $t$  is

$$\hat{X}_t = (\sigma_u^{-2} + \sigma_{wt}^{-2})^{-1}(\sigma_u^{-2}Y_t + \sigma_{wt}^{-2}\alpha\hat{X}_{t-1}), \tag{4.6.10}$$

where  $w_t = \alpha v_{t-1} - e_t$ ,  $v_t = \hat{X}_t - X_t$ ,

$$\sigma_{wt}^2 = \sigma_e^2 + \alpha^2 \sigma_{v,t-1}^2,$$

and

$$\sigma_{vt}^2 = (\sigma_u^{-2} + \sigma_{wt}^{-2})^{-1}.$$

Equation (4.6.10) can be rearranged to yield

$$\hat{X}_t = \alpha\hat{X}_{t-1} + (\sigma_u^{-2} + \sigma_{wt}^{-2})^{-1}\sigma_u^{-2}(Y_t - \alpha\hat{X}_{t-1}) \tag{4.6.11}$$

or

$$\hat{X}_t = \alpha \hat{X}_{t-1} + \sigma_{w_t}^2 (\sigma_u^2 + \sigma_{w_t}^2)^{-1} (Y_t - \alpha \hat{X}_{t-1}). \quad (4.6.12)$$

The first term on the right of (4.6.12) is the estimator of  $X_t$  based upon  $\hat{X}_{t-1}$ . The second term is the negative of an estimator of the error made in predicting  $X_t$  with  $\alpha \hat{X}_{t-1}$ . We can give a direct justification of (4.6.12) as follows. Because  $u_t - w_t = Y_t - \alpha \hat{X}_{t-1}$  is a linear combination of  $Y_t$  and  $\hat{X}_{t-1}$ , the linear estimator of  $X_t$  based on

$$(u_t - w_t, \hat{X}_{t-1}) = (Y_t - \alpha \hat{X}_{t-1}, \hat{X}_{t-1})$$

is the same as the linear estimator based on  $(Y_t, \hat{X}_{t-1})$ . Now  $u_t - w_t$  is uncorrelated with  $\hat{X}_{t-1}$ . It follows, by the properties of regression (or of Hilbert spaces), that the best linear estimator of  $X_t$  based on  $(\hat{X}_{t-1}, Y_t)$  is the best linear estimator of  $X_t$  based on  $\hat{X}_{t-1}$ , plus the best linear estimator of  $X_t - \alpha \hat{X}_{t-1}$  based on  $Y_t - \alpha \hat{X}_{t-1}$ . Hence, the best estimator of  $X_t$  is given by (4.6.12). See Exercise 4.25.

An equation equivalent to (4.6.12) is

$$\hat{X}_t = Y_t - \sigma_u^2 (\sigma_u^2 + \sigma_{w_t}^2)^{-1} (Y_t - \alpha \hat{X}_{t-1}). \quad (4.6.13)$$

The form (4.6.12) appears in the original signal extraction literature, and the form (4.6.13) appears in the random model prediction literature. In the form (4.6.13), an estimator of the error  $u_t$  is subtracted from  $Y_t$  to obtain the estimator of  $X_t$ .

The variance of  $v_t$  satisfies the equation

$$\sigma_{v_t}^2 = \sigma_{w_t}^2 - (\sigma_{w_t}^2)^2 (\sigma_u^2 + \sigma_{w_t}^2)^{-1}. \quad (4.6.14)$$

Hence, the data required at time  $t+1$  to construct the estimator of  $X_{t+1}$  and to construct the variance of the estimation error are the elements of  $(Y_{t+1}, \hat{X}_t, \sigma_{v_t}^2)$ . The equations (4.6.12) and (4.6.14) are sometimes called the *updating equations* of the Kalman filter.

The results for the simple model (4.6.1)–(4.6.2) generalize to  $p$ -dimensional vector time series  $\mathbf{X}_t$ , and to the situation wherein the observed vector  $\mathbf{Y}_t$  is the sum of a known linear function of  $\mathbf{X}_t$  and measurement error. Let

$$\mathbf{X}_t = \mathbf{A}_t \mathbf{X}_{t-1} + \mathbf{e}_t, \quad t = 1, 2, \dots, \quad (4.6.15)$$

$$\mathbf{Y}_t = \mathbf{H}_t \mathbf{X}_t + \mathbf{u}_t, \quad t = 1, 2, \dots, \quad (4.6.16)$$

$$\hat{\mathbf{X}}_0 = \mathbf{X}_0 + \mathbf{v}_0, \quad (4.6.17)$$

where  $\mathbf{X}_t$  is a  $p$ -dimensional column vector,  $\mathbf{Y}_t$  is an  $r$ -dimensional column vector,  $\{\mathbf{H}_t\}$  is a sequence of known  $r \times p$  matrices,  $\mathbf{A}_t$  is a sequence of known  $p \times p$  matrices,  $\hat{\mathbf{X}}_0$  is a known initial vector, and  $\{(\mathbf{u}_t', \mathbf{e}_t')\}$  is a sequence of uncorrelated,

zero mean, vector random variables with known covariance matrix

$$E\{(\mathbf{u}_t', \mathbf{e}_t')'(\mathbf{u}_t', \mathbf{e}_t')\} = \text{block diag}(\Sigma_{uut}, \Sigma_{eet}) .$$

Equation (4.6.15) is the state equation, equation (4.6.16) is the measurement equation, and equation (4.6.17) is the initial equation. As before, we assume  $\mathbf{v}_0$  of the initial equation to be uncorrelated with  $\mathbf{u}_t$  and  $\mathbf{e}_t$ ,  $t = 1, 2, \dots$ . Considerable generalization of the model is obtained by permitting the variances to be functions of  $t$  and by the inclusion of the matrices  $\mathbf{H}_t$  in the measurement equation. Many different forms of state space representations appear in the literature.

For the vector model, the system of equations analogous to (4.6.8) is

$$\begin{aligned} \mathbf{Y}_t &= \mathbf{H}_t \mathbf{X}_t + \mathbf{u}_t, \\ (4.6.18) \end{aligned}$$

$$\mathbf{A}_t \hat{\mathbf{X}}_{t-1} = \mathbf{X}_t + \mathbf{w}_t$$

for  $t = 1, 2, \dots$ , where  $\mathbf{w}_t = \mathbf{A}_t \mathbf{v}_{t-1} - \mathbf{e}_t$  and  $\mathbf{v}_t = \hat{\mathbf{X}}_t - \mathbf{X}_t$ . If we assume that  $\Sigma_{uut}$  is nonsingular, the best estimator of  $\mathbf{X}_t$  in (4.6.18) is

$$\hat{\mathbf{X}}_t = (\mathbf{H}'_t \Sigma_{uut}^{-1} \mathbf{H}_t + \Sigma_{wwt}^{-1})^{-1} (\mathbf{H}'_t \Sigma_{uut}^{-1} \mathbf{Y}_t + \Sigma_{wwt}^{-1} \mathbf{A}_t \hat{\mathbf{X}}_{t-1}), \quad (4.6.19)$$

where

$$\Sigma_{wwt} = E\{\mathbf{w}_t \mathbf{w}'_t\} = \Sigma_{eet} + \mathbf{A}_t \Sigma_{vv,t-1,t-1} \mathbf{A}'_t, \quad (4.6.20)$$

$$\begin{aligned} \Sigma_{vvtt} &= E\{(\hat{\mathbf{X}}_t - \mathbf{X}_t)(\hat{\mathbf{X}}_t - \mathbf{X}_t)'\} = (\mathbf{H}'_t \Sigma_{uut}^{-1} \mathbf{H}_t + \Sigma_{wwt}^{-1})^{-1} \\ &= \Sigma_{wwt} - \Sigma_{wwt} \mathbf{H}'_t \mathbf{D}_t^{-1} \mathbf{H}_t \Sigma_{wwt}, \end{aligned} \quad (4.6.21)$$

$$\mathbf{D}_t = \Sigma_{uut} + \mathbf{H}_t \Sigma_{wwt} \mathbf{H}'_t. \quad (4.6.22)$$

The estimator (4.6.19) can also be written as

$$\hat{\mathbf{X}}_t = \mathbf{A}_t \hat{\mathbf{X}}_{t-1} + \Sigma_{wwt} \mathbf{H}'_t \mathbf{D}_t^{-1} (\mathbf{Y}_t - \mathbf{H}_t \mathbf{A}_t \hat{\mathbf{X}}_{t-1}), \quad (4.6.23)$$

where  $\mathbf{D}_t$  is the covariance matrix of  $\mathbf{Y}_t - \mathbf{H}_t \mathbf{A}_t \hat{\mathbf{X}}_{t-1}$  and  $\Sigma_{wwt} \mathbf{H}'_t$  is the covariance between  $\mathbf{w}_t$  and  $\mathbf{Y}_t - \mathbf{H}_t \mathbf{A}_t \hat{\mathbf{X}}_{t-1}$ . Therefore, the estimator (4.6.23) is the difference between the unbiased estimator  $\mathbf{A}_t \hat{\mathbf{X}}_{t-1}$  of  $\mathbf{X}_t$  and an unbiased estimator of the error  $\mathbf{w}_t$  in  $\mathbf{A}_t \hat{\mathbf{X}}_{t-1}$ .

Equation (4.6.20), the second equation of (4.6.21), and equation (4.6.23) form a set of equations that can be used to construct  $\hat{\mathbf{X}}_t$  given  $\mathbf{Y}_t$ ,  $\hat{\mathbf{X}}_{t-1}$ , and  $\Sigma_{vv,t-1,t-1}$ . If this system of equations is used for updating, only the matrix  $\mathbf{D}_t$  is required to be nonsingular. The matrix  $\mathbf{D}_t$  should always be nonsingular because there is little reason for the subject matter specialist to consider singular observation vectors  $\mathbf{Y}_t$ .

The vector of updating equations analogous to (4.6.13) is

$$\hat{\mathbf{X}}_t = \mathbf{Y}_t - \Sigma_{uuu} \mathbf{D}_t^{-1} (\mathbf{Y}_t - \mathbf{H}_t \mathbf{A}_t \hat{\mathbf{X}}_{t-1})$$

and an alternative expression for  $\Sigma_{vvv}$  is

$$\Sigma_{vvv} = \Sigma_{uuu} - \Sigma_{uuu} \mathbf{D}_t^{-1} \Sigma_{uuu}.$$

**Example 4.6.1.** In this example, we construct the Kalman filter for the Des Moines River example of Section 4.5. The model for the data is

$$\begin{aligned} X_t - 5.28 &= 0.81(X_{t-1} - 5.28) + e_t, \\ Y_t &= X_t + u_t, \end{aligned} \tag{4.6.24}$$

where  $e_t \sim \text{II}(0, 0.172)$  and  $u_t \sim \text{II}(0, 0.053)$ . The sediment time series has a nonzero mean, so the difference  $X_t - 5.28$  plays the role of  $X_t$  of (4.6.1).

We begin with the first observation of Table 4.6.1, which we call  $Y_1$ . If  $X_t$  is a stationary process, as we assume for the Des Moines River example, we can use the population mean to initiate the filter. If we use the population mean as an estimator of  $X_0$ , the variance of the estimation error is 0.500, which is the variance of the  $X_t$  process. The system of equations (4.6.4) becomes

$$\begin{aligned} Y_1 - 5.28 &= 0.16 = (X_1 - 5.28) + u_1, \\ 0 &= (X_1 - 5.28) + w_1, \end{aligned}$$

**Table 4.6.1. Estimates of Sediment Constructed with Kalman Filter**

$t$	$Y_t$	$\hat{X}_t$	$\sigma_{v1}^2$	$\sigma_{w1}^2$
1	5.44	5.42467	0.04792	0.50000
2	5.38	5.38355	0.04205	0.20344
3	5.43	5.41613	0.04188	0.19959
4	5.22	5.25574	0.04187	0.19948
5	5.28	5.27588	0.04187	0.19947
6	5.21	5.22399	0.04187	0.19947
7	5.23	5.23097	0.04187	0.19947
8	5.33	5.31117	0.04187	0.19947
9	5.58	5.52232	0.04187	0.19947
10	6.18	6.03227	0.04187	0.19947
11	6.16	6.10318	0.04187	0.19947
12	6.07	6.04413	0.04187	0.19947
13	6.56	6.42123	0.04187	0.19947
14	5.93	5.98760	0.04187	0.19947
15	5.70	5.73215	0.04187	0.19947

where  $\mathbf{V}\{(u_1, w_1)\} = \text{diag}(0.053, 0.500)$ . Therefore, by (4.6.10),

$$\begin{aligned}\hat{X}_1 &= 5.28 + (18.868 + 2.000)^{-1}[18.868(Y_1 - 5.28) + 0] \\ &= 5.28 + 0.145 = 5.425.\end{aligned}$$

The reader may verify that the coefficient for  $Y_1 - 5.28$  is the covariance between  $X_1$  and  $Y_1$  divided by the variance of  $Y_1$ . The variance of the error in the estimator of  $X_1$  is  $\sigma_{v1}^2 = 0.0479$ .

The estimate of  $X_2$  is

$$\begin{aligned}\hat{X}_2 &= 5.28 + [(0.053)^{-1} + (0.2034)^{-1}]^{-1} \\ &\quad \times [(0.053)^{-1}(Y_2 - 5.28) + (0.2034)^{-1}(0.81)(\hat{X}_1 - 5.28)] \\ &= 0.2075 + 0.7933Y_2 + 0.1674\hat{X}_1 \\ &= 5.3836,\end{aligned}$$

where  $\sigma_{w2}^2 = \sigma_e^2 + \alpha^2 \sigma_{v1}^2 = 0.2034$ . The variance of  $\hat{X}_2 - X_2$  is

$$\begin{aligned}\sigma_{v2}^2 &= 0.2034 - (0.2034)^2(0.053 + 0.2034)^{-1} \\ &= 0.0420.\end{aligned}$$

The estimate for  $X_3$  is

$$\begin{aligned}\hat{X}_3 &= 5.28 + \frac{(0.053)^{-1}(Y_3 - 5.28) + (4.0581)(\hat{X}_2 - 5.28)}{(0.053)^{-1} + (0.1996)^{-1}} \\ &= 5.4161,\end{aligned}$$

and the variance of the estimation error is

$$\sigma_{v3}^2 = 0.1996 - (0.1996)^2(0.053 + 0.1996)^{-1} = 0.0419.$$

The estimates and variances for the remaining observations are given in Table 4.6.1. Note that the variance of the estimation error is approaching 0.0419. This limiting variance, denoted by  $\sigma_v^2$ , was derived in Example 4.5.1 as the mean square error based on an infinite past. The variance of  $w_t$  stabilizes at

$$\begin{aligned}\sigma_w^2 &= \sigma_e^2 + \alpha^2 \sigma_v^2 = 0.172 + (0.81)^2(0.0419) \\ &= 0.1995.\end{aligned}$$

It follows that equation (4.6.12) stabilizes at

$$\hat{X}_t = 5.28 + 0.81(\hat{X}_{t-1} - 5.28) + 0.7901[Y_t - 5.28 - 0.81(\hat{X}_{t-1} - 5.28)],$$

where  $0.7901 = (\sigma_u^2 + \sigma_w^2)^{-1} \sigma_w^2$  and  $\alpha = 0.81$ . ▲▲

**Example 4.6.2.** We investigate estimation of  $X_t$  for the Des Moines River

example under the assumption that the mean of the process is unknown. We retain the assumption that the other parameters of the model are known. We write the model as

$$\begin{aligned} Z_t &= \alpha Z_{t-1} + e_{tt}, \\ \mu_t &= \mu, \\ Y_t &= \mu + Z_t + u_t, \end{aligned} \quad (4.6.25)$$

where  $\mu$  is the mean of the process. The first two equations of (4.6.25) are the state equations and the last equation is the measurement equation. In terms of the model (4.6.15)–(4.6.17),  $\sigma_{uuu} = \sigma_{uu} = 0.053$ ,

$$\begin{aligned} A_t &= A = \text{diag}(\alpha, 1) = \text{diag}(0.81, 1), \\ \Sigma_{eet} &= \Sigma_{ee} = \text{diag}(0.172, 0), \end{aligned}$$

$X'_t = (Z_t, \mu)$ ,  $e'_t = (e_{tt}, 0)$ , and  $H_t = (1, 1)$ . Under the model, each  $Y_t$  is unbiased for  $\mu$ , and the variance of  $Y_t - \mu$  is  $\sigma_{zz} + \sigma_{uu}$ .

To initiate the filter, we use the knowledge that  $Z_t$  is a random variable with mean zero and variance  $\sigma_{zz} = \sigma_z^2$ . Letting  $Y_0$  be the first observation of Table 4.6.2, we can form the system of equations

$$0 = Z_0 + v_0, \quad (4.6.26)$$

$$Y_0 = \mu + Z_0 + u_0. \quad (4.6.27)$$

We are treating  $\mu$  as a fixed unknown parameter, so we do not have an initial equation of the type (4.6.26) for  $\mu$ . The first real observation furnishes the first information about  $\mu$ . From the system (4.6.26)–(4.6.27), we obtain the estimator

$$\hat{X}'_0 = (\hat{Z}_0, \hat{\mu}_0) = (0, Y_0)$$

with covariance matrix

$$\Sigma_{vv00} = \begin{pmatrix} \sigma_{zz} & -\sigma_{zz} \\ -\sigma_{zz} & \sigma_{zz} + \sigma_{uu} \end{pmatrix},$$

where  $\sigma_{zz} = 0.500$  and  $\sigma_{uu} = 0.053$ .

Using  $\hat{X}'_0 = (0, 5.44)$  and equation (4.6.19), we have

$$\begin{aligned} \begin{pmatrix} \hat{Z}_1 \\ \hat{\mu}_1 \end{pmatrix} &= \begin{pmatrix} 0.46953 & -0.45252 \\ -0.45252 & 0.47901 \end{pmatrix} \begin{pmatrix} 18.868Y_1 + 19.5884 \\ 18.868Y_1 + 24.1832 \end{pmatrix} \\ &= (-0.0193, 5.4100)', \end{aligned}$$

and from (4.6.20),

$$\Sigma_{vv11} = \begin{pmatrix} 0.172 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0.328 & -0.405 \\ -0.405 & 0.553 \end{pmatrix} = \begin{pmatrix} 0.500 & -0.405 \\ -0.405 & 0.553 \end{pmatrix}.$$

If we use (4.6.23),

$$\begin{aligned}\hat{\mathbf{x}}_1 &= \begin{pmatrix} 0.81 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 5.44 \end{pmatrix} + \begin{pmatrix} 0.500 & -0.405 \\ -0.405 & 0.553 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \end{pmatrix} (0.1430)^{-1} (5.38 - 5.44) \\ &= (-0.0193, 5.4100)',\end{aligned}$$

where

$$D_1 = 0.053 + (1, 1) \Sigma_{ww11} (1, 1)' = 0.29605.$$

It follows that

$$\Sigma_{vv11} = \begin{pmatrix} 0.46953 & -0.45252 \\ -0.45252 & 0.47901 \end{pmatrix}.$$

The estimate of the sediment at  $t = 1$  is

$$\hat{S}_1 = \hat{Z}_1 + \hat{\mu}_1 = -0.0193 + 5.4100 = 5.3907.$$

The variance of the error in the estimated sediment at time one is

$$(1, 1) \hat{\Sigma}_{vv11} (1, 1)' = 0.04351.$$

The estimates for  $t = 2$  are

$$\begin{pmatrix} \hat{Z}_2 \\ \hat{\mu}_2 \end{pmatrix} = \begin{pmatrix} 0.43387 & -0.41230 \\ -0.41230 & 0.43367 \end{pmatrix} \begin{pmatrix} 123.116 \\ 129.558 \end{pmatrix} = \begin{bmatrix} -0.0004 \\ 5.4247 \end{bmatrix},$$

where

$$\Sigma_{ww22} = \begin{pmatrix} 0.4801 & -0.3665 \\ -0.3665 & 0.4790 \end{pmatrix}.$$

The estimates of  $S_t = Z_t + \mu_t$ , based on  $Y_t, Y_{t-1}, \dots, Y_0$ , are given in Table 4.6.2. Note that the estimation of the mean contributes modestly to the variance of  $\hat{S}_t - S_t$ . The variance of  $\hat{\mu}_t$  is declining approximately at the rate  $t^{-1}$ . While the variance of the estimator of  $S_t = \mu + Z_t$  will eventually approach 0.0419, the approach is slowed by the estimation error in  $\hat{\mu}_t$ .  $\blacktriangle\blacktriangle$

**Example 4.6.3.** In the previous examples, we applied the Kalman filter to a stationary time series. Stationarity made it relatively easy to find starting values for the filter. In this example, we consider filtering for a time series in which the autoregressive part has a unit root. Let the model be

$$Z_t = \begin{cases} 0, & t = 0, \\ Z_{t-1} + e_t, & t = 1, 2, \dots, \end{cases} \quad (4.6.28)$$

$$Y_t = \theta + Z_t + u_t, \quad t = 0, 1, \dots, \quad (4.6.29)$$

**Table 4.6.2. Estimates of Sediment Constructed with Kalman Filter, Unknown Mean**

$t$	$Y_t$	$\hat{Z}_t + \hat{\mu}_t = \hat{S}_t$	$\hat{\mu}_t$	$V\{\hat{S}_t - S_t\}$	$V\{\hat{\mu}_t - \mu\}$
0	5.44	5.44000	5.44000	0.05300	0.55300
1	5.38	5.39074	5.41001	0.04351	0.47901
2	5.43	5.42324	5.42436	0.04293	0.43367
3	5.22	5.25915	5.35070	0.04280	0.39757
4	5.28	5.27933	5.35185	0.04272	0.36722
5	5.21	5.22621	5.32613	0.04266	0.34121
6	5.23	5.23298	5.32174	0.04261	0.31864
7	5.33	5.31422	5.34347	0.04256	0.29887
8	5.58	5.52856	5.40987	0.04252	0.28141
9	6.18	6.04632	5.57235	0.04249	0.26588
10	6.16	6.11947	5.61890	0.04246	0.25198
11	6.07	6.06090	5.62881	0.04243	0.23945
12	6.56	6.44377	5.74903	0.04240	0.22811
13	5.93	6.00652	5.67363	0.04238	0.21780
14	5.70	5.74886	5.62767	0.04236	0.20838

where  $(e_t, u_t)' \sim NI[0, \text{diag}(\sigma_e^2, \sigma_u^2)]$ . We treat the true part of  $Y_0$ , denoted by  $\theta$ , as a fixed unknown constant to be estimated. The  $Y$ -data of Table 4.6.3 were generated by the model with  $(\sigma_e^2, \sigma_u^2) = (0.25, 0.16)$ . In terms of the model (4.6.15)–(4.6.17),  $\sigma_{uuu} = \sigma_{uu} = 0.16$ ,  $A_t = A = I$ ,

$$X'_t = (Z_t, \theta), \quad e'_t = (e_t, 0),$$

$$H_t = H = (1, 1), \quad \text{and} \quad \Sigma_{ee} = \text{diag}(0.25, 0).$$

**Table 4.6.3. Kalman Filter Applied to a Unit Root Process**

$t$	$Y_t$	$\hat{y}_t = \hat{\theta}_t + \hat{Z}_t$	$\hat{\theta}_t$	$V\{\hat{y}_t - y_t\}$	$V\{\hat{\theta}_t\}$	$C\{\hat{\theta}_t, \hat{y}_t - y_t\}$
0	2.30797	2.30797	2.30797	0.16000	0.16000	0.16000
1	2.54141	2.47588	2.37350	0.11509	0.11509	0.04491
2	3.08044	2.89622	2.42521	0.11125	0.11125	0.01369
3	1.35846	1.83049	2.38483	0.11089	0.11089	0.00420
4	1.55019	1.63629	2.38257	0.11085	0.11085	0.00129
5	2.34068	2.12430	2.38432	0.11085	0.11085	0.00040
6	1.33786	1.57945	2.38372	0.11085	0.11085	0.00012
7	0.98497	1.16759	2.38358	0.11085	0.11085	0.00004
8	1.17314	1.17143	2.38358	0.11085	0.11085	0.00001
9	0.65385	0.81285	2.38357	0.11085	0.11085	0.00000
10	0.35140	0.49315	2.38357	0.11085	0.11085	0.00000
11	0.47546	0.48090	2.38357	0.11085	0.11085	0.00000
12	-0.56643	-0.24470	2.38356	0.11085	0.11085	0.00000
13	0.04359	-0.04497	2.38356	0.11085	0.11085	0.00000
14	-0.25374	-0.18961	2.38356	0.11085	0.11085	0.00000

The model specifies  $Z_0 = 0$  and  $Y_0 = \theta + u_0$ . Therefore, we initiate the filter with

$$\hat{\mathbf{X}}'_0 = (\hat{Z}_0, \hat{\theta}_0) = (0, Y_0) = (0, 2.308)$$

and

$$\Sigma_{vv00} = \text{diag}(0, 0.16).$$

Then

$$\begin{aligned}\Sigma_{ww11} &= \begin{pmatrix} 0.25 & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & 0.16 \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \\ &= \text{diag}(0.25, 0.16), \\ D_1 &= 0.16 + (1, 1)[\text{diag}(0.25, 0.16)](1, 1)' = 0.57,\end{aligned}$$

and

$$\begin{aligned}\begin{pmatrix} \hat{Z}_1 \\ \hat{\theta}_1 \end{pmatrix} &= \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} 0 \\ 2.308 \end{pmatrix} + \begin{pmatrix} 0.4386 \\ 0.2807 \end{pmatrix} (Y_1 - \mathbf{H}\hat{\mathbf{X}}_0) \\ &= (0.1024, 2.3735').\end{aligned}$$

The estimate of  $\theta + Z_1$  is  $\hat{y}_1 = \hat{\theta}_1 + \hat{Z}_1 = 2.4759$ .

The remaining estimates and the variances of the estimators are given in Table 4.6.3. There are several interesting aspects of the results. First, the variance of the estimator of  $\theta$  stabilizes rather quickly. Observations after the first five add very little information about the true value at time zero. Associated with the stabilization of the variance of  $\hat{\theta}_t$  is the fact that the estimate of  $\theta + Z_t$  at time  $t$  depends very little on values of  $Y_\tau$  for  $t - \tau < 5$ . Also, the variance of the estimator of  $\theta$  at time  $t$  is equal to the variance of the estimator of  $\theta + Z_t$ . This is because the estimator of  $\theta$ , expressed as a function of  $(Y_0, Y_1, \dots, Y_t)$ , is the mirror image of the estimator of  $\theta + Z_t$ .

After a few initial observations, the estimation equation stabilizes. For this problem, the limiting updating equation for  $y_t = \theta + Z_t$  can be written as

$$\hat{y}_t = \hat{y}_{t-1} + c(Y_t - \hat{y}_{t-1}), \quad (4.6.30)$$

where

$$\begin{aligned}c &= (\sigma_u^2 + \sigma_w^2)^{-1} \sigma_w^2, \\ \sigma_w^2 &= 0.5 \sigma_{ee11}^{-1} [1 + (1 + 4\sigma_{ee11}^{-1} \sigma_u^2)^{1/2}], \\ \sigma_v^2 &= \sigma_w^2 - (\sigma_u^2 + \sigma_w^2)^{-1} \sigma_w^4.\end{aligned}$$

The expression for  $\sigma_w^2$  was obtained from the expression following (4.6.10) by setting  $\sigma_w^2 = \sigma_{w1}^2$  and  $\sigma_v^2 = \sigma_{v,t-1}^2$ .

The  $\hat{y}_t$  of (4.6.30) is a convex combination of  $\hat{y}_{t-j}$  and  $Y_t$ . Hence, the sum of the weights on the  $Y_{t-j}$ ,  $j = 0, 1, \dots, t$ , that define  $\hat{y}_t$  is one. This means that estimates constructed with this model maintain the level of the original series. If the true  $Y_t$  contains a positive time trend, the predictor (4.6.30) will have a negative bias.  $\blacktriangle\blacktriangle$

In our discussion to this point, we have used information on the  $Y$ -process through  $Y_t$  to construct an estimator for  $X_t$ . It follows from (4.6.15) that predictors of future values can be constructed from the estimator of  $X_t$ . For example, the predictor of  $X_{t+1}$  constructed with data through  $Y_t$  is

$$\hat{X}_{t+1|t} = A_{t+1} \hat{X}_t \quad (4.6.31)$$

and the variance of the prediction error  $V\{\hat{X}_{t+1|t} - X_{t+1}\}$  is

$$\Sigma_{t+1|t} = \Sigma_{ww,t+1,t+1} = A_{t+1} \Sigma_{vvtt} A'_{t+1} + \Sigma_{ee,t+1,t+1}. \quad (4.6.32)$$

The formula (4.6.31) can be applied recursively to obtain predictions for any number of periods. Thus,

$$\hat{X}_{t+l|t} = A_{t+l} \hat{X}_{t+l-1|t},$$

and the variance of the prediction error is

$$\Sigma_{t+l|t} = A_{t+l} \Sigma_{t+l-1|t} A'_{t+l} + \Sigma_{ee,t+l,t+l}.$$

The prediction formulas can also be used in constructing estimators of  $X_r$  when data are missing. Assume that  $(Y_1, Y_2, \dots, Y_{r-1}, Y_{r+1})$  is available, that the Kalman filter has been initiated prior to time  $r-1$ , and that the objective is to estimate  $X_{r+1}$ . At time  $r$ , the best estimator of  $X_r$  is

$$\hat{X}_r = \hat{X}_{r|r-1} = A_r \hat{X}_{r-1},$$

and  $V\{X_r - A_r \hat{X}_{r-1}\} = \Sigma_{wwrr}$ . Because there is no  $Y$ -information at time  $r$ , we have  $\Sigma_{wwrr} = \Sigma_{vurr}$ . That is, the error in the estimator of  $X_r$  constructed with  $\hat{X}_{r-1}$  is the final estimation error. At time  $r+1$ , when  $Y_{r+1}$  is available, the best estimator of  $X_{r+1}$  is

$$\hat{X}_{r+1} = A_{r+1} \hat{X}_r + \Sigma_{ww,r+1,r+1} H'_{r+1} D_{r+1}^{-1} (Y_{r+1} - H_{r+1} A_{r+1} \hat{X}_r),$$

where  $\Sigma_{ww,r+1,r+1}$ ,  $D_{r+1}$ , and  $\Sigma_{vv,r+1,r+1}$  are given by (4.6.20), (4.6.22), and (4.6.21), respectively.

**Example 4.6.4.** We use the Kalman filter and the data of Example 4.6.1 to construct estimators in the presence of missing data. We assume that the mechanism that causes data to be missing is independent of the  $(Y_t, X_t)$  process.

Table 4.6.4 contains the data of Table 4.6.1 with observations 7, 11, and 12

**Table 4.6.4. Estimates of Sediment Constructed with Kalman Filter**

<i>t</i>	<i>Y<sub>t</sub></i>	$\hat{X}_t$	$\sigma_{v_t}^2$	$\sigma_w^2$
1	5.44	5.42467	0.04792	0.50000
2	5.38	5.38355	0.04205	0.20344
3	5.43	5.41613	0.04188	0.19959
4	5.22	5.25574	0.04187	0.19948
5	5.28	5.27588	0.04187	0.19947
6	5.21	5.22399	0.04187	0.19947
7		5.23463	0.19947	0.19947
8	5.33	5.31708	0.04511	0.30287
9	5.58	5.52380	0.04197	0.20160
10	6.18	6.03256	0.04188	0.19954
11		5.88957	0.19948	0.19948
12		5.77375	0.30288	0.30288
13	6.56	6.44992	0.04637	0.37072
14	5.93	5.99176	0.04200	0.20242
15	5.70	5.73285	0.04188	0.19956

missing. Recall that the model of Example 4.6.1 is

$$\begin{aligned} X_t - 5.28 &= 0.81(X_{t-1} - 5.28) + e_t, \\ Y_t &= X_t + u_t, \end{aligned}$$

where  $(e_t, u_t)' \sim \text{II}[\mathbf{0}, \text{diag}(0.172, 0.053)]$ . Given the previous data, the predictor of  $X_7$  is

$$\hat{X}_7 = 5.28 + 0.81(5.22399 - 5.28) = 5.23463,$$

and the variance of the estimation error is  $\sigma_{v7}^2 = 0.19947$ . The estimator for  $X_8$  is

$$\begin{aligned} \hat{X}_8 &= 5.28 + 0.81(-0.04537) + (0.30287)(0.35587)^{-1}(0.05 + 0.03675) \\ &= 5.31708, \end{aligned}$$

where  $Y_8 - 5.28 = 0.05$ ,

$$\begin{aligned} \sigma_{w8}^2 &= \sigma_e^2 + \alpha^2 \sigma_{v7}^2 = 0.172 + (0.81)^2 0.19947 = 0.30287, \\ D_8 &= 0.053 + 0.30287 = 0.35587, \\ \sigma_{v8}^2 &= 0.30287 - (0.30287)^2 (0.35587)^{-1} = 0.04511. \end{aligned}$$

The calculations are analogous for  $\hat{X}_{11}$ . The estimator for  $X_{12}$  is

$$\hat{X}_{12} = 5.28 + 0.81(5.88957 - 5.28) = 5.77375,$$

where

$$\sigma_{w,12}^2 = \sigma_e^2 + \alpha^2 \sigma_{w,11}^2 = 0.172 + (0.81)^2 0.19948 = 0.30288.$$

The estimator for  $X_{13}$  given in Table 4.6.4 is constructed with

$$(\sigma_{w,13}^2, D_{13}, \sigma_{v,13}^2) = (0.37072, 0.42372, 0.04637).$$

These calculations illustrate the importance of  $Y_t$  in the estimation of  $X_t$  for these data.  $\blacktriangle\blacktriangle$

In some situations, it is useful to express an autoregressive moving average in the state space form. The vector first order autoregressive process is in the form (4.6.15) with a fixed matrix  $A$  and zero measurement error. Therefore, because any autoregressive model can be put in the vector first order form (see Section 2.8), any pure autoregressive model can be put in the state space form. We shall see that there are alternative state space representations.

Akaike (1974) suggested a state space representation for autoregressive moving averages. To introduce the ideas, consider a second order moving average

$$Y_t = \beta_2 \epsilon_{t-2} + \beta_1 \epsilon_{t-1} + \epsilon_t, \quad (4.6.33)$$

where the  $\epsilon_t$  are independent  $(0, \sigma_e^2)$  random variables. Let

$$\begin{aligned} \mathbf{X}'_t &= (E\{Y_t | t\}, E\{Y_{t+1} | t\}, E\{Y_{t+2} | t\}), \\ &= (Y_t, \beta_2 \epsilon_{t-1} + \beta_1 \epsilon_t, \beta_2 \epsilon_t), \end{aligned} \quad (4.6.34)$$

where  $E\{Y_t | t\}$  is the expected value of  $Y_t$  given an infinite past  $(Y_t, Y_{t-1}, \dots)$ . The vector  $\mathbf{X}_t$  contains information equivalent to the vector  $(\epsilon_t, \epsilon_{t-1})$ , which is the information required for any future predictions of  $Y_t$ . Furthermore,  $\mathbf{X}_t$  satisfies the first order vector autoregression

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{e}_t, \quad (4.6.35)$$

where

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix},$$

and  $\mathbf{e}'_t = (1, \beta_1, \beta_2)\epsilon_t$ . The state space representation of (4.6.18) is completed by adding the "observation equation"

$$Y_t = (1, 0, 0)\mathbf{X}_t = \mathbf{H}\mathbf{X}_t, \quad (4.6.36)$$

to (4.6.35).

To initiate the filter for the second order moving average, we begin with the

vector  $\hat{\mathbf{X}}'_0 = (0, 0, 0)$ . The covariance matrix of  $\hat{\mathbf{X}}_0 - \mathbf{X}_0$  is

$$\Sigma_{vv00} = \begin{pmatrix} 1 + \beta_1^2 + \beta_2^2 & \beta_1 + \beta_1\beta_2 & \beta_2 \\ \beta_1 + \beta_1\beta_2 & \beta_2^2 + \beta_1^2 & \beta_1\beta_2 \\ \beta_2 & \beta_1\beta_2 & \beta_2^2 \end{pmatrix} \sigma_\epsilon^2.$$

The reader may verify that

$$\begin{aligned} \Sigma_{ww11} &= \Sigma_{\epsilon\epsilon} + \mathbf{A}\Sigma_{vv00}\mathbf{A}' = \Sigma_{vv00}, \\ D_1 &= \mathbf{H}\Sigma_{ww11}\mathbf{H}' = (1 + \beta_1^2 + \beta_2^2)\sigma_\epsilon^2, \\ \Sigma_{\epsilon\epsilon} &= (1, \beta_1, \beta_2)'(1, \beta_1, \beta_2)\sigma_\epsilon^2, \\ \hat{\mathbf{X}}'_1 &= (1 + \beta_1^2 + \beta_2^2)^{-1}(1 + \beta_1^2 + \beta_2^2, \beta_1 + \beta_1\beta_2, \beta_2)Y_1, \end{aligned} \quad (4.6.37)$$

and

$$\hat{\mathbf{X}}_{2|1} = \mathbf{A}\hat{\mathbf{X}}_1 = (1 + \beta_1^2 + \beta_2^2)^{-1}(\beta_1 + \beta_1\beta_2, \beta_2, 0)'Y_1. \quad (4.6.38)$$

The vector  $\hat{\mathbf{X}}'_1$  is the best predictor of  $(Y_1, Y_2, Y_3)$  given  $Y_1$ , and the vector  $\hat{\mathbf{X}}_{2|1}$  is the best predictor of  $(Y_2, Y_3, Y_4)$  given  $Y_1$ . The predictor of  $(Y_2, Y_3, Y_4)$  could have been obtained directly from the regression of  $(Y_2, Y_3, Y_4)$  on  $Y_1$ . The predictor for  $\mathbf{X}_2$  given  $(Y_2, Y_1)$ , and subsequent predictions can be constructed using the equations (4.6.20), (4.6.21), and (4.6.23).

We now give a state space representation for a stationary autoregressive moving average of order  $(p, q)$ . Let

$$Y_t = \sum_{j=1}^p \alpha_j Y_{t-j} + \sum_{i=1}^q \beta_i \epsilon_{t-i} + \epsilon_t, \quad (4.6.39)$$

where the  $\epsilon_t$  are independent  $(0, \sigma_\epsilon^2)$  random variables. The vector of conditional expectations,  $\mathbf{X}_t$ , of (4.6.34) becomes

$$\mathbf{X}_t = (E\{Y_t | t\}, E\{Y_{t+1} | t\}, \dots, E\{Y_{t+m-1} | t\})', \quad (4.6.40)$$

where  $m = \max(p, q + 1)$ . From Theorem 2.7.1, we have

$$Y_t = \sum_{i=0}^{\infty} \psi_i \epsilon_{t-i}, \quad (4.6.41)$$

where the  $\psi_i$  are defined in that theorem. It follows that

$$E\{Y_{t+j} | t\} = \sum_{i=j}^{\infty} \psi_i \epsilon_{t-i+j}, \quad (4.6.42)$$

for  $j = 1, 2, \dots, m - 1$ . From equations (4.6.41) and (4.6.42), we have

$$\mathbf{X}_t = \mathbf{A}\mathbf{X}_{t-1} + \mathbf{e}_t, \quad (4.6.43)$$

where

$$\begin{aligned} \mathbf{e}_t &= (1, v_1, v_2, \dots, v_{m-1})' \epsilon_t, \\ \mathbf{A} &= \begin{pmatrix} 0 & 1 & 0 & \cdots & 0 \\ 0 & 0 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \\ \alpha_m & \alpha_{m-1} & \alpha_{m-2} & \cdots & \alpha_1 \end{pmatrix}, \end{aligned}$$

and it is understood that  $\alpha_m, \dots, \alpha_{p+1}$  are zero if  $m > p$ . Equation (4.6.43) and the equation

$$Y_t = (1, 0, \dots, 0)\mathbf{X}_t, \quad (4.6.44)$$

form a state space representation for a stationary autoregressive moving average process.

**Example 4.6.5.** In this example, we use the Kalman filter to construct predictions for the stationary autoregressive moving average model

$$Y_t = 0.80Y_{t-1} + \epsilon_t + 0.60\epsilon_{t-1} + 0.58\epsilon_{t-2}, \quad (4.6.45)$$

where  $\epsilon_t \sim NI(0, 1)$ . The second column of Table 4.6.5 contains observations generated by the process (4.6.45). By Theorem 2.7.1,

$$Y_t = \sum_{i=0}^{\infty} v_i \epsilon_{t-i},$$

where  $(v_0, v_1, v_2) = (1.00, 1.40, 1.70)$ . The state space representation of the model (4.6.45) is

$$\begin{aligned} \mathbf{X}_t &= \mathbf{A}\mathbf{X}_{t-1} + \mathbf{e}_t, \\ Y_t &= (1, 0, 0)\mathbf{X}_t, \end{aligned} \quad (4.6.46)$$

where

$$\begin{aligned} \mathbf{X}'_t &= (Y_t, E\{Y_{t+1} | t\}, E\{Y_{t+2} | t\}), \\ \mathbf{e}'_t &= (1.00, 1.40, 1.70)\epsilon_t, \end{aligned}$$

**Table 4.6.5. Kalman Filter Used to Construct Predictors for an Autoregressive Moving Average**

$t$	$Y_t$	$\hat{X}_{t+1,1 t}$	$\hat{X}_{t+1,2 t}$	$V\{\hat{X}_{t+1,1 t} - Y_{t+1}\}$	$V\{\hat{X}_{t+1,2 t} - Y_{t+2}\}$
1	3.240	3.008	2.578	1.515	4.033
2	1.643	0.699	0.036	1.162	3.200
3	2.521	2.457	2.875	1.150	3.170
4	3.122	3.779	3.359	1.049	3.074
5	3.788	3.371	2.702	1.032	3.001
6	2.706	1.782	1.052	1.024	3.001
7	4.016	4.169	4.601	1.008	2.977
8	5.656	6.680	6.199	1.007	2.969
9	6.467	5.903	4.600	1.004	2.968
10	7.047	6.201	5.622	1.002	2.963
11	4.284	2.939	1.241	1.002	2.962
12	2.587	0.749	0.395	1.001	2.961
13	-0.421	-1.242	-1.671	1.000	2.960
14	0.149	0.276	1.028	1.000	2.960
15	-1.012	-0.776	-1.368	1.000	2.960

NOTE:  $\hat{X}_{t+1,1|t} = \hat{Y}_{t+1}(Y_1, \dots, Y_t)$  and  $\hat{X}_{t+1,2|t} = \hat{Y}_{t+2}(Y_1, \dots, Y_t)$ .

and

$$\mathbf{A} = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0.8 \end{pmatrix}.$$

Multiplying (4.6.45) by  $Y_t$  and  $Y_{t-1}$ , and taking expectations, we have

$$\begin{aligned} \gamma_Y(0) - 0.80\gamma_Y(1) &= 2.8260, \\ \gamma_Y(1) - 0.80\gamma_Y(0) &= 1.4120. \end{aligned}$$

It follows that  $[\gamma_Y(0), \gamma_Y(1)] = [10.9878, 10.2022]$ . If we initiate the filter with

$$\hat{\mathbf{X}}'_0 = (0, 0, 0),$$

the covariance matrix of  $\hat{\mathbf{X}}_0 - \mathbf{X}_0$  is the variance of  $[Y_t, E\{Y_{t+1}|t\}, E\{Y_{t+2}|t\}]$ , namely,

$$\boldsymbol{\Sigma}_{vv00} = \begin{pmatrix} 10.988 & 10.202 & 8.742 \\ 10.202 & 9.988 & 8.802 \\ 8.742 & 8.802 & 8.028 \end{pmatrix}.$$

To construct the estimator of  $\mathbf{X}_1$  given  $\hat{\mathbf{X}}_0$  and  $Y_t$ , we compute

$$\begin{aligned} \boldsymbol{\Sigma}_{ww11} &= \boldsymbol{\Sigma}_{ee} + \mathbf{A}\boldsymbol{\Sigma}_{vv00}\mathbf{A}' = \boldsymbol{\Sigma}_{vv00}, \\ D_1 &= (1, 0, 0)\boldsymbol{\Sigma}_{ww11}(1, 0, 0)' = \gamma_Y(0) = 10.988, \end{aligned}$$

and

$$\begin{aligned}\hat{\mathbf{X}}'_1 &= (1.000, 0.929, 0.796)\mathbf{Y}_1 \\ &= (3.240, 3.008, 2.578).\end{aligned}\quad (4.6.47)$$

The reader can verify that the coefficients in (4.6.47) are  $\gamma_Y^{-1}(0)[\gamma_Y(0), \gamma_Y(1), \gamma_Y(2)]$ . The covariance matrix of  $\hat{\mathbf{X}}_1 - \mathbf{X}_1$  is

$$\Sigma_{vv11} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0.515 & 0.685 \\ 0 & 0.685 & 1.073 \end{pmatrix}.$$

The predictor of  $\mathbf{X}_2$  given  $\hat{\mathbf{X}}_1$  is

$$\hat{\mathbf{X}}_{2|1} = \mathbf{A}\hat{\mathbf{X}}_1,$$

and the covariance matrix of the prediction error is

$$\Sigma_{2|1} = \begin{pmatrix} 1.515 & 2.085 & 2.248 \\ 2.085 & 3.033 & 3.238 \\ 2.248 & 3.238 & 3.577 \end{pmatrix},$$

where  $\Sigma_{2|1} = \mathbf{A}\Sigma_{vv11}\mathbf{A}' + \Sigma_{ee}$ . The quantities required to construct  $\hat{\mathbf{X}}_2$  are

$$\Sigma_{ww22} = \begin{pmatrix} 1.515 & 2.085 & 2.248 \\ 2.085 & 3.033 & 3.238 \\ 2.248 & 3.238 & 3.577 \end{pmatrix},$$

and

$$D_2 = \mathbf{H}_2 \Sigma_{ww22} \mathbf{H}'_2 = 1.515.$$

The predictions and variances of the prediction errors are given in the third and fifth columns, respectively, of Table 4.6.5. As the initial effects die out, the estimator of  $\mathbf{X}_t$  stabilizes at

$$\begin{aligned}\hat{\mathbf{X}}_t &= \mathbf{A}\hat{\mathbf{X}}_{t-1} + (1.0, 1.4, 1.7)'[\mathbf{Y}_t - (1, 0, 0)\mathbf{A}\hat{\mathbf{X}}_{t-1}] \\ &= \mathbf{A}\mathbf{X}_{t-1} + (1.0, 1.4, 1.7)'e_t.\end{aligned}$$

The predictor stabilizes at

$$\hat{\mathbf{X}}_{t+1|t} = \mathbf{A}\hat{\mathbf{X}}_t = \mathbf{A}\mathbf{X}_t,$$

and the covariance matrix of the prediction error stabilizes at

$$\Sigma_{t+1|t} = \mathbf{V}\{\hat{\mathbf{X}}_{t+1|t} - \mathbf{X}_{t+1}\} = \Sigma_{ee}.$$

The second entry in  $\hat{\mathbf{X}}_{t+1|t}$  is the predictor of  $E\{Y_{t+2}|t+1\}$ , based on

observations through time  $t$ . Thus, it is also the predictor of  $Y_{t+2}$  based on observations through time  $t$ . The limiting variance of  $\hat{X}_{t+1,2|t} - Y_{t+2}$  is  $(1 + v_1^2)\sigma^2$ , while the second entry on the diagonal of  $\Sigma_{ee}$  is  $v_1^2\sigma^2$ . In the last column of Table 4.6.5 we give

$$V\{\hat{X}_{t+1,2|t} - Y_{t+2}\} = V\{\hat{X}_{t+1,2|t} - X_{t+1,2}\} + 1,$$

where  $V\{\hat{X}_{t+1,2|t} - X_{t+1,2}\}$  is the second entry on the diagonal of  $\Sigma_{t+1|t}$  and  $\sigma^2 = 1$ .  $\blacktriangle \blacktriangle$

## REFERENCES

**Sections 4.1, 4.3–4.5.** Anderson (1971), Beran (1992), Blackman and Tukey (1959), Bloomfield (1976), Brillinger (1975), Brockwell and Davis (1991), Granger and Hatanaka (1964), Hannan (1960, 1970), Harris (1967), Jenkins and Watts (1968), Kendall and Stuart (1966), Koopmans (1974), Whittle (1963), Yaglom (1962).

**Section 4.2.** Amemiya and Fuller (1967), Bellman (1960), Wahba (1968).

**Section 4.6.** Akaike (1974), Anderson and Moore (1979), Diderrich (1985), Duncan and Horn (1972), Jones (1980), Kalman (1960, 1963), Kalman and Bucy (1961), Meinholt and Singpurwalla (1983), Sage (1968), Sallas and Harville (1981), Sorenson (1966).

## EXERCISES

1. Which of the following functions is the spectral density of a stationary time series? Explain why or why not.
  - $f(\omega) = 1 - \frac{1}{2}\omega^2$ ,  $-\pi \leq \omega \leq \pi$ .
  - $f(\omega) = 1 + \frac{1}{2}\omega$ ,  $-\pi \leq \omega \leq \pi$ .
  - $f(\omega) = 476 + \cos 13\omega$ ,  $-\pi \leq \omega \leq \pi$ .
2. Let  $\{X_t : t \in (0, \pm 1, \pm 2, \dots)\}$  be defined by  $X_t = e_t + 0.4e_{t-1}$ . Compute the autocovariance function  $\gamma(h)$  and the spectral density  $f(\omega)$ , given that the  $e_t$  are independently and identically distributed  $(0, \sigma^2)$  random variables.
3. Give the spectral density for the time series defined by

$$X_t - \beta X_{t-1} = e_t + \alpha_1 e_{t-1} + \alpha_2 e_{t-2}, \quad t = 0, \pm 1, \pm 2, \dots,$$

where  $|\beta| < 1$ , the roots of  $m^2 + \alpha_1 m + \alpha_2 = 0$  are less than one in absolute value, and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables.

4. Let

$$X_t + \alpha_1 X_{t-1} + \alpha_2 X_{t-2} = Z_t$$

and

$$Z_t + \beta_1 Z_{t-1} + \beta_2 Z_{t-2} = e_t,$$

where the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables, and the roots of both  $m^2 + \alpha_1 m + \alpha_2 = 0$  and  $r^2 + \beta_1 r + \beta_2 = 0$  are less than one in absolute value. Give an expression for the spectral density of  $X_t$ . How do you describe the time series  $X_t$ ?

5. Find the covariance function and spectral distribution function for the time series

$$X_t = u_1 \cos t + u_2 \sin t + Y_t,$$

where  $(u_1, u_2)'$  is distributed as a bivariate normal random variable with zero mean and diagonal covariance matrix  $\text{diag}(2, 2)$ ,  $Y_t = e_t - e_{t-1}$ , and the  $e_t$  are independent  $(0, 3)$  random variables, independent of  $(u_1, u_2)'$ .

6. Given the following spectral distribution function:

$$F_X(\omega) = \begin{cases} \pi + \omega, & -\pi \leq \omega < -\pi/2, \\ 5\pi + \omega, & -\pi/2 \leq \omega < \pi/2, \\ 9\pi + \omega, & \pi/2 \leq \omega \leq \pi. \end{cases}$$

What is the variance of  $X_t$ ? What is the spectral distribution function of  $X_t - X_{t-1}$ ? Is there a  $k$  such that  $X_t - X_{t-k}$  will have a continuous spectral distribution function?

7. Prove Corollary 4.3.1.3.

8. Let  $\{e_t; t \in (0, \pm 1, \pm 2, \dots)\}$  be a time series of uncorrelated  $(0, \sigma^2)$  random variables. Let  $X_t = e_{t-2} + 0.5e_{t-3}$ . Give the covariance matrix  $\Gamma(h)$  and the spectral matrix  $f(\omega)$  for  $(X_t, e_t)'$ .

9. Let  $\{e_{1,t}\}$  and  $\{e_{2,t}\}$  be two independent sequences of uncorrelated random variables with variances 0.34 and 0.50 respectively. Let

$$\begin{aligned} X_{1,t} &= 0.8X_{1,t-1} + e_{1,t}, \\ X_{2,t} &= X_{1,t} + e_{2,t}. \end{aligned}$$

Find  $\gamma_{22}(h)$ ,  $\gamma_{12}(h)$ ,  $f_{12}(\omega)$ , and  $f_{22}(\omega)$ .

10. The complex vector random variable  $\mathbf{X}$  of dimension  $q$  is distributed as a complex multivariate normal if the real vector  $[(\text{Re } \mathbf{X})', (\text{Im } \mathbf{X})']'$  is distributed as a multivariate normal with mean  $[(\text{Re } \boldsymbol{\mu})', (\text{Im } \boldsymbol{\mu})']'$  and covariance

matrix

$$\frac{1}{2} \begin{pmatrix} \operatorname{Re} \Sigma & -\operatorname{Im} \Sigma \\ \operatorname{Im} \Sigma & \operatorname{Re} \Sigma \end{pmatrix}.$$

where  $\mu = E\{\mathbf{X}\}$  and  $\Sigma$  is a positive semidefinite Hermitian matrix. Let  $(A_1, B_1, A_2, B_2)'$  be the multivariate normal random variable defined following (4.4.9). Show that  $\mathbf{X} = (A_1 + eB_1, A_2 + eB_2)'$  is distributed as a complex bivariate normal random variable. Give the Hermitian matrix  $\Sigma$ .

11. Let  $\{a_j\}_{j=-\infty}^{\infty}$  and  $\{b_j\}_{j=-\infty}^{\infty}$  be absolutely summable, and let  $\{X_t: t \in (0, \pm 1, \pm 2, \dots)\}$  be a stationary time series with absolutely summable covariance function. Consider the following filtering operation: (1) apply the filter  $\{a_j\}$  to the time series  $X_t$ , to obtain a time series  $Z_t$ , and then (2) apply the filter  $\{b_j\}$  to  $Z_t$ , to obtain a time series  $Y_t$ . What is the transfer function of this filtering operation? Express the spectral density of  $Y_t$  as a function of the spectral density of  $X_t$  and of the transfer function.
12. Let  $\{a_j\}$ ,  $\{b_j\}$ , and  $\{X_t\}$  be as defined in Exercise 11, and define

$$Y_t = \sum_{j=-\infty}^{\infty} (a_j + b_j) X_{t-j}.$$

Express the spectral density of  $Y_t$  as a function of the spectral density of  $X_t$  and the transfer functions of  $\{a_j\}$  and  $\{b_j\}$ .

13. Let  $X_t$  and  $Y_t$  be defined by

$$\begin{aligned} X_t &= 0.9X_{t-1} + e_t, \\ Y_t &= X_t + u_t, \end{aligned}$$

where  $\{e_t\}$  is a sequence of normal independent  $(0, 1)$  random variables independent of the sequence  $\{u_t\}$ . Let  $u_t$  satisfy the difference equation

$$u_t + 0.5u_{t-1} = v_t,$$

where  $\{v_t\}$  is a sequence of normal independent  $(0, 0.3)$  random variables. Assuming that only  $Y_t$  is observed, construct the filter  $\{a_j: j = -2, -1, 0, 1, 2\}$  so that

$$\sum_{j=-2}^2 a_j Y_{t-j}$$

is the minimum mean square error estimator of  $X_t$ . Construct the one-sided filter  $\{b_j: j = 0, 1, \dots, 5\}$  to estimate  $X_t$ . How does the mean square error of these filters compare with the lower bound for linear filters?

14. Let  $f(\omega)$  be an even nonnegative continuous periodic function of period  $2\pi$ . Let

$$a(h) = \int_{-\pi}^{\pi} f(\omega) e^{-i\omega h} d\omega .$$

Show that, for  $q$  a positive integer,

$$\gamma(h) = \begin{cases} \frac{q - |h|}{q} a(h), & h = 0, \pm 1, \pm 2, \dots, \pm q, \\ 0 & \text{otherwise} \end{cases}$$

is the covariance function of a stationary time series. (*Hint:* See Exercise 3.15 and Theorem 3.1.10.)

15. Let  $X_t$  be a time series with continuous spectral density  $f_X(\omega)$ . Let  $Y_t$  be a time series satisfying

$$\sum_{j=0}^p \alpha_j Y_{t-j} = e_t ,$$

$$|f_Y(\omega) - f_X(\omega)| < \epsilon ,$$

where the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables,  $\alpha_0 = 1$ , and  $Y_t$  is defined by Theorem 4.3.4.

- (a) Show that  $|\gamma_X(h) - \gamma_Y(h)| < 2\pi\epsilon$  for all  $h$ .
- (b) Let  $f_X(\omega)$  be strictly positive. Prove that given  $\epsilon > 0$  there is a  $p$  and a set  $\{\alpha_j : j = 0, 1, \dots, p\}$  with  $\alpha_0 = 1$  such that the time series  $Z_t$  defined by

$$Z_t = \sum_{j=0}^p \alpha_j X_{t-j}$$

satisfies

$$\left| f_Z(\omega) - \frac{\sigma^2}{2\pi} \right| < \epsilon$$

for all  $\omega$ , where  $f_Z(\omega)$  is the spectral density of  $Z_t$  and  $\gamma_Z(0) = \sigma^2$ . Show that

$$\sum_{h=1}^{\infty} \gamma_Z^2(h) \leq 2\pi\epsilon^2 .$$

- (c) Let  $f_X(\omega)$  be strictly positive. Show that, given  $\epsilon > 0$ , one may define two autoregressive time series  $Y_{1t}$  and  $Y_{2t}$  with spectral densities

$$f_{Y_1}(\omega) = \frac{\sigma_1^2}{2\pi |\sum_{j=0}^{p_1} \alpha_{1j} e^{-i\omega j}|^2} ,$$

$$f_{Y_2}(\omega) = \frac{\sigma_2^2}{2\pi |\sum_{j=0}^{p_2} \alpha_{2j} e^{-i\omega j}|^2} ,$$

such that

$$f_X(\omega) - \epsilon \leq f_{Y_1}(\omega) \leq f_X(\omega) \leq f_{Y_2}(\omega) \leq f_X(\omega) + \epsilon.$$

- (d) For the three time series defined in part (c) prove that

$$\text{Var}\left\{\sum_{t=1}^n a_t Y_{1t}\right\} \leq \text{Var}\left\{\sum_{t=1}^n a_t X_t\right\} \leq \text{Var}\left\{\sum_{t=1}^n a_t Y_{2t}\right\}$$

for any fixed real numbers  $\{a_t: t = 1, 2, \dots, n\}$ .

16. Let

$$\begin{aligned} X_{1t} &= e_t - 0.8e_{t-1}, \\ X_{2t} &= u_t - 0.9u_{t-4}, \end{aligned}$$

where  $\{e_t\}$  is a sequence of independent  $(0, 1)$  random variables independent of  $\{u_t\}$ , a sequence of independent  $(0, 6)$  random variables. Express

$$Y_t = X_{1t} + X_{2t}$$

as a moving average process.

17. Let  $Y_t$  be defined by

$$\begin{aligned} Y_t &= S_t + Z_t, \\ S_t &= 0.9 S_{t-4} + u_t, \\ Z_t &= 0.8 Z_{t-1} + e_t, \end{aligned}$$

where the sequence  $\{(e_t, u_t)'\}$  is a sequence of normal independent  $(0, \Sigma)$  random variables with  $\Sigma = \text{diag}(0.1, 0.6)$ . Construct the optimum filter  $\{a_j: j = -9, -8, \dots, 8, 9\}$  to estimate  $S_t$ , where the estimator is defined by  $\sum_{j=-9}^9 a_j Y_{t-j}$ . Construct the best one-sided filter  $\{b_j: j = 0, 1, \dots, 8, 9\}$  to estimate  $S_t$ . Compare the mean square error of these filters with the lower bound for linear filters.

18. Let  $\mathbf{C}$  and  $\mathbf{D}$  be  $k \times k$  nonsingular matrices. Show that

$$(\mathbf{C}^{-1} + \mathbf{D}^{-1})^{-1} \mathbf{D}^{-1} = \mathbf{C}(\mathbf{C} + \mathbf{D})^{-1}.$$

19. Let  $X \sim N(0, \sigma_x^2)$ , and let  $Y = X + u$ , where  $u \sim N(0, \sigma_u^2)$  independent of  $X$ . Give the covariance matrix of  $(Y, X)$  and the conditional expected value of  $X$  given  $Y$ . Consider the regression problem

$$(Y, 0)' = (1, 1)'X + (u, \epsilon)',$$

where  $(u, \epsilon)' \sim N[(0, 0)', \text{diag}\{\sigma_u^2, \sigma_x^2\}]$ . Show that the generalized regression

estimator of  $X$  constructed with known  $(\sigma_x^2, \sigma_u^2)$  is the conditional expected value of  $X$  given  $Y$ .

20. Let  $\Sigma_{uu}$  be an  $r \times r$  nonsingular matrix,  $\Sigma_{ww}$  a  $p \times p$  nonsingular matrix, and  $H$  an  $r \times p$  matrix. Show that

$$[H'\Sigma_{uu}^{-1}H + \Sigma_{ww}^{-1}]^{-1} = \Sigma_{ww} - \Sigma_{ww}H'(\Sigma_{uu} + H\Sigma_{ww}H')^{-1}H\Sigma_{ww},$$

and hence verify (4.6.21). [See Duncan and Horn (1972).]

21. Using Exercise 20, verify equation (4.6.23).

22. Let the model (4.6.1)–(4.6.2) hold, and assume it is desired to construct a recursive estimator of  $X_{t+s}$  given the observations  $Y_1, Y_2, \dots, Y_t$ . Construct such an estimator. Consider both  $s > 0$  and  $s < 0$ .

23. Let the following linear model hold

$$\begin{aligned} Y_1 &= X_1\beta + u_1, \\ Y_2 &= X_2\beta + u_2, \end{aligned}$$

where  $Y_1$  is  $n_1 \times 1$ ,  $X_1$  is  $n_1 \times k$ ,  $\beta$  is  $k \times 1$ ,  $Y_2$  is  $n_2 \times 1$ ,  $X_2$  is  $n_2 \times k$ , and  $(u'_1, u'_2)' \sim N[\mathbf{0}, I\sigma^2]$ . Let  $X'_1X_1$  be nonsingular. Let  $X'_1Y_1$  and  $\hat{\beta}_1$ , where

$$\hat{\beta}_1 = (X'_1X_1)^{-1}X'_1Y_1,$$

be given. Construct the best linear unbiased estimator of  $\beta$  as a function of  $\hat{\beta}_1$ ,  $X'_1X_1$ ,  $X_2$ , and  $Y_2$ .

24. Prove Corollary 4.2.2.

25. Prove the following lemma.

**Lemma.** Let  $\Sigma$  be the nonsingular covariance matrix of the vector  $\mathbf{Y}'_t = (\mathbf{Y}'_{t1}, \mathbf{Y}'_{t2}, \mathbf{Y}'_{t3})$ , where

$$\Sigma = \begin{pmatrix} \Sigma_{11} & \Sigma_{12} & \Sigma_{13} \\ \Sigma_{21} & \Sigma_{22} & \Sigma_{23} \\ \Sigma_{31} & \Sigma_{32} & \Sigma_{33} \end{pmatrix}$$

and  $E\{\mathbf{Y}_t\} = \mathbf{0}$ . Let

$$(\hat{\mathbf{Y}}'_{t3|1}, \hat{\mathbf{Y}}'_{t2|1}) = (\mathbf{Y}'_{t1}\Sigma_{11}^{-1}\Sigma_{13}, \mathbf{Y}'_{t1}\Sigma_{11}^{-1}\Sigma_{12})$$

be the best linear estimator of  $(Y'_{t3}, Y'_{t2})$  based upon  $\mathbf{Y}_{t1}$ . Let

$$(\mathbf{e}'_{t2}, \mathbf{e}'_{t3}) = [(\mathbf{Y}_{t2} - \hat{\mathbf{Y}}_{t2|1})', (\mathbf{Y}_{t3} - \hat{\mathbf{Y}}_{t3|1})']'$$

and let  $\mathbf{V}$  be the covariance matrix of  $(\mathbf{e}'_{t2}, \mathbf{e}'_{t3})'$ , where

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{22} & \mathbf{V}_{23} \\ \mathbf{V}_{32} & \mathbf{V}_{33} \end{pmatrix} = \begin{pmatrix} \Sigma_{22} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{12} & \Sigma_{23} - \Sigma_{21}\Sigma_{11}^{-1}\Sigma_{13} \\ \Sigma_{32} - \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{12} & \Sigma_{33} - \Sigma_{31}\Sigma_{11}^{-1}\Sigma_{13} \end{pmatrix}.$$

Then the best linear estimator of  $\mathbf{Y}_{t3}$  given  $(\mathbf{Y}'_{t1}, \mathbf{Y}'_{t2})$  is the best linear estimator of  $\mathbf{Y}_{t3}$  given  $\mathbf{Y}_{t1}$ , plus the best linear estimator of  $\mathbf{e}_{t3|1}$  given  $\mathbf{e}_{t2|1}$ . That is,

$$\hat{Y}_{t3|(1,2)} = \hat{Y}_{t3|1} + \hat{\mathbf{e}}_{t3|2},$$

where  $\hat{\mathbf{e}}'_{t3|2} = \mathbf{e}'_{t2} \mathbf{V}_{22}^{-1} \mathbf{V}_{23}$ , and

$$\hat{Y}_{t3|(1,2)} = (\Sigma_{31}, \Sigma_{32}) \begin{pmatrix} \Sigma_{11} & \Sigma_{12} \\ \Sigma_{21} & \Sigma_{22} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{Y}_{t1} \\ \mathbf{Y}_{t2} \end{pmatrix}.$$

26. Show that the predictor of  $Y_{t+2}$  given  $(Y_1, \dots, Y_t)$  of Example 4.6.4 is the first element of  $\mathbf{A}^2 \hat{\mathbf{X}}_t$ . Show that the variance of the prediction error is the upper left element of

$$\mathbf{A}^2 \Sigma_{vvtt} \mathbf{A}^{2'} + \mathbf{A} \Sigma_{\epsilon\epsilon} \mathbf{A}' + \Sigma_{\epsilon\epsilon}.$$

27. Assume that the time series  $Y_t$  has a spectral density  $f_Y(\omega)$  such that  $f_Y(\omega) > 0$  for all  $\omega$  and such that the Fourier coefficients of  $[f_Y(\omega)]^{-1}$  are absolutely summable. Cleveland (1972) defines the *inverse autocorrelation function* by  $\rho_i(h) = [\gamma_i(0)]^{-1} \gamma_i(h)$ , where

$$\gamma_i(h) = \int_{-\pi}^{\pi} (4\pi^2)^{-1} [f_Y(\omega)]^{-1} e^{i\omega h} d\omega.$$

- (a) Show that if  $Y_t$  is a stationary autoregressive process satisfying

$$Y_t + \sum_{i=1}^p \alpha_i Y_{t-i} = e_t,$$

then the  $\rho_i(h)$  are the autocorrelations of the moving average process

$$X_t = e_t + \sum_{i=1}^p \alpha_i e_{t-i}.$$

- (b) Show that if  $Y_t$  is an ARMA( $p, q$ ) process, then  $\rho_i(h)$  is the autocorrelation of the ARMA( $q, p$ ) process with the autoregressive and moving average coefficients interchanged.

28. Let  $X_t$  have the spectral density

$$f_X(\omega) = \begin{cases} (2\pi)^{-1}, & \pi/4 \leq \omega \leq \pi/2, \\ (2\pi)^{-1}, & -\pi/2 \leq \omega \leq -\pi/4, \\ 0 & \text{otherwise.} \end{cases}$$

The time series  $X_t$  is sometimes called "bandpassed white noise." Give the covariance function of  $X_t$ . Plot the covariance function. Using the covariance function (or otherwise), approximate  $X_t$  by a third order autoregressive process. Plot the spectral density of the third order process with the spectral density of the original process. What would be the error made in predicting  $X_t$  one period ahead using the third order autoregressive process?

Given the availability of a random number generator, explain how you would create 200 observations from the time series  $X_t$  using the "bandpass" idea. Assume you wish to create a normal time series.

29. Let

$$Y_t = S_t + u_t,$$

where  $S_t$  is a stationary time series satisfying

$$S_t = 0.9S_{t-2} + e_t,$$

where the  $e_t$  are  $N(0, 1.9)$  random variables. Let  $u_t$  be a sequence of  $N(0, 1)$  random variables where  $u_t$  is independent of  $S_j$  for all  $t, j$ . Construct a centered two-sided filter of length seven to estimate  $S_t$  given  $(Y_{t-3}, Y_{t-2}, \dots, Y_{t+3})$ . Construct a one-sided filter of length 6 using  $(Y_t, Y_{t-1}, \dots, Y_{t-5})$  to estimate  $S_t$ . For the two-sided filter define

$$X_t = S_t - \hat{S}_t,$$

where  $\hat{S}_t$  is the estimate of  $S_t$  computed using the filter. Compare the variance of  $X_t$  with the lower bound for a two-sided filter.

30. Let  $Y_t$  be a stationary autoregressive process satisfying

$$\sum_{i=0}^p \alpha_i Y_{t-i} = e_t, \quad \text{or} \quad B(\mathcal{B})Y_t = e_t,$$

where  $\alpha_0 = 1$ ,  $B(\mathcal{B}) = \sum_{i=0}^p \alpha_i \mathcal{B}^i$ ,  $e_t$  are independent  $(0, \sigma^2)$  random variables, and  $\mathcal{B}$  is the backshift operator. The process also satisfies

$$\sum_{i=0}^p \alpha_i Y_{t+i} = a_t, \quad \text{or} \quad A(\mathcal{B})Y_t = a_t,$$

where  $A(\mathcal{B}) = \sum_{i=0}^p \alpha_i \mathcal{B}^{p-i}$  and  $a_t$  are uncorrelated  $(0, \sigma^2)$  random variables.

Show that the partial correlation between  $Y_t$  and  $Y_{t-p}$  given  $(Y_{t-p}, \dots, Y_{t-1}, Y_{t+1}, \dots, Y_{t+p})$  is zero for  $j = 1, 2, \dots$ , and that the partial correlation between  $Y_t$  and  $Y_{t-p-j}$  given  $(Y_{t-p}, \dots, Y_{t-1}, Y_{t+1}, \dots, Y_{t+p})$  is zero for  $j = 1, 2, \dots$ .

**31.** Let

$$\begin{aligned} Y_t &= X_t + u_t, \\ X_t &= \theta X_{t-1} + e_t, \end{aligned}$$

where  $e_t \sim NI(0, \sigma_e^2)$ ,  $u_t \sim NI(0, \sigma_u^2)$ ,  $\{e_t\}$  is independent of  $\{u_t\}$ , and  $X_t$  is stationary with  $|\theta| < 1$ . Show that

$$Y_t = \theta Y_{t-1} + v_t + \beta v_{t-1},$$

where  $\beta = -[\sigma_e^2 + (1 + \theta^2)\sigma_u^2]^{-1}\theta\sigma_u^2$ ,  $v_t$  are  $NI(0, \sigma_v^2)$ , and  $(1 + \beta^2)\sigma_v^2 = \sigma_e^2 + (1 + \theta^2)\sigma_u^2$ .

## CHAPTER 5

# Some Large Sample Theory

So far, we have been interested in ways of representing time series and describing their properties. In most practical situations we have a portion of a realization, or of several realizations, and we wish a description (an estimate of the parameters) of the time series.

Most of the presently available results on the estimation of the covariance function, the parameters of autoregressive and moving average processes, and the spectral density rest on large sample theory. Therefore, we shall present some results in large sample statistics.

### 5.1. ORDER IN PROBABILITY

Concepts of relative magnitude or *order of magnitude* are useful in investigating limiting behavior of random variables. We first define the concepts of order as used in real analysis. Let  $\{a_n\}_{n=1}^{\infty}$  be a sequence of real numbers and  $\{g_n\}_{n=1}^{\infty}$  be a sequence of positive real numbers.

**Definition 5.1.1.** We say  $a_n$  is of smaller order than  $g_n$  and write

$$a_n = o(g_n)$$

if

$$\lim_{n \rightarrow \infty} g_n^{-1} a_n = 0.$$

**Definition 5.1.2.** We say  $a_n$  is at most of order  $g_n$  and write

$$a_n = O(g_n)$$

if there exists a real number  $M$  such that  $g_n^{-1}|a_n| \leq M$  for all  $n$ .

The properties of Lemma 5.1.1 are easily established using the definitions of order and the properties of limits.

**Lemma 5.1.1.** Let  $\{a_n\}$  and  $\{b_n\}$  be sequences of real numbers. Let  $\{f_n\}$  and  $\{g_n\}$  be sequences of positive real numbers.

(i) If  $a_n = o(f_n)$  and  $b_n = o(g_n)$ , then

$$\begin{aligned} a_n b_n &= o(f_n g_n), \\ |a_n|^s &= o(f_n^s) \quad \text{for } s > 0, \\ a_n + b_n &= o(\max\{f_n, g_n\}). \end{aligned}$$

(ii) If  $a_n = O(f_n)$  and  $b_n = O(g_n)$ , then

$$\begin{aligned} a_n b_n &= O(f_n g_n), \\ |a_n|^s &= O(f_n^s) \quad \text{for } s \geq 0, \\ a_n + b_n &= O(\max\{f_n, g_n\}). \end{aligned}$$

(iii) If  $a_n = o(f_n)$  and  $b_n = O(g_n)$ , then

$$a_n b_n = o(f_n g_n).$$

**Proof.** Reserved for the reader. ▲

The concepts of order when applied to random variables are closely related to *convergence in probability*.

**Definition 5.1.3.** The sequence of random variables  $\{X_n\}$  converges in probability to the random variable  $X$ , and we write

$$p\lim X_n = X$$

(the *probability limit* of  $X_n$  is  $X$ ), if for every  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} P\{|X_n - X| > \epsilon\} = 0.$$

An equivalent definition is that for every  $\epsilon > 0$  and  $\delta > 0$  there exists an  $N$  such that for all  $n > N$ ,

$$P\{|X_n - X| > \epsilon\} < \delta.$$

The notation

$$X_n \xrightarrow{P} X$$

is also frequently used to indicate that  $X_n$  converges in probability to  $X$ .

For sequences of random variables, definitions of *order in probability* were introduced by Mann and Wald (1943b). Let  $\{X_n\}$  be a sequence of random variables and  $\{g_n\}$  a sequence of positive real numbers.

**Definition 5.1.4.** We say  $X_n$  is of smaller order in probability than  $g_n$  and write

$$X_n = o_p(g_n)$$

if

$$\text{plim } g_n^{-1} X_n = 0.$$

**Definition 5.1.5.** We say  $X_n$  is at most of order in probability  $g_n$  and write

$$X_n = O_p(g_n)$$

if, for every  $\epsilon > 0$ , there exists a positive real number  $M_\epsilon$  such that

$$P\{|X_n| \geq M_\epsilon g_n\} \leq \epsilon$$

for all  $n$ .

If  $X_n = O_p(g_n)$ , we sometimes say that  $X_n$  is *bounded in probability* by  $g_n$ . We define a vector random variable to be  $O_p(g_n)$  if every element of the vector is  $O_p(g_n)$  as follows.

**Definition 5.1.6.** If  $\mathbf{X}_n$  is a  $k$ -dimensional random variable, then  $\mathbf{X}_n$  is at most of order in probability  $g_n$  and we write

$$\mathbf{X}_n = O_p(g_n)$$

if, for every  $\epsilon > 0$ , there exists a positive real number  $M_\epsilon$  such that

$$P\{|X_{jn}| \geq M_\epsilon g_n\} \leq \epsilon, \quad j = 1, 2, \dots, k,$$

for all  $n$ . We say  $\mathbf{X}_n$  is of smaller order in probability than  $g_n$  and write

$$\mathbf{X}_n = o_p(g_n)$$

if, for every  $\epsilon > 0$  and  $\delta > 0$ , there exists an  $N$  such that for all  $n > N$ ,

$$P\{|X_{jn}| > \epsilon g_n\} < \delta, \quad j = 1, 2, \dots, k.$$

Note that  $k$  might be a function of  $n$ , and  $\mathbf{X}_n$  could still satisfy the definition. However, it is clear that the  $M_\epsilon$  of the definition is a function of  $\epsilon$  only (and not of  $n$ ).

A matrix random variable may be viewed as a vector random variable with the elements displayed in a particular manner, or as a collection of vector random variables. Therefore, we shall define the order of matrix random variables in an analogous manner.

**Definition 5.1.7.** A  $k \times r$  matrix  $\mathbf{B}_n$  of random variables is at most of order in

probability  $g_n$ , and we write

$$\mathbf{B}_n = O_p(g_n),$$

if for every  $\epsilon > 0$  there exists a positive real number  $M_\epsilon$  such that

$$P\{|b_{ijn}| \geq M_\epsilon g_n\} \leq \epsilon, \quad i = 1, 2, \dots, k, \quad j = 1, 2, \dots, r,$$

for all  $n$ , where the  $b_{ijn}$  are the elements of  $\mathbf{B}_n$ . We say that  $\mathbf{B}_n$  is of smaller order in probability than  $g_n$  and write

$$\mathbf{B}_n = o_p(g_n)$$

if for every  $\epsilon > 0$  and  $\delta > 0$  there exists an  $N$  such that for all  $n > N$ ,

$$P\{|b_{ijn}| > \epsilon g_n\} < \delta, \quad i = 1, 2, \dots, k, \quad j = 1, 2, \dots, r.$$

For real numbers  $a_i$ ,  $i = 1, \dots, n$ , we know that  $|\sum_{i=1}^n a_i| \leq \sum_{i=1}^n |a_i|$ . The following lemma, resting on this property, furnishes a useful bound on the probability that the absolute value of a sum of random variables exceeds a given number.

**Lemma 5.1.2.** Let  $\mathbf{X}_i$ ,  $i = 1, 2, \dots, n$ , be  $k$ -dimensional random variables. Then, for every  $\epsilon > 0$ ,

$$P\left\{\left|\sum_{i=1}^n \mathbf{X}_i\right| \geq \epsilon\right\} \leq \sum_{i=1}^n P\left\{|\mathbf{X}_i| \geq \frac{\epsilon}{n}\right\}.$$

**Proof.** Let  $\epsilon > 0$  be arbitrary. We see that if  $\sum_{i=1}^n |\mathbf{X}_i| \geq \epsilon$ , then  $|\mathbf{X}_i| \geq \epsilon/n$  for at least one  $i \in \{1, 2, \dots, n\}$ . Therefore,

$$\begin{aligned} P\left\{\left|\sum_{i=1}^n \mathbf{X}_i\right| \geq \epsilon\right\} &\leq P\left\{\sum_{i=1}^n |\mathbf{X}_i| \geq \epsilon\right\} \\ &\leq \sum_{i=1}^n P\left\{|\mathbf{X}_i| \geq \frac{\epsilon}{n}\right\}. \end{aligned} \quad \blacktriangle$$

Definition 5.1.3 applies for vector random variables of fixed dimension as well as for scalar random variables, if it is understood that  $|\mathbf{X}_n - \mathbf{X}|$  is the common Euclidean distance.

**Lemma 5.1.3.** Let  $\mathbf{X}_n$  be a  $k$ -dimensional random variable such that

$$p\lim X_{jn} = X_j, \quad j = 1, 2, \dots, k,$$

where  $X_{jn}$  is the  $j$ th element of  $\mathbf{X}_n$ . Then, for  $k$  fixed,

$$p\lim \mathbf{X}_n = \mathbf{X}.$$

**Proof.** By hypothesis, for each  $j$  and for every  $\epsilon > 0$  and  $\delta > 0$ , there exists an integer  $N_j$  such that for all  $n > N_j$

$$P\{|X_{jn} - X_j| > k^{-1/2}\epsilon\} \leq k^{-1}\delta.$$

Let  $N$  be the maximum of  $\{N_1, N_2, \dots, N_k\}$ . Using Lemma 5.1.2, we have

$$P\{|\mathbf{X}_n - \mathbf{X}| > \epsilon\} \leq \sum_{j=1}^k P\{|X_{jn} - X_j| > k^{-1/2}\epsilon\} \leq \delta$$

for  $n > N$ . ▲

The proof of Lemma 5.1.3 should also help to make it clear that if  $k$  is not fixed, then the fact that  $\mathbf{X}_n = o_p(1)$  does not necessarily imply that  $\text{plim}|\mathbf{X}_n| = 0$ . The vector random variable composed of  $n$  entries all equal to  $n^{-1/2}$  furnishes a counterexample for  $k = n$ .

We shall demonstrate later (Theorems 5.1.5 and 5.1.6) that operations valid for order are also valid for order in probability. Since it is relatively easy to establish the properties analogous to those of Lemma 5.1.1, we do so at this time.

**Lemma 5.1.4.** Let  $\{f_n\}$  and  $\{g_n\}$  be sequences of positive real numbers, and let  $\{X_n\}$  and  $\{Y_n\}$  be sequences of random variables.

(i) If  $X_n = o_p(f_n)$  and  $Y_n = o_p(g_n)$ , then

$$\begin{aligned} X_n Y_n &= o_p(f_n g_n), \\ |X_n|^s &= o_p(f_n^s) \quad \text{for } s > 0, \\ X_n + Y_n &= o_p(\max\{f_n, g_n\}). \end{aligned}$$

(ii) If  $X_n = O_p(f_n)$  and  $Y_n = O_p(g_n)$ , then

$$\begin{aligned} X_n Y_n &= O_p(f_n g_n), \\ |X_n|^s &= O_p(f_n^s) \quad \text{for } s \geq 0, \\ X_n + Y_n &= O_p(\max\{f_n, g_n\}). \end{aligned}$$

(iii) If  $X_n = o_p(f_n)$  and  $Y_n = O_p(g_n)$ , then

$$X_n Y_n = o_p(f_n g_n).$$

**Proof.** We investigate only part i, leaving parts ii and iii as an exercise. By arguments similar to those of Lemma 5.1.3,  $|X_n Y_n| > f_n g_n$  implies that  $|X_n/f_n| > 1$  or (and)  $|Y_n/g_n| > 1$ . By hypothesis, given  $\epsilon > 0$  and  $\delta > 0$ , there is an  $N$  such

that

$$\begin{aligned} P\{|X_n| > \epsilon f_n\} &< 0.5\delta, \\ P\{|Y_n| > \epsilon g_n\} &< 0.5\delta \end{aligned}$$

for  $n > N$ . Therefore,

$$\begin{aligned} P\{|X_n Y_n| > \epsilon^2 f_n g_n\} &\leq P\{|f_n^{-1} X_n| > \epsilon \text{ or } |g_n^{-1} Y_n| > \epsilon\} \\ &\leq P\{|f_n^{-1} X_n| > \epsilon\} + P\{|g_n^{-1} Y_n| > \epsilon\} \\ &< \delta \end{aligned}$$

for  $n > N$ .

The second equality in part i follows from

$$P\{|X_n| > \epsilon f_n\} = P\{|X_n|^s > \epsilon^s f_n^s\},$$

which holds for all  $\epsilon > 0$ .

Let  $q_n = \max\{f_n, g_n\}$ . Given  $\epsilon > 0$  and  $\delta > 0$ , there exists an  $n$  such that

$$\begin{aligned} P\{|X_n| > \frac{1}{2}\epsilon q_n\} &< \frac{1}{2}\delta, \\ P\{|Y_n| > \frac{1}{2}\epsilon q_n\} &< \frac{1}{2}\delta \end{aligned}$$

for  $n > N$ . Hence, the third result of part i follows by Lemma 5.1.2. ▲

One of the most useful tools for establishing the order in probability of random variables is Chebyshev's inequality.

**Theorem 5.1.1 (Chebyshev's inequality).** Let  $r > 0$ , let  $X$  be a random variable such that  $E\{|X|^r\} < \infty$ , and let  $F(x)$  be the distribution function of  $X$ . Then, for every  $\epsilon > 0$  and finite  $A$ ,

$$P\{|X - A| \geq \epsilon\} \leq \frac{E\{|X - A|^r\}}{\epsilon^r}.$$

**Proof.** Let us denote by  $S$  the set of  $x$  for which  $|x - A| \geq \epsilon$  and by  $\tilde{S}$  the set of  $x$  for which  $|x - A| < \epsilon$ . Then,

$$\begin{aligned} \int |x - A|^r dF(x) &= \int_S |x - A|^r dF(x) + \int_{\tilde{S}} |x - A|^r dF(x) \\ &\geq \epsilon^r \int_S dF(x) = \epsilon^r P\{|X - A| \geq \epsilon\}. \end{aligned}$$
▲

It follows from Chebyshev's inequality that any random variable with finite variance is bounded in probability by the square root of its second moment about the origin.

**Corollary 5.1.1.1.** Let  $\{X_n\}$  be a sequence of random variables and  $\{a_n\}$  a sequence of positive real numbers such that

$$E\{X_n^2\} = O(a_n^2).$$

Then

$$X_n = O_p(a_n).$$

**Proof.** By assumption there exists an  $M_1$  such that

$$E\{X_n^2\} < M_1^2 a_n^2$$

for all  $n$ . By Chebyshev's inequality, for any  $M_2 > 0$ ,

$$P\{|X_n| \geq M_2 a_n\} \leq \frac{E\{X_n^2\}}{M_2^2 a_n^2}.$$

Hence, given  $\epsilon > 0$ , we choose  $M_2 \geq M_1 \epsilon^{-1/2}$ , and the result follows.  $\blacktriangleleft$

If the sequence  $\{X_n\}$  has zero mean or a mean whose order is less than or equal to the order of the standard error, then the order in probability of the sequence is the order of the standard error.

**Corollary 5.1.1.2.** Let the sequence of random variables  $\{X_n\}$  satisfy

$$E\{(X_n - E\{X_n\})^2\} = O(a_n^2)$$

and

$$E\{X_n\} = O(a_n),$$

where  $\{a_n\}$  is a sequence of positive real numbers. Then

$$X_n = O_p(a_n).$$

**Proof.** By the assumptions and by property ii of Lemma 5.1.1,

$$E\{X_n^2\} = E\{(X_n - E\{X_n\})^2\} + (E\{X_n\})^2 = O(a_n^2),$$

and the result follows by Corollary 5.1.1.1.  $\blacktriangleleft$

Let the probability limits of two sequences of random variables be defined. We now demonstrate that the sequences have a common probability limit if the probability limit of the sequence of differences is zero.

**Theorem 5.1.2.** Let  $\{X_n\}$  and  $\{Y_n\}$  be sequences of random variables such that

$$p\lim |X_n - Y_n| = 0.$$

If there exists a random variable  $X$  such that  $p\lim X_n = X$ , then

$$p\lim Y_n = X.$$

**Proof.** Given  $\epsilon > 0$  and  $\delta > 0$ , there exists, by hypothesis, an  $N$  such that for  $n > N$ ,

$$P\{|Y_n - X_n| \geq 0.5\epsilon\} \leq 0.5\delta \quad \text{and} \quad P\{|X_n - X| \geq 0.5\epsilon\} \leq 0.5\delta.$$

Applying Lemma 5.1.2, for  $n > N$ ,

$$P\{|Y_n - X| \geq \epsilon\} \leq \delta. \quad \blacktriangle$$

**Definition 5.1.8.** For  $r \geq 1$ , the sequence of random variables  $\{X_n\}$  converges in  $r$ th mean to the random variable  $X$  if  $E\{|X_n|^r\} < \infty$  for all  $n$  and

$$E\{|X_n - X|^r\} \rightarrow 0$$

as  $n \rightarrow \infty$ . We denote convergence in  $r$ th mean by writing  $X_n \xrightarrow{r} X$ .

We note that if  $E\{|X_n - X_m|^r\} \rightarrow 0$  as  $n \rightarrow \infty$  and  $m \rightarrow \infty$ , then there exists a random variable  $X$  such that  $X_n \xrightarrow{r} X$ . Using Chebyshev's inequality it is easy to demonstrate that convergence in  $r$ th mean implies convergence in probability.

**Theorem 5.1.3.** Let  $\{X_n\}$  be a sequence of random variables with finite  $r$ th moments. If there exists a random variable  $X$  such that  $X_n \xrightarrow{r} X$ , then  $X_n \xrightarrow{P} X$ .

**Proof.** Given  $\epsilon > 0$ ,

$$P\{|X_n - X| > \epsilon\} \leq \epsilon^{-r} E\{|X_n - X|^r\}$$

by Chebyshev's inequality. For  $\delta > 0$  there is, by hypothesis, an integer  $N = N(\epsilon, \delta)$  such that for all  $n > N$ ,

$$E\{|X_n - X|^r\} < \delta\epsilon^r$$

and therefore, for  $n > N$ ,

$$P\{|X_n - X| > \epsilon\} < \delta. \quad \blacktriangle$$

One useful consequence is Corollary 5.1.3.1, which can be paraphrased as follows. If the sequence of differences of two sequences of random variables converges in squared mean to zero, then the two sequences of random variables have a common probability limit if the limit exists.

**Corollary 5.1.3.1.** Let  $\{X_n\}$  and  $\{Y_n\}$  be sequences of random variables such that

$$\lim_{n \rightarrow \infty} E\{(X_n - Y_n)^2\} = 0.$$

If there exists a random variable  $X$  such that  $p\lim X_n = X$ , then  $p\lim Y_n = X$ .

**Proof.** By Theorem 5.1.3, we have that  $p\lim(X_n - Y_n) = 0$ . The conclusion follows by Theorem 5.1.2.  $\blacktriangle$

**Corollary 5.1.3.2.** If the sequence of random variables  $\{Y_n\}$  is such that

$$\lim_{n \rightarrow \infty} E\{Y_n\} = \mu$$

and

$$\lim_{n \rightarrow \infty} E\{(Y_n - E\{Y_n\})^2\} = 0,$$

then  $p\lim Y_n = \mu$ .

**Proof.** The proof follows directly by letting the constants  $\mu$  and  $\{E\{Y_n\}\}$  be, respectively, the  $X$  and  $\{X_n\}$  of Corollary 5.1.3.1.  $\blacktriangle$

Since we often work with functions of sequences of random variables, the following theorem is very important. The theorem states that if the function  $g(x)$  is continuous, then "the probability limit of the function is the function of the probability limit."

**Theorem 5.1.4.** Let  $\{X_n\}$  be a sequence of real valued  $k$ -dimensional random variables such that  $p\lim X_n = X$ . Let  $g(x)$  be a function mapping the real  $k$ -dimensional vector  $x$  into a real  $p$ -dimensional space. Let  $g(x)$  be continuous. Then  $p\lim g(X_n) = g(X)$ .

**Proof.** Given  $\epsilon > 0$  and  $\delta > 0$ , let  $A$  be a closed and bounded  $k$ -dimensional set such that

$$P\{X \in A\} \geq 1 - 0.5\delta.$$

Since  $g(x)$  is continuous, it is uniformly continuous on  $A$ , and there exists a  $\delta_\epsilon$  such that

$$|g(x_1) - g(x_2)| < \epsilon$$

if  $|x_1 - x_2| < \delta_\epsilon$  and  $x_1$  is in  $A$ . Since  $p\lim X_n = X$ , there exists an  $N$  such that for  $n > N$ ,

$$P\{|X_n - X| > \delta_\epsilon\} < \frac{1}{2}\delta.$$

Therefore, for  $n > N$ ,

$$\begin{aligned}
 P\{|g(\mathbf{X}_n) - g(\mathbf{X})| > \epsilon\} &= P\{|g(\mathbf{X}_n) - g(\mathbf{X})| > \epsilon | \mathbf{X} \notin A\}P\{\mathbf{X} \notin A\} \\
 &\quad + P\{|g(\mathbf{X}_n) - g(\mathbf{X})| > \epsilon | \mathbf{X} \in A\}P\{\mathbf{X} \in A\} \\
 &\leq P\{\mathbf{X} \notin A\} + P\{|\mathbf{X}_n - \mathbf{X}| > \delta_\epsilon | \mathbf{X} \in A\}P\{\mathbf{X} \in A\} \\
 &\leq P\{\mathbf{X} \notin A\} + P\{|\mathbf{X}_n - \mathbf{X}| > \delta_\epsilon\} \\
 &< \delta.
 \end{aligned}$$
▲

Theorem 5.1.4 can be extended to functions that are continuous except on a set  $D$  where  $P\{\mathbf{X} \in D\} = 0$ . See, for example, Tucker (1967, p. 104).

Mann and Wald (1943b) demonstrated that the algebra of the common order relationships holds for order in probability. The following two theorems are similar to a paraphrase of Mann and Wald's result given by Pratt (1959). The proof follows more closely that of Mann and Wald, however.

**Theorem 5.1.5.** Let  $\{\mathbf{X}_n\}$  be a sequence of  $k$ -dimensional random variables with elements  $\{X_{jn}; j = 1, 2, \dots, k\}$ , and let  $\{\mathbf{r}_n\}$  be a sequence of  $k$ -dimensional vectors with positive real elements  $\{r_{jn}; j = 1, 2, \dots, k\}$  such that

$$\begin{aligned}
 X_{jn} &= O_p(r_{jn}), \quad j = 1, 2, \dots, t, \\
 X_{jn} &= o_p(r_{jn}), \quad j = t+1, t+2, \dots, k.
 \end{aligned}$$

Let  $g_n(\mathbf{x})$  be a sequence of real valued (Borel measurable) functions defined on  $k$ -dimensional Euclidian space, and let  $\{s_n\}$  be a sequence of positive real numbers. Let  $\{\mathbf{a}_n\}$  be a nonrandom sequence of  $k$ -dimensional vectors. If

$$g_n(\mathbf{a}_n) = O(s_n)$$

for all sequences  $\{\mathbf{a}_n\}$  such that

$$\begin{aligned}
 a_{jn} &= O(r_{jn}), \quad j = 1, 2, \dots, t, \\
 a_{jn} &= o(r_{jn}), \quad j = t+1, t+2, \dots, k,
 \end{aligned}$$

then

$$g_n(\mathbf{X}_n) = O_p(s_n).$$

**Proof.** Set  $\epsilon > 0$ . By assumption there exist real numbers  $M_1, M_2, \dots, M_t$  and sequences  $\{M_{jn}\}$ ,  $j = t+1, t+2, \dots, k$ , such that  $\lim_{n \rightarrow \infty} M_{jn} = 0$  and

$$\begin{aligned}
 P\{|X_{jn}| \geq M_j r_{jn}\} &< \frac{\epsilon}{k}, \quad j = 1, 2, \dots, t, \\
 P\{|X_{jn}| \geq M_{jn} r_{jn}\} &< \frac{\epsilon}{k}, \quad j = t+1, t+2, \dots, k,
 \end{aligned}$$

for all  $n$ . Let  $\{A_n\}$  be a sequence of  $k$ -dimensional sets defined by

$$\begin{aligned} A_n = \{(y_1, y_2, \dots, y_k) : -M_j r_{jn} \leq y_j \leq M_j r_{jn} \text{ for } 1 \leq j \leq t \\ \text{and } -M_{jn} r_{jn} \leq y_j \leq M_{jn} r_{jn} \text{ for } t+1 \leq j \leq k\}. \end{aligned}$$

Then, for  $\mathbf{a} \in A_n$ , there exists an  $M$  such that  $|g_n(\mathbf{a})| < Ms_n$ . Hence,

$$|g_n(\mathbf{X}_n)| < Ms_n$$

for all  $\mathbf{X}_n$  contained in  $A_n$ , and the result follows, since the  $A_n$  were constructed so that  $P\{\mathbf{X}_n \in A_n\} > 1 - \epsilon$  for all  $n$ .  $\blacktriangle$

**Theorem 5.1.6.** If we replace

$$g_n(\mathbf{a}_n) = O(s_n)$$

by

$$g_n(\mathbf{a}_n) = o(s_n)$$

in the hypothesis of Theorem 5.1.5, we may replace

$$g_n(\mathbf{X}_n) = O_p(s_n)$$

by

$$g_n(\mathbf{X}_n) = o_p(s_n)$$

in the conclusion.

**Proof.** The set  $A_n$  is constructed exactly as in the proof of Theorem 5.1.5. There then exists a sequence  $\{b_n\}$  such that  $\lim_{n \rightarrow \infty} b_n = 0$  and

$$|g_n(\mathbf{a})| < b_n s_n$$

for  $\mathbf{a}$  contained in  $A_n$ . Therefore,  $|g_n(\mathbf{X}_n)| < b_n s_n$  for all  $\mathbf{X}_n$  contained in  $A_n$ , and the result follows from the construction of the  $A_n$ .  $\blacktriangle$

**Corollary 5.1.5.** Let  $\{X_n\}$  be a sequence of scalar random variables such that

$$X_n = a + O_p(r_n),$$

where  $r_n \rightarrow 0$  as  $n \rightarrow \infty$ . If  $g(x)$  is a function with  $s$  continuous derivatives at  $x = a$ ,

then

$$\begin{aligned} g(X_n) &= g(a) + g^{(1)}(a)(X_n - a) \\ &\quad + \cdots + \frac{1}{(s-1)!} g^{(s-1)}(a)(X_n - a)^{s-1} + O_p(r_n^s), \end{aligned}$$

where  $g^{(j)}(a)$  is the  $j$ th derivative of  $g(x)$  evaluated at  $x = a$ .

**Proof.** Since the statement holds for a sequence of real numbers, the result follows from Theorem 5.1.5. A direct proof can be obtained by expanding  $g(x)$  in a Taylor series with remainder

$$\frac{g^{(s)}(b)(X_n - a)^s}{s!},$$

where  $b$  is between  $X_n$  and  $a$ . Since  $g^{(s)}(x)$  is continuous at  $a$ , for  $n$  sufficiently large,  $g^{(s)}(b)$  is bounded in probability [i.e.,  $g^{(s)}(b) = O_p(1)$ ]. Therefore,

$$\frac{g^{(s)}(b)(X_n - a)^s}{s!} = O_p(r_n^s). \quad \blacktriangle$$

**Corollary 5.1.6.** If

$$X_n = a + O_p(r_n)$$

in the hypothesis of Corollary 5.1.5 is replaced by

$$X_n = a + o_p(r_n),$$

then the remainder  $O_p(r_n^s)$  is replaced by  $o_p(r_n^s)$ .

**Proof.** The proof is nearly identical to that of Corollary 5.1.5.  $\blacktriangle$

Note that the condition on the derivative defining the remainder can be weakened. As we saw in the proof of Corollary 5.1.5, we need only that  $g^{(s)}(b)$  is bounded in probability. Also see Exercise 5.30.

The corollaries generalize immediately to vector random variables. For example, let

$$\mathbf{X}_n = \mathbf{a} + o_p(r_n),$$

where  $\mathbf{X}_n = (X_{1n}, X_{2n}, \dots, X_{kn})'$ ,  $\mathbf{a} = (a_1, a_2, \dots, a_k)'$ , and  $r_n \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $g(\mathbf{x})$  be a real valued function defined on  $k$ -dimensional Euclidean space with

continuous partial derivatives of order three at  $\mathbf{a}$ ; then, for example,

$$\begin{aligned} g(\mathbf{X}_n) &= g(\mathbf{a}) + \sum_{j=1}^k \frac{\partial g(\mathbf{a})}{\partial x_j} (X_{jn} - a_j) \\ &\quad + \sum_{j=1}^k \sum_{i=1}^k \frac{1}{2!} \frac{\partial^2 g(\mathbf{a})}{\partial x_j \partial x_i} (X_{jn} - a_j)(X_{in} - a_i) + o_p(r_n^3), \end{aligned}$$

where  $\partial g(\mathbf{a})/\partial x_j$  is the partial derivative of  $g(\mathbf{x})$  with respect to  $x_j$  evaluated at  $\mathbf{x} = \mathbf{a}$  and  $\partial^2 g(\mathbf{a})/\partial x_j \partial x_i$  is the second partial derivative of  $g(\mathbf{x})$  with respect to  $x_j$  and  $x_i$  evaluated at  $\mathbf{x} = \mathbf{a}$ .

The Taylor expansions of Corollaries 5.1.5 and 5.1.6 are about a fixed point  $\mathbf{a}$ . Expansions about a random point are also valid under certain conditions.

**Theorem 5.1.7.** Let  $\{\mathbf{X}_n\}$ ,  $\{\mathbf{W}_n\}$  be two sequences of  $k$ -dimensional vector random variables defined on the same probability space. Let  $l(\mathbf{X}_n, \mathbf{W}_n)$  be the line segment joining  $\mathbf{X}_n$  and  $\mathbf{W}_n$ . For every  $\epsilon > 0$ , suppose there is an  $N_\epsilon$  and a set  $B_\epsilon$  in  $k$ -dimensional Euclidean space, such that

$$P\{l(\mathbf{X}_n, \mathbf{W}_n) \in B_\epsilon\} \geq 1 - \epsilon$$

for all  $n > N_\epsilon$ .

Let  $g(\mathbf{x})$  be a real valued function that is continuous with continuous partial derivatives through order  $s$  on  $B_\epsilon$  for every  $\epsilon > 0$ . Suppose the absolute values of the  $s$ th order partial derivatives are bounded on  $B_\epsilon$  and that

$$\mathbf{X}_n - \mathbf{W}_n = O_p(r_n),$$

where  $r_n \rightarrow 0$  as  $n \rightarrow \infty$ . Then

$$\begin{aligned} g(\mathbf{X}_n) &= g(\mathbf{W}_n) + \sum_{i=1}^k g^{(i)}(\mathbf{W}_n)(X_{in} - W_{in}) \\ &\quad + 2^{-1} \sum_{i=1}^k \sum_{j=1}^k g^{(ij)}(\mathbf{W}_n)(X_{in} - W_{in})(X_{jn} - W_{jn}) + \dots \\ &\quad + [(s-1)!]^{-1} \sum_{i=1}^k \dots \sum_{t=1}^k g^{(i\dots t)}(\mathbf{W}_n)(X_{in} - W_{in}) \dots (X_{tn} - W_{tn}) \\ &\quad + O_p(r_n^s), \end{aligned}$$

where  $g^{(i)}(\mathbf{W}_n)$  is the first partial derivative of  $g(\mathbf{x})$  with respect to  $x_i$  evaluated at  $\mathbf{x} = \mathbf{W}_n$ , and  $g^{(i\dots t)}(\mathbf{W}_n)$  is the  $s$ th partial derivative of  $g(\mathbf{x})$  with respect to the elements of  $\mathbf{x}$  whose indexes are  $i, \dots, t$ , evaluated at  $\mathbf{x} = \mathbf{W}_n$ .

**Proof.** Let  $\epsilon > 0$  be given, and let  $B_\epsilon$  be the associated set. Then  $g(\mathbf{x})$  and the first  $s$  derivatives of  $g(\mathbf{x})$  are continuous on  $B_\epsilon$ . For  $l(\mathbf{X}_n, \mathbf{W}_n) \in B_\epsilon$ , one can expand the function in a Taylor series with exact remainder. For scalar  $X_n$ , the expansion

is

$$\begin{aligned} g(X_n) &= g(W_n) + g^{(1)}(W_n)(X_n - W_n) + \dots \\ &\quad + [(s-1)!]^{-1} g^{(s-1)}(W_n)(X_n - W_n)^{s-1} + (s!)^{-1} g^{(s)}(W_n^{\dagger})(X_n - W_n)^s, \end{aligned}$$

where  $W_n^{\dagger}$  is on the line segment joining  $X_n$  and  $W_n$ , and for scalar arguments,  $g^{(j)}(W_n)$  denotes the  $j$ th derivative. Now  $g^{(s)}(W_n^{\dagger})$  is bounded on  $B_{\epsilon}$ , and  $X_n - W_n = O_p(r_n)$ . It follows that

$$g^{(s)}(W_n^{\dagger})(X_n - W_n)^s = O_p(r_n^s).$$

The arguments are analogous for the vector case, and the result is established.  $\blacktriangle$

**Example 5.1.1.** Let  $W_n = W$ , where  $W \sim N(0, 1)$ , and let

$$X_n = W + Z_n,$$

where  $Z_n \sim U(-n^{-1/2}, n^{-1/2})$ , independent of  $W$ . Let  $g(x) = x^{-2}$ . Let  $1 > \epsilon > 0$  be given, and set

$$B_{\epsilon} = \{x: |x| > 2^{-1}\epsilon\},$$

$$B_{2\epsilon} = \{x: |x| > \epsilon\}.$$

The ordinate of the standard normal distribution is about 0.40 at zero. Therefore,  $P\{W \in B_{\epsilon}\} \geq 1 - 0.4\epsilon$  and  $P\{W \in B_{2\epsilon}\} \geq 1 - 0.8\epsilon$ . If  $W \in B_{2\epsilon}$  and  $n > 4\epsilon^{-2}$ , then  $X_n \in B_{\epsilon}$ . Hence, if  $n > 4\epsilon^{-2}$  the probability is greater than  $1 - \epsilon$  that  $l(X_n, W_n)$  is in  $B_{\epsilon}$ . The function  $x^{-2}$  is continuous with continuous derivatives on  $B_{\epsilon}$ . The absolute value of the first derivative and of the second derivative are bounded by finite multiples of  $\epsilon^{-3}$  and  $\epsilon^{-4}$ , respectively, for  $x \in B_{\epsilon}$ . Therefore, the conditions of Theorem 5.1.7 are met and

$$\begin{aligned} X_n^{-2} &= W^{-2} - 2W^{-3}(X_n - W) + O_p(n^{-1}) \\ &= W^{-2} + O_p(n^{-1/2}). \end{aligned}$$

Because  $W^2$  is a chi-square random variable,  $g(X_n)$  converges in probability to the reciprocal of a chi-square random variable, by Theorem 5.1.4. Theorem 5.1.7 enables us to establish the order of the remainder in the approximation.  $\blacktriangle\blacktriangle$

## 5.2. CONVERGENCE IN DISTRIBUTION

In the preceding section we discussed conditions under which a sequence of random variables converges in probability to a limit random variable. A second type of convergence important in statistics is the convergence of a sequence of distribution functions to a limit function. The classical example of such conver-

gence is given by the central limit theorem wherein the sequence of distribution functions converges pointwise to the normal distribution function.

**Definition 5.2.1.** If  $\{X_n\}$  is a sequence of random variables with distribution functions  $\{F_{X_n}(x)\}$ , then  $\{X_n\}$  is said to converge in distribution (or in law) to the random variable  $X$  with distribution function  $F_X(x)$ , and we write  $X_n \xrightarrow{d} X$ , if

$$\lim_{n \rightarrow \infty} F_{X_n}(x) = F_X(x)$$

at all  $x$  for which  $F_X(x)$  is continuous.

Note that the sequence of distribution functions is converging to a function that is itself a distribution function. Some authors define this type of convergence by saying the sequence  $\{F_{X_n}(x)\}$  converges *completely* to  $F_X(x)$ . Thus our symbolism

$$X_n \xrightarrow{d} X$$

is understood to mean that  $F_{X_n}(x)$  converges to the distribution function of the random variable  $X$ . The notation

$$F_{X_n}(x) \xrightarrow{C} F_X(x)$$

is also used.

**Theorem 5.2.1.** Let  $\{X_n\}$  and  $\{Y_n\}$  be sequences of random variables such that

$$p\lim(X_n - Y_n) = 0.$$

If there exists a random variable  $X$  such that

$$X_n \xrightarrow{d} X,$$

then

$$Y_n \xrightarrow{d} X.$$

**Proof.** Let  $W$  and  $Z$  be random variables with distribution functions  $F_W(w)$  and  $F_Z(z)$ , respectively, and fix  $\epsilon > 0$  and  $\delta > 0$ . We first show that

$$P\{|Z - W| > \epsilon\} \leq \delta$$

implies that

$$F_Z(z - \epsilon) - \delta \leq F_W(z) \leq F_Z(z + \epsilon) + \delta$$

for all  $z$ . This result holds because

$$\begin{aligned} F_Z(z - \epsilon) - F_W(z) &= P\{Z \leq z - \epsilon\} - P\{W \leq z\} \\ &\leq P\{Z \leq z - \epsilon\} - P\{Z \leq z - \epsilon \text{ and } W \leq z\} \\ &= P\{Z \leq z - \epsilon \text{ and } W > z\} \\ &\leq P\{(W - Z) > \epsilon\} \\ &\leq P\{|W - Z| > \epsilon\} \leq \delta, \end{aligned}$$

and, in a similar manner,

$$\begin{aligned} F_W(z) - F_Z(z + \epsilon) &= P\{W \leq z\} - P\{Z \leq z + \epsilon\} \\ &\leq P\{W \leq z\} - P\{Z \leq z + \epsilon \text{ and } W \leq z\} \\ &= P\{W \leq z \text{ and } Z > z + \epsilon\} \\ &\leq P\{|Z - W| > \epsilon\} \leq \delta. \end{aligned}$$

Let  $x_0$  be a continuity point for  $F_X(x)$ . Then, given  $\delta > 0$ , there is an  $\eta > 0$  such that  $|F_X(x) - F_X(x_0)| < \frac{1}{4}\delta$  for  $|x - x_0| \leq \eta$  and  $F_X(x)$  is continuous at  $x_0 - \eta$  and at  $x_0 + \eta$ . Furthermore, for this  $\delta$  and  $\eta$ , there is an  $N_1$  such that, for  $n > N_1$ ,

$$P\{|X_n - Y_n| > \eta\} < \frac{1}{2}\delta$$

and therefore

$$F_{X_n}(x - \eta) - \frac{1}{2}\delta \leq F_{Y_n}(x) \leq F_{X_n}(x + \eta) + \frac{1}{2}\delta$$

for all  $x$ . Also, there is an  $N_2$  such that, for  $n > N_2$ ,

$$|F_{X_n}(x_0 - \eta) - F_X(x_0 - \eta)| < \frac{1}{4}\delta$$

and

$$|F_{X_n}(x_0 + \eta) - F_X(x_0 + \eta)| < \frac{1}{4}\delta.$$

Therefore, given the continuity point  $x_0$  and  $\delta > 0$ , there is an  $\eta > 0$  and an  $N = \max(N_1, N_2)$  such that for  $n > N$ ,

$$\begin{aligned} F_X(x_0) - \delta &< F_X(x_0 - \eta) - \frac{3}{4}\delta < F_{X_n}(x_0 - \eta) - \frac{1}{2}\delta \\ &\leq F_{Y_n}(x_0) \leq F_{X_n}(x_0 + \eta) + \frac{1}{2}\delta < F_X(x_0 + \eta) + \frac{3}{4}\delta \\ &< F_X(x_0) + \delta. \end{aligned}$$
▲

As a corollary we have the result that convergence in probability implies convergence in law.

**Corollary 5.2.1.1.** Let  $\{X_n\}$  be a sequence of random variables. If there exists a random variable  $X$  such that  $p\lim X_n = X$ , then  $X_n \xrightarrow{P} X$ .

**Corollary 5.2.1.2.** Let  $\{X_n\}$  and  $X$  be random variables such that  $p\lim X_n = X$ . If  $g(x)$  is a continuous function, then the distribution of  $g(X_n)$  converges to the distribution of  $g(X)$ .

**Proof.** This follows immediately, since by Theorem 5.1.4,

$$p\lim g(X_n) = g(X). \quad \blacktriangle$$

We state the following two important theorems without proof.

**Theorem 5.2.2 (Helly–Bray).** If  $\{F_n(x)\}$  is a sequence of distribution functions over  $k$ -dimensional Euclidean space  $\mathcal{R}^{(k)}$  such that  $F_n(x) \xrightarrow{C} F(x)$ , then

$$\int g(x) dF_n(x) \rightarrow \int g(x) dF(x) \quad \text{as } n \rightarrow \infty$$

for every bounded continuous function  $g(x)$ .

**Theorem 5.2.3.** Let  $\{F_n(x)\}$  be a sequence of distribution functions over  $\mathcal{R}^{(k)}$  with corresponding characteristic functions  $\{\varphi_n(u)\}$ .

- (i) If  $F_n(x) \xrightarrow{C} F(x)$ , then  $\varphi_n(u) \rightarrow \varphi(u)$  at all  $u \in \mathcal{R}^{(k)}$ , where  $\varphi(u)$  is the characteristic function associated with  $F(x)$ .
- (ii) Continuity theorem. If  $\varphi_n(u)$  converges pointwise to a function  $\varphi(u)$  that is continuous at  $(0, 0, \dots, 0) \in \mathcal{R}^{(k)}$ , then  $\varphi(u)$  is the characteristic function of a distribution function  $F(x)$  and  $F_n(x) \xrightarrow{C} F(x)$ .

**Theorem 5.2.4.** Let  $\{X_n\}$  be a sequence of  $k$ -dimensional random variables with distribution functions  $\{F_{X_n}(x)\}$  such that  $F_{X_n}(x) \xrightarrow{C} F_X(x)$ , and let  $T$  be a continuous mapping from  $\mathcal{R}^{(k)}$  to  $\mathcal{R}^{(p)}$ . Then

$$F_{T(X_n)}(y) \xrightarrow{C} F_{T(X)}(y).$$

**Proof.** By the Helly–Bray theorem, the characteristic function of  $T(X_n)$  converges to the characteristic function of  $T(X)$ , and the result follows.  $\blacktriangle$

**Theorem 5.2.5.** Let  $\{X_n\}$  and  $\{Y_n\}$  be two sequences of  $k$ -dimensional random variables such that  $X_n$  is independent of  $Y_n$  for all  $n$ . If there exist random variables  $X$  and  $Y$  such that  $F_{X_n}(x) \xrightarrow{C} F_X(x)$  and  $F_{Y_n}(y) \xrightarrow{C} F_Y(y)$ , then

$$F_{X_n Y_n}(x, y) \xrightarrow{C} F_X(x)F_Y(y).$$

**Proof.** The characteristic function of  $(X'_n, Y'_n)'$  is given by

$$\begin{aligned} \int \int e^{\epsilon u' x} e^{\epsilon v' y} dF_{X'_n Y'_n}(x, y) &= \int e^{\epsilon u' x} dF_{X'_n}(x) \int e^{\epsilon v' y} dF_{Y'_n}(y) \\ &= \varphi_{X'_n}(u) \varphi_{Y'_n}(v). \end{aligned}$$

Now, by the Helly–Bray theorem,  $\varphi_{X'_n}(u) \rightarrow \varphi_X(u)$  and  $\varphi_{Y'_n}(v) \rightarrow \varphi_Y(v)$ . Therefore,

$$\varphi_{XY}(u, v) = \lim_{n \rightarrow \infty} \varphi_{X'_n}(u) \varphi_{Y'_n}(v) = \varphi_X(u) \varphi_Y(v).$$

By the continuity theorem, this implies that  $F_{X'_n Y'_n}(x, y) \xrightarrow{C} F_{XY}(x, y)$ , where  $F_{XY}(x, y)$  is the distribution function of independent random variables associated with the characteristic function  $\varphi_{XY}(u, v)$ .  $\blacktriangle$

From Corollary 5.2.1.1, we know that convergence in probability implies convergence in law. For the special case wherein a sequence of random variables converges in law to a constant random variable, the converse is also true.

**Lemma 5.2.1.** Let  $\{Y_n\}$  be a sequence of  $p$ -dimensional random variables with corresponding distribution functions  $\{F_{Y_n}(y)\}$ . Let  $Y$  be a  $p$ -dimensional random variable with distribution function  $F_Y(y)$  such that  $P\{Y = b\} = 1$ ,  $b$  is a constant vector, and

$$F_{Y_n}(y) \xrightarrow{C} F_Y(y).$$

Then, given  $\epsilon > 0$ , there exists an  $N$  such that, for  $n > N$ ,

$$P\{|Y_n - b| \geq \epsilon\} < \epsilon.$$

**Proof.** Let  $B = \{y: y_1 > b_1 - \epsilon/p, y_2 > b_2 - \epsilon/p, \dots, y_p > b_p - \epsilon/p\}$ . Then  $F_Y(y) = 0$  on the complement of  $B$ . Fix  $\epsilon > 0$ . As  $F_{Y_n}(y) \rightarrow F_Y(y)$ , there exists an  $N_0$  such that, for  $n > N_0$ ,

$$F_{Y_n}(g_1) = F_{Y_n}\left(b_1 - \frac{\epsilon}{p}, b_2 + \frac{\epsilon}{p}, \dots, b_p + \frac{\epsilon}{p}\right) < \frac{\epsilon}{2p},$$

$$F_{Y_n}(g_2) = F_{Y_n}\left(b_1 + \frac{\epsilon}{p}, b_2 - \frac{\epsilon}{p}, \dots, b_p + \frac{\epsilon}{p}\right) < \frac{\epsilon}{2p},$$

⋮

$$F_{Y_n}(g_p) = F_{Y_n}\left(b_1 + \frac{\epsilon}{p}, b_2 + \frac{\epsilon}{p}, \dots, b_p - \frac{\epsilon}{p}\right) < \frac{\epsilon}{2p}.$$

There also exists an  $N_1$  such that, for  $n > N_1$ ,  $1 - F_{Y_n}(b_1 + \epsilon/p, b_2 + \epsilon/p, \dots, b_p +$

$\epsilon/p < \epsilon/2$ . Therefore, for  $n > \max(N_0, N_1)$ ,

$$\begin{aligned} P\left\{ b_1 - \frac{\epsilon}{p} \leq y_1 \leq b_1 + \frac{\epsilon}{p}, b_2 - \frac{\epsilon}{p} \leq y_2 \leq b_2 + \frac{\epsilon}{p}, \dots, \right. \\ \left. b_p - \frac{\epsilon}{p} \leq y_p \leq b_p + \frac{\epsilon}{p} \right\} \\ \geq F_{Y_n}\left(b_1 + \frac{\epsilon}{p}, b_2 + \frac{\epsilon}{p}, \dots, b_p + \frac{\epsilon}{p}\right) - \sum_{i=1}^p F_{Y_n}(b_i) \geq 1 - \epsilon, \end{aligned}$$

and it follows that

$$P\{|Y_n - b| \geq \epsilon\} < \epsilon. \quad \blacktriangle$$

**Theorem 5.2.6.** Let  $\{(X'_n, Y'_n)\}'$  be a sequence of  $(k+p)$ -dimensional random variables where  $X'_n$  is  $k$ -dimensional. Let the sequence of joint distribution functions be denoted by  $\{F_{X'_n Y'_n}(x, y)\}$  and the sequences of marginal distribution functions by  $\{F_{X'_n}(x)\}$  and  $\{F_{Y'_n}(y)\}$ . If there exists a  $k$ -dimensional random variable  $X$  and a  $p$ -dimensional random variable  $Y$  such that  $F_{X'_n}(x) \xrightarrow{c} F_X(x)$  and  $F_{Y'_n}(y) \xrightarrow{c} F_Y(y)$ , where  $P\{Y = b\} = 1$  and  $b = (b_1, b_2, \dots, b_p)'$  is a constant vector, then

$$F_{X'_n Y'_n}(x, y) \xrightarrow{c} F_{XY}(x, y).$$

**Proof.** Now  $P\{Y = b\} = 1$  implies that

$$F_X(x) = F_{XY}(x, b),$$

that  $F_{XY}(x, y) = 0$  if any element of  $y$  is less than the corresponding element of  $b$ , and that  $F_{XY}(x, y) = F_X(x)$  if every element of  $y$  is greater than or equal to the corresponding element of  $b$ . Fix  $\epsilon > 0$ , and consider a point  $(x_0, y_0)$  where at least one element of  $y_0$  is less than the corresponding element of  $b$  by an amount  $\epsilon$ . Then  $F_{XY}(x_0, y_0) = 0$ . However, there is an  $N_0$  such that, for  $n > N_0$ ,

$$F_{X'_n Y'_n}(x_0, y_0) < \frac{\epsilon}{2},$$

by Lemma 5.2.1. Let  $(x_0, y_1)$  be a continuity point of  $F_{XY}(x, y)$  where every element of  $y_1$  exceeds the corresponding element of  $b$  by  $\epsilon/p^{1/2} > 0$ . Because  $F_{X'_n}(x) \xrightarrow{c} F_X(x)$  and  $F_{Y'_n}(y) \xrightarrow{c} F_Y(y)$ , we can choose  $N_1$  such that, for  $n \geq N_1$ ,  $|F_{X'_n}(x_0) - F_X(x_0)| < \epsilon/2$  and  $|F_{Y'_n}(y_1) - F_Y(y_1)| < \epsilon/2$ . Hence,

$$\begin{aligned} |F_{X'_n Y'_n}(x_0, y_1) - F_{XY}(x_0, y_1)| &\leq |F_{X'_n Y'_n}(x_0, y_1) - F_{X'_n}(x_0)| \\ &\quad + |F_{X'_n}(x_0) - F_X(x_0)| \\ &< 1 - F_{Y'_n}(y_1) + \frac{\epsilon}{2} < \epsilon. \end{aligned} \quad \blacktriangle$$

Utilizing Theorems 5.2.4 and 5.2.6, we obtain the following corollary.

**Corollary 5.2.6.1.** Let  $\{X_n\}$  and  $\{Y_n\}$  be two sequences of  $k$ -dimensional random variables. If there exists a  $k$ -dimensional random variable  $Y$  and a fixed vector  $b$  such that  $Y_n \xrightarrow{\mathcal{L}} Y$  and  $X_n \xrightarrow{\mathcal{L}} b$ , then

- (i)  $X_n + Y_n \xrightarrow{\mathcal{L}} b + Y$ ,
- (ii)  $X'_n Y_n \xrightarrow{\mathcal{L}} b' Y$ .

**Corollary 5.2.6.2.** Let  $\{Y_n\}$  be a sequence of  $k$ -dimensional random variables, and let  $\{A_n\}$  be a sequence of  $k \times k$  random matrices. If there exists a random vector  $Y$  and a fixed nonsingular matrix  $A$  such that  $Y_n \xrightarrow{\mathcal{L}} Y$ ,  $A_n \xrightarrow{\mathcal{L}} A$ , then

$$A_n^{-1} Y_n \xrightarrow{\mathcal{L}} A^{-1} Y.$$

### 5.3. CENTRAL LIMIT THEOREMS

The exact distributions of many statistics encountered in practice have not been obtained. Fortunately, many statistics in the class of continuous functions of means or of sample moments converge in distribution to normal random variables. We give without proof the following central limit theorem.

**Theorem 5.3.1 (Lindeberg central limit theorem).** Let  $\{Z_t: t = 1, 2, \dots\}$  be a sequence of independent random variables with distribution functions  $\{F_t(z)\}$ . Let  $E\{Z_t\} = \mu_t$ ,  $E\{(Z_t - \mu_t)^2\} = \sigma_t^2$ , and assume

$$\lim_{n \rightarrow \infty} V_n^{-1} \sum_{t=1}^n \int_{|z - \mu_t| > \epsilon V_n^{1/2}} (z - \mu_t)^2 dF_t(z) = 0 \quad (5.3.1)$$

for all  $\epsilon > 0$ , where  $V_n = \sum_{t=1}^n \sigma_t^2$ . Then,

$$V_n^{-1/2} \sum_{t=1}^n (Z_t - \mu_t) \xrightarrow{\mathcal{L}} N(0, 1),$$

where  $N(0, 1)$  denotes the normal distribution with mean zero and variance one.

A form of the central limit theorem whose conditions are often more easily verified is the following theorem.

**Theorem 5.3.2 (Liapounov central limit theorem).** Let  $\{Z_t: t = 1, 2, \dots\}$  be a sequence of independent random variables with distribution functions  $\{F_t(z)\}$ . Let

$E\{Z_i\} = \mu_i$ ,  $E\{(Z_i - \mu_i)^2\} = \sigma_i^2$ , and  $V_n = \sum_{i=1}^n \sigma_i^2$ . If

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \int |z - \mu_i|^{2+\delta} dF_i(z)}{V_n^{1+\delta/2}} = 0$$

for some  $\delta > 0$ , then

$$V_n^{-1/2} \sum_{i=1}^n (Z_i - \mu_i) \xrightarrow{\mathcal{L}} N(0, 1).$$

**Proof.** Let  $\delta > 0$  and  $\epsilon > 0$ , and define the set  $A_\epsilon$  by

$$A_\epsilon = \{z: |z - \mu_i| > \epsilon V_n^{1/2}\}.$$

Then

$$\begin{aligned} \frac{1}{V_n} \sum_{i=1}^n \int_{A_\epsilon} (z - \mu_i)^2 dF_i(z) &\leq \frac{1}{V_n (\epsilon V_n^{1/2})^\delta} \sum_{i=1}^n \int_{A_\epsilon} |z - \mu_i|^{2+\delta} dF_i(z) \\ &\leq \frac{1}{V_n^{1+\delta/2} \epsilon^\delta} \sum_{i=1}^n \int |z - \mu_i|^{2+\delta} dF_i(z), \end{aligned}$$

which, by hypothesis, goes to zero as  $n \rightarrow \infty$ . Therefore, the condition on the  $2 + \delta$  moment implies the condition of the Lindeberg theorem.  $\blacktriangle$

The reader is referred to the texts of Tucker (1967), Gnedenko (1967), and Loève (1963) for discussions of these theorems.

For a proof of the following extension of the central limit theorems to the multivariate case, see Varadarajan (1958).

**Theorem 5.3.3.** Let  $\{Z_n: n = 1, 2, \dots\}$  be a sequence of  $k$ -dimensional random variables with distribution functions  $\{F_{Z_n}(z)\}$ . Let  $F_{\lambda, n}(x)$  be the distribution function of  $X_n = \lambda' Z_n$ , where  $\lambda$  is a fixed vector. A necessary and sufficient condition for  $F_{Z_n}(z)$  to converge to the  $k$ -variate distribution function  $F(z)$  is that  $F_{\lambda, n}(x)$  converges to a limit for each  $\lambda$ .

In most of our applications of Theorem 5.3.3, each  $F_{\lambda, n}(x)$  will be converging to a normal distribution function and hence the vector random variable  $Z_n$  will converge in distribution to a multivariate normal.

The Lindeberg and Liapounov central limit theorems are for independent random variables. It is possible to obtain the limiting normal distribution for sequences that satisfy weaker conditions. One type of sequence that has been studied is the *martingale* process.

**Definition 5.3.1.** Let  $\{X_i\}_{i=1}^\infty$  be a sequence of random variables defined on the

space  $\{\Omega, \mathcal{A}, P\}$ . The sequence is a martingale if, for all  $t$ ,

$$E\{|X_t|\} < \infty$$

and

$$E\{X_t | \mathcal{A}_{t-1}\} = X_{t-1} \quad \text{a.s.},$$

where  $\mathcal{A}_t$  is the sigma-field generated by  $\{X_1, X_2, \dots, X_t\}$ .

It is an immediate consequence of the definition that for  $s \leq t-1$ ,

$$E\{X_t | \mathcal{A}_s\} = X_s \quad \text{a.s.}$$

**Definition 5.3.2.** Let  $\{X_t\}_{t=1}^\infty$  be a martingale sequence defined on  $\{\Omega, \mathcal{A}, P\}$ . Let  $Z_t = X_t - X_{t-1}$ . Then the sequence  $\{Z_t\}_{t=1}^\infty$  is called a sequence of *martingale differences*.

We give a central limit theorem for martingale differences. Theorem 5.3.4 is due to Brown (1971). Related results have been obtained by Dvoretzky (1972), Scott (1973), and McLeish (1974). Also see Hall and Heyde (1980) and Pollard (1984).

**Theorem 5.3.4.** Let  $\{Z_{tn}: 1 \leq t \leq n, n \geq 1\}$  denote a triangular array of random variables defined on the probability space  $(\Omega, \mathcal{A}, P)$ , and let  $\{\mathcal{A}_{tn}: 0 \leq t \leq n, n \geq 1\}$  be any triangular array of sub-sigma-fields of  $\mathcal{A}$  such that for each  $n$  and  $1 \leq t \leq n$ ,  $Z_{tn}$  is  $\mathcal{A}_{tn}$ -measurable and  $\mathcal{A}_{t-1,n}$  is contained in  $\mathcal{A}_{tn}$ . For  $1 \leq k \leq n$ ,  $1 \leq j \leq n$ , and  $n \geq 1$ , let

$$\begin{aligned} S_{kn} &= \sum_{t=1}^k Z_{tn}, \\ \delta_{tn}^2 &= E\{Z_{tn}^2 | \mathcal{A}_{t-1,n}\}, \\ V_{jn}^2 &= \sum_{t=1}^j \delta_{tn}^2, \end{aligned}$$

and

$$s_{nn}^2 = E\{V_{nn}^2\}.$$

Assume

- (i)  $E(Z_{tn} | \mathcal{A}_{t-1,n}) = 0 \quad \text{a.s. for } 1 \leq t \leq n,$
- (ii)  $V_{nn}^2 s_{nn}^{-2} \xrightarrow{P} 1,$
- (iii)  $\lim_{n \rightarrow \infty} s_{nn}^{-2} \sum_{j=1}^n E[Z_{jn}^2 I(|Z_{jn}| \geq \epsilon s_{nn})] | \mathcal{A}_{t-1,n} = 0 \text{ for all } \epsilon > 0,$

where  $I(A)$  denotes the indicator function of a set  $A$ . Then, as  $n \rightarrow \infty$ ,

$$s_{nn}^{-1} S_{nn} \xrightarrow{\mathcal{L}} N(0, 1).$$

**Proof.** Omitted. ▲

**Corollary 5.3.4.** Let  $\{e_{tn}\}$  be a triangular array satisfying

$$\begin{aligned} E\{(e_{tn}, e_{tn}^2) | \mathcal{A}_{t-1,n}\} &= (0, \sigma^2) \quad \text{a.s.,} \\ E\{|e_{tn}|^{2+\delta} | \mathcal{A}_{t-1,n}\} &< M < \infty \quad \text{a.s.} \end{aligned}$$

for some  $\delta > 0$ , where  $\mathcal{A}_{t,n}$  is the sigma-field generated by  $\{e_{jn}; j \leq t\}$ . Let  $\{w_{tn}; 1 \leq t \leq n, n \geq 1\}$  be a triangular array of constants,  $\sum_{i=1}^n w_{in}^2 \neq 0$  for all  $n$ , satisfying

$$\lim_{n \rightarrow \infty} \sup_{1 \leq j \leq n} \left( \sum_{i=1}^n w_{in}^2 \right)^{-1} w_{jn}^2 = 0.$$

Then

$$\left( \sum_{i=1}^n w_{in}^2 \right)^{-1/2} \sum_{i=1}^n w_{in} e_{in} \xrightarrow{\mathcal{L}} N(0, \sigma^2).$$

**Proof.** Reserved for the reader. ▲

We now give a functional central limit theorem which is an extension of Theorem 5.3.4. For related results, we refer the reader to Billingsley (1968). In the following, we let  $D = D[0, 1]$  denote the space of functions  $f(u)$  on  $[0, 1]$  that are right continuous and have left limits. Let  $(\Omega, \mathcal{A}, P)$  be a probability space, and let  $X(u, \omega)$  be a random function on  $D$  defined by an element  $\omega$  of  $\Omega$ . We will often abbreviate  $X(u, \omega)$  to  $X(u)$  or to  $X$ .

An important random function is the *Wiener process*. If  $W$  is a standard Wiener process, then  $W(u)$  is normally distributed for every  $u$  in  $[0, 1]$ , where  $W(u) \sim N(0, u)$ . Furthermore,  $W(u)$  has normal independent increments. A process on  $[0, 1]$  has independent increments if, for

$$0 \leq u_1 \leq u_2 \leq \cdots \leq u_k \leq 1,$$

the random variables

$$W(u_2) - W(u_1), \quad W(u_3) - W(u_2), \dots, \quad W(u_k) - W(u_{k-1})$$

are independent. The weak convergence or convergence in distribution of a sequence of random elements  $X_n$  in  $D$  to a random element  $X$  in  $D$  will be denoted by  $X_n \Rightarrow X$  or by  $X_n \xrightarrow{\mathcal{D}} X$ .

**Theorem 5.3.5.** Let  $\{e_i\}_{i=1}^\infty$  be a sequence of random variables, and let  $\mathcal{A}_{t-1}$

be the sigma-field generated by  $\{e_1, e_2, \dots, e_{t-1}\}$ . Assume that

$$E\{(e_t, e_t^2) | \mathcal{A}_{t-1}\} = (0, \sigma^2) \quad \text{a.s.}$$

and

$$E[|e_t|^{2+\delta} | \mathcal{A}_{t-1}] < M < \infty \quad \text{a.s.}$$

for some  $\delta > 0$ . Let

$$S_n(u) = \sigma^{-1} n^{-1/2} \sum_{j=1}^{\lfloor nu \rfloor} e_j, \quad 0 \leq u \leq 1,$$

where  $\lfloor nu \rfloor$  denotes the integer part of  $nu$  and the sum is zero if  $\lfloor nu \rfloor = 0$ . Then

$$S_n \Rightarrow W,$$

where  $W$  is a standard Wiener process.

**Proof.** See Billingsley (1968) or Chan and Wei (1988).  $\blacktriangle$

The conclusion of Theorem 5.3.5 holds for  $e_t$  that are  $\text{iid}(0, \sigma^2)$  random variables. That form of the theorem is known as Donsker's theorem. See Billingsley (1968, p. 68).

We now give a result for the limiting distribution of continuous functionals of  $S_n$ . The result follows from the continuous mapping theorem. See Theorem 5.1 of Billingsley (1968).

**Theorem 5.3.6.** Let  $S_n$  and  $S$  be random elements in  $D$ . For  $0 \leq s \leq 1$ , let

$$Z_{in}(s) = \int_0^s u^{k_i} f_i[S_n(u)] du$$

and

$$Z_i(s) = \int_0^s u^{k_i} f_i[S(u)] du, \quad i = 1, \dots, m,$$

where  $f_1, \dots, f_m$  are real valued continuous functions on the real line and  $k_i \geq 0$ . If  $S_n \Rightarrow S$ , then

$$(S_n, Z_{1n}, \dots, Z_{mn}) \Rightarrow (S, Z_1, \dots, Z_m).$$

**Proof.** Omitted.  $\blacktriangle$

The following corollary is useful in deriving the limiting distribution of estimators of the parameters of nonstationary autoregressive processes.

**Corollary 5.3.6.** Let  $\{e_i\}$  be a sequence satisfying the conditions of Theorem 5.3.5. Define

$$Y_t = Y_{t-1} + e_t = \sum_{i=1}^t e_i,$$

for  $t \geq 1$  with  $Y_0 = 0$ . Then

$$\begin{aligned} & \left( n^{-1/2} Y_n, n^{-3/2} \sum_{t=2}^n Y_{t-1}, n^{-2} \sum_{t=2}^n Y_{t-1}^2, n^{-5/2} \sum_{t=2}^n t Y_{t-1} \right) \\ & \xrightarrow{\mathcal{L}} \left( \sigma W(1), \sigma \int_0^1 W(t) dt, \sigma^2 \int_0^1 W^2(t) dt, \sigma \int_0^1 t W(t) dt \right). \end{aligned}$$

**Proof.** From Theorem 5.3.5, we have that

$$S_n(u) = n^{-1/2} \sum_{j=1}^{\lfloor nu \rfloor} e_j$$

converges weakly to  $\sigma W(u)$ , where  $W(u)$  is the Wiener process. If we let

$$f_1(x) = f_3(x) = x$$

and  $f_2(x) = x^2$ , then

$$\begin{aligned} n^{-3/2} \sum_{t=2}^n Y_{t-1} &= \int_0^1 f_1[S_n(u)] du, \\ n^{-2} \sum_{t=2}^n Y_{t-1}^2 &= \int_0^1 f_2[S_n(u)] du, \end{aligned}$$

and

$$n^{-5/2} \sum_{t=2}^n t Y_{t-1} = \int_0^1 u f_3[S_n(u)] du,$$

The result follows from Theorem 5.3.6. ▲

Theorem 5.3.7 is the vector version of the functional central limit theorem. We define a standard vector Wiener process of dimension  $k$  to be a process such that the elements of  $W(u)$  are independent standard Wiener processes.

**Theorem 5.3.7.** Let  $\{\mathbf{e}_i\}_{i=1}^\infty$  be a sequence of random vectors, and let  $\mathcal{A}_{t-1}$  be the sigma-field generated by  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{t-1}\}$ . Assume

$$E\{(\mathbf{e}_t, \mathbf{e}_t \mathbf{e}'_t) | \mathcal{A}_{t-1}\} = (\mathbf{0}, \Sigma_{ee}) \quad \text{a.s.}$$

and

$$E\{|e_i|^{2+\delta} | \mathcal{A}_{i-1}\} < M < \infty \quad \text{a.s.}$$

for some  $\delta > 0$ , or assume the  $e_i$  are iid(0,  $\Sigma_{ee}$ ), where  $\Sigma_{ee}$  is positive definite. Let

$$\mathbf{Y}_t = \sum_{j=1}^t \mathbf{e}_j$$

and

$$\mathbf{S}_n(u) = n^{-1/2} \Sigma_{ee}^{-1/2} \mathbf{Y}_{[nu]}, \quad 0 \leq u \leq 1,$$

where  $[nu]$  is the integer part of  $nu$ , and  $\Sigma_{ee}^{1/2}$  is the symmetric square root of  $\Sigma_{ee}$ . Then

$$\mathbf{S}_n(u) \Rightarrow \mathbf{W},$$

where  $\mathbf{W}$  is a standard vector Wiener process. Also,

$$\begin{aligned} n^{-3/2} \Sigma_{ee}^{-1/2} \sum_{t=1}^n \mathbf{Y}_t &\Rightarrow \int_0^1 \mathbf{W}(u) du \stackrel{\text{def}}{=} \zeta, \\ n^{-2} \Sigma_{ee}^{-1/2} \sum_{t=1}^n \mathbf{Y}_t \mathbf{Y}_t' \Sigma_{ee}^{-1/2} &\Rightarrow \int_0^1 \mathbf{W}(u) \mathbf{W}'(u) du \stackrel{\text{def}}{=} \mathbf{G}, \\ n^{-2} \Sigma_{ee}^{-1/2} \sum_{t=1}^n (\mathbf{Y}_t - \bar{\mathbf{y}})(\mathbf{Y}_t - \bar{\mathbf{y}})' \Sigma_{ee}^{-1/2} &\Rightarrow \mathbf{G} - \zeta \zeta', \\ n^{-1} \Sigma_{ee}^{-1/2} \sum_{t=1}^n \mathbf{Y}_{t-1} \mathbf{e}_t' \Sigma_{ee}^{-1/2} &\Rightarrow \int_0^1 \mathbf{W}(u) d\mathbf{W}'(u) \stackrel{\text{def}}{=} \mathbf{Y}, \end{aligned}$$

and

$$n^{-1} \Sigma_{ee}^{-1/2} \sum_{t=1}^n (\mathbf{Y}_{t-1} - \bar{\mathbf{y}}) \mathbf{e}_t' \Sigma_{ee}^{-1/2} = \mathbf{Y} - \zeta \mathbf{W}'(1).$$

**Proof.** See Phillips and Durlauf (1986). ▲

We shall have use for a strong consistency result for martingales.

**Theorem 5.3.8.** Let  $\{S_n = \sum_{i=1}^n X_i, \mathcal{A}_n, n \geq 1\}$  be a martingale, and let  $\{M_n, n \geq 1\}$  be a nondecreasing sequence of positive random variables where  $M_n$  is  $\mathcal{A}_{n-1}$  measurable for each  $n$ . Let

$$\lim_{n \rightarrow \infty} M_n = \infty \quad \text{a.s.}$$

and

$$\sum_{i=1}^{\infty} M_i^{-p} E\{|X_i|^p | \mathcal{A}_{i-1}\} < \infty \quad \text{a.s.}$$

for some  $1 \leq p \leq 2$ . Then

$$\lim_{n \rightarrow \infty} M_n^{-1} S_n = 0 \quad \text{a.s.}$$

**Proof.** See Hall and Heyde (1980, p. 36).  $\blacktriangle$

**Corollary 5.3.8.** Let  $\{e_i\}_{i=1}^{\infty}$  be a sequence of random variables, and let  $\mathcal{A}_{i-1}$  be the sigma-field generated by  $\{e_1, e_2, \dots, e_{i-1}\}$ . Assume

$$E\{e_i | \mathcal{A}_{i-1}\} = 0, \quad E\{|e_i|^{1+\delta} | \mathcal{A}_{i-1}\} < K_F < \infty \quad \text{a.s.,}$$

where  $K_F$  is a fixed constant and  $\delta > 0$ . Then

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n e_i = 0 \quad \text{a.s.}$$

**Proof.** By assumption

$$\sum_{i=1}^n i^{-(1+\delta)} E\{|e_i|^{1+\delta} | \mathcal{A}_{i-1}\} < \sum_{i=1}^n i^{-(1+\delta)} K_F < \infty.$$

The conclusion follows by letting  $M_i = i$  in Theorem 5.3.8.  $\blacktriangle$

#### 5.4. APPROXIMATING A SEQUENCE OF EXPECTATIONS

Taylor expansions and the order in probability concepts introduced in Section 5.1 are very important in investigating the limiting behavior of sample statistics. Care must be taken, however, in understanding the meaning of statements about such behavior. To illustrate, consider the sequence  $\{\bar{x}_n^{-1}\}$ , where  $\bar{x}_n$  is the mean of  $n$  normal independent  $(\mu, 1)$  random variables,  $\mu \neq 0$ . By Corollary 5.1.5, we may write

$$\bar{x}_n^{-1} = \mu^{-1} - \mu^{-2}(\bar{x}_n - \mu) + \mu^{-3}(\bar{x}_n - \mu)^2 - \mu^{-4}(\bar{x}_n - \mu)^3 + O_p(n^{-2}) \quad (5.4.1)$$

and

$$n^{1/2}(\bar{x}_n^{-1} - \mu^{-1}) = -n^{1/2}\mu^{-2}(\bar{x}_n - \mu) + O_p(n^{-1/2}).$$

It follows that

$$p\lim n^{1/2}[(\bar{x}_n^{-1} - \mu^{-1}) + \mu^{-2}(\bar{x}_n - \mu)] = 0.$$

Therefore, by Theorem 5.2.1, the limiting distribution of  $n^{1/2}(\bar{x}_n^{-1} - \mu^{-1})$  is the

same as the limiting distribution of  $-n^{1/2}\mu^{-2}(\bar{x}_n - \mu)$ . The distribution of  $-n^{1/2}\mu^{-2}(\bar{x}_n - \mu)$  is  $N(0, \mu^{-4})$  for all  $n$ , and it follows that the limiting distribution of  $n^{1/2}(\bar{x}_n^{-1} - \mu^{-1})$  is  $N(0, \mu^{-4})$ .

On the other hand, it can be demonstrated that  $E\{\bar{x}_n^{-1}\}$  exists for no finite  $n$ . Since the expectation of  $(\bar{x}_n^{-1} - \mu^{-1})$  is not defined, it is clear that one cannot speak of the sequence of expectations  $\{E\{\bar{x}_n^{-1}\}\}$ , and it is incorrect to say that  $E\{\bar{x}_n^{-1}\}$  converges to  $\mu^{-1}$ .

The example illustrates that a random variable  $Y_n$  may converge in probability and hence in distribution to a random variable  $Y$  that possesses finite moments even though  $E\{Y_n\}$  is not defined. If we know that  $Y_n$  has finite moments of order  $r > 1$ , we may be able to determine that the sequence of expectations differs from a given sequence by an amount of specified order. The conditions required to permit such statements are typically more stringent than those required to obtain convergence in distribution. In this section we investigate these conditions and develop approximations to the expectation of functions of mean or "meanlike" statistics. In preparation for that investigation we consider the expectations of integer powers of sample means of random variables with zero population means. Let  $(\bar{x}_n, \bar{y}_n, \bar{z}_n, \bar{w}_n)'$  be a vector of sample means computed from a random sample selected from a distribution function with zero mean vector and finite fourth moments. Then

$$\begin{aligned} E\{\bar{x}_n \bar{y}_n \bar{z}_n\} &= \frac{1}{n^3} E\left\{ \sum_i \sum_j \sum_k X_i Y_j Z_k \right\} \\ &= \frac{1}{n^3} E\left\{ \sum_i X_i Y_i Z_i + \sum_{i \neq j} X_i Y_j Z_j + \sum_{i \neq j} X_j Y_i Z_i \right. \\ &\quad \left. + \sum_{i \neq j} X_j Y_j Z_i + \sum_{i \neq j \neq k} X_i Y_j Z_k \right\} \\ &= \frac{1}{n^2} E\{XYZ\}, \end{aligned}$$

where  $E\{XYZ\}$  is the expectation of the product of the original random variables. Similarly,

$$\begin{aligned} E\{\bar{x}_n \bar{y}_n \bar{z}_n \bar{w}_n\} &= \frac{1}{n^4} E\left\{ \sum_i \sum_j \sum_k \sum_m X_i Y_j Z_k W_m \right\} \\ &= \frac{1}{n^4} E\left\{ \sum_i X_i Y_i Z_i W_i \right. \\ &\quad \left. + \sum_{i \neq j} (X_i Y_j Z_j W_i + X_j Y_i Z_i W_j + X_i Y_j Z_i W_j + X_j Y_i Z_j W_i) \right. \\ &\quad \left. + \sum_{i \neq j} (X_i Y_i Z_j W_j + X_i Y_j Z_i W_j + X_i Y_j Z_i W_j + X_i Y_j Z_j W_i) \right. \\ &\quad \left. + \sum_{i \neq j \neq k} (X_i Y_j Z_k W_k + X_i Y_j Z_i W_k + X_i Y_j Z_k W_i + X_i Y_j Z_i W_k) \right\} \end{aligned}$$

$$\begin{aligned}
 & + X_i Y_j Z_k W_l + X_j Y_k Z_l W_i) + \sum_{i \neq j \neq k \neq m} X_i Y_j Z_k W_m \Big\} \\
 & = \frac{1}{n^3} E\{XYZW\} + \frac{n-1}{n^3} (\sigma_{xy}\sigma_{zw} + \sigma_{xz}\sigma_{yw} + \sigma_{xw}\sigma_{yz}),
 \end{aligned}$$

where  $\sigma_{xy}$  is the covariance between  $X$  and  $Y$ ,  $\sigma_{zw}$  is the covariance between  $Z$  and  $W$ , and so forth.

We note that the expectation of a product of either three or four means is  $O(n^{-2})$ . This is an example of a general result that we state as a theorem. Our proof follows closely that of Hansen, Hurwitz, and Madow (1953).

**Theorem 5.4.1.** Let  $\bar{\mathbf{x}}_n = (\bar{x}_{1n}, \bar{x}_{2n}, \dots, \bar{x}_{mn})'$  be the mean of a random sample of  $n$  vector random variables selected from a distribution function with mean vector zero and finite  $B$ th moment. Consider the sequence  $\{\bar{\mathbf{x}}_n\}_{n=1}^\infty$ , and let  $b_1, b_2, \dots, b_m$  be nonnegative integers such that  $B = \sum_{i=1}^m b_i$ . Then

$$E\{\bar{x}_{1n}^{b_1} \bar{x}_{2n}^{b_2} \cdots \bar{x}_{mn}^{b_m}\} = \begin{cases} O(n^{-B/2}) & \text{if } B \text{ is even,} \\ O(n^{-(B+1)/2}) & \text{if } B \text{ is odd.} \end{cases}$$

**Proof.** Now  $E\{\bar{x}_{1n}^{b_1} \bar{x}_{2n}^{b_2} \cdots \bar{x}_{mn}^{b_m}\}$  can be expanded into a sum of terms such as

$$n^{-B} X_{1i_1} X_{1i_2} \cdots X_{1i_{b_1}} X_{2i_{b_1+1}} X_{2i_{b_1+2}} \cdots X_{2i_{b_1+b_2}} \cdots X_{mi_{B-b_m+1}} X_{mi_{B-b_m+2}} \cdots X_{mi_B}.$$

If there is a subscript matched by no other subscript, the expected value of the product is zero. (Recall that in the four-variable case this included terms of the form  $X_i Y_j Z_l W_j$ ,  $X_i Y_j Z_k Z_k$ , and  $X_i Y_j Z_k Z_r$ .) If every subscript agrees with at least one other subscript, we group the terms with common subscripts to obtain, say,  $H$  groups. The expected value is then the product of the  $H$  expected values. The sum contains  $n(n-1) \cdots (n-H+1)$  terms for a particular configuration of  $H$  different subscripts. The order of  $n^{-B}[n(n-1) \cdots (n-H+1)]$  is  $n^{-B+H}$  and will be maximized if we choose  $H$  as large as possible. If  $B$  is even, the largest  $H$  that gives a nonzero expectation is  $B/2$ , in which case we have  $B/2$  groups, each containing two indexes. If  $B$  is odd, the largest  $H$  that gives a nonzero expectation is  $(B-1)/2$ , in which case we have  $(B-1)/2 - 1$  groups of two and one group of three.  $\blacktriangle$

The following lemma is used in the proof of the principal results of this section.

**Lemma 5.4.1.** Let  $\{\mathbf{X}_n\}$  be a sequence of  $k$ -dimensional random variables with corresponding distribution functions  $\{F_n(\mathbf{x})\}$  such that

$$\int |x_i - \mu_i|^s dF_n(\mathbf{x}) = O(a_n^s), \quad i = 1, 2, \dots, k,$$

where the integral is over  $\mathcal{R}^{(k)}$ ,  $a_n > 0$ ,  $s$  is a positive integer,  $E\{\mathbf{X}_n\} = \boldsymbol{\mu} =$

$(\mu_1, \mu_2, \dots, \mu_k)', |\mathbf{x}| = [\sum_{i=1}^k x_i^2]^{1/2}$  is the Euclidean norm, and

$$\lim_{n \rightarrow \infty} a_n = 0.$$

Then

$$\int |x_1 - \mu_1|^{p_1} |x_2 - \mu_2|^{p_2} \cdots |x_k - \mu_k|^{p_k} dF_n(\mathbf{x}) = O(1),$$

where the  $p_i$ ,  $i = 1, 2, \dots, k$ , are nonnegative real numbers satisfying

$$\sum_{i=1}^k p_i \leq s.$$

**Proof.** Without loss of generality, set all  $\mu_i = 0$ . Define  $A = [-1, 1]$ , and let  $I_A(x)$  be the indicator function with value one for  $x \in A$  and zero for  $x \notin A$ . Then, for  $0 \leq q \leq s$ ,

$$|x_i|^q \leq I_A(x_i) + |x_i|^s,$$

so that

$$\int |x_i|^q dF_n(\mathbf{x}) \leq \int [I_A(x_i) + |x_i|^s] dF_n(\mathbf{x}) \leq 1 + O(a_n^s) = O(1),$$

where the integrals are over  $\mathcal{R}^{(k)}$ . By the Hölder inequality,

$$\begin{aligned} & \int |x_1|^{p_1} |x_2|^{p_2} \cdots |x_k|^{p_k} dF_n(\mathbf{x}) \\ & \leq \left[ \int |x_1|^r dF_n(\mathbf{x}) \right]^{p_1/r} \left[ \int |x_2|^r dF_n(\mathbf{x}) \right]^{p_2/r} \cdots \left[ \int |x_k|^r dF_n(\mathbf{x}) \right]^{p_k/r} \\ & = O(1), \end{aligned}$$

where

$$r = \sum_{i=1}^k p_i.$$

▲

**Theorem 5.4.2.** Let  $\{X_n\}$  be a sequence of real valued random variables with corresponding distribution functions  $\{F_n(x)\}$ , and let  $\{f_n(x)\}$  be a sequence of real valued functions. Assume that for some positive integers  $s$  and  $N_0$ :

- (i)  $\int |x - \mu|^{2s} dF_n(x) = a_n^{2s}$ , where  $a_n \rightarrow 0$  as  $n \rightarrow \infty$ .
- (ii)  $\int |f_n(x)|^2 dF_n(x) = O(1)$ .
- (iii)  $f_n^{(s)}(x)$  is continuous in  $x$  over a closed and bounded interval  $S$  for  $n$  greater than  $N_0$ , where  $f_n^{(j)}(x)$  denotes the  $j$ th derivative of  $f_n(x)$  evaluated at  $x$  and  $f_n^{(0)}(x) = f_n(x)$ .

- (iv)  $\mu$  is an interior point of  $S$ .  
(v) There is a  $K$  such that, for  $n > N_0$ ,

$$|f_n^{(s)}(x)| \leq K \quad \text{for all } x \in S$$

and

$$|f_n^{(r)}(\mu)| \leq K \quad \text{for } r = 0, 1, \dots, s-1.$$

Then

$$\int f_n(x) dF_n(x) = f_n(\mu) + Q(n, s) + O(a_n^2),$$

where

$$Q(n, s) = \begin{cases} 0, & s = 1, \\ \sum_{j=1}^{s-1} (1/j!) f_n^{(j)}(\mu) \int (x - \mu)^j dF_n(x), & s > 1. \end{cases}$$

**Proof.** See the proof of Theorem 5.4.3. ▲

**Theorem 5.4.3.** Let  $\{\mathbf{X}_n\}$  be a sequence of  $k$ -dimensional random variables with corresponding distribution functions  $\{F_n(\mathbf{x})\}$ , and let  $\{f_n(\mathbf{x})\}$  be a sequence of functions mapping  $\mathcal{R}^{(k)}$  into  $\mathcal{R}$ . Let  $\delta \in (0, \infty)$ , and define  $\alpha = \delta^{-1}(1 + \delta)$ . Assume that for some positive integers  $s$  and  $N_0$ :

- (i)  $\int |\mathbf{x} - \mu|^{\alpha s} dF_n(\mathbf{x}) = a_n^{\alpha s}$ , where  $a_n \rightarrow 0$  as  $n \rightarrow \infty$ .
- (ii)  $\int |f_n(\mathbf{x})|^{1+\delta} dF_n(\mathbf{x}) = O(1)$ .
- (iii)  $f_n^{(i_1, \dots, i_s)}(\mathbf{x})$  is continuous in  $\mathbf{x}$  over a closed and bounded sphere  $S$  for all  $n$  greater than  $N_0$ , where

$$f_n^{(i_1, \dots, i_r)}(\mathbf{x}_0) = \frac{\partial^r}{\partial x_{i_1} \cdots \partial x_{i_r}} f_n(\mathbf{x}) \Big|_{\mathbf{x}=\mathbf{x}_0}.$$

- (iv)  $\mu$  is an interior point of  $S$ .
- (v) There is a finite number  $K$  such that, for  $n > N_0$ ,

$$|f_n^{(i_1, \dots, i_s)}(\mathbf{x})| \leq K \quad \text{for all } \mathbf{x} \in S,$$

$$|f_n^{(i_1, \dots, i_r)}(\mu)| \leq K \quad \text{for } r = 1, 2, \dots, s-1,$$

and

$$|f_n(\mu)| \leq K.$$

Then

$$\int f_n(\mathbf{x}) dF_n(\mathbf{x}) = f_n(\boldsymbol{\mu}) + \sum_{j=1}^{s-1} \frac{1}{j!} \int D^j f_n(\boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^j dF_n(\mathbf{x}) + O(a_n^s),$$

where

$$D^r f_n(\boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^r = \sum_{i_1=1}^k \sum_{i_2=1}^k \cdots \sum_{i_r=1}^k f_n^{(i_1, \dots, i_r)}(\boldsymbol{\mu}) \prod_{j=1}^r (x_{i_j} - \mu_{i_j})$$

and, for  $s = 1$ , it is understood that

$$\int f_n(\mathbf{x}) dF_n(\mathbf{x}) = f_n(\boldsymbol{\mu}) + O(a_n).$$

The result also holds if we replace (ii) with the condition that the  $f_n(\mathbf{x})$  are uniformly bounded for  $n$  sufficiently large and assume that (i), (iii), (iv), and (v) hold for  $\alpha = 1$ .

**Proof.** We consider only those  $n$  greater than  $N_0$ . By Taylor's theorem there is a sequence of functions  $\{\mathbf{Y}_n\}$  mapping  $S$  into  $S$  such that

$$f_n(\mathbf{x}) = f_n(\boldsymbol{\mu}) + I_S(\mathbf{x}) \sum_{j=1}^{s-1} \frac{1}{j!} D^j f_n(\boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^j + R_n(\mathbf{x}),$$

where

$$I_S(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in S, \\ 0 & \text{otherwise} \end{cases}$$

and

$$R_n(\mathbf{x}) = \begin{cases} (s!)^{-1} D^s f_n(\mathbf{Y}_n(\mathbf{x})) (\mathbf{x} - \boldsymbol{\mu})^s & \text{if } \mathbf{x} \in S, \\ f_n(\mathbf{x}) - f_n(\boldsymbol{\mu}) & \text{otherwise.} \end{cases}$$

For  $\mathbf{x} \in S$  we have  $\mathbf{Y}_n(\mathbf{x}) \in S$ , so that

$$\begin{aligned} |R_n(\mathbf{x})| &\leq (s!)^{-1} \sum_{i_1=1}^k \cdots \sum_{i_s=1}^k \{|f_n^{(i_1, \dots, i_s)}(\mathbf{Y}_n(\mathbf{x}))| |x_{i_1} - \mu_{i_1}| \cdots |x_{i_s} - \mu_{i_s}|\} \\ &\leq (s!)^{-1} \sum_{i_1=1}^k \cdots \sum_{i_s=1}^k K |\mathbf{x} - \boldsymbol{\mu}|^s \\ &= (s!)^{-1} K k^s |\mathbf{x} - \boldsymbol{\mu}|^s. \end{aligned}$$

Thus,

$$\begin{aligned} \int_S |R_n(\mathbf{x})| dF_n(\mathbf{x}) &\leq (s!)^{-1} K k^s \int_S |\mathbf{x} - \boldsymbol{\mu}|^s dF_n(\mathbf{x}) \\ &= O(a_n^s). \end{aligned}$$

Now, for finite  $\delta$ , and letting  $\tilde{S}$  denote the complement of  $S$ ,

$$\begin{aligned} \int_S |R_n(\mathbf{x})| dF_n(\mathbf{x}) &= \int I_S(\mathbf{x}) |f_n(\mathbf{x}) - f_n(\boldsymbol{\mu})| dF_n(\mathbf{x}) \\ &\leq \left[ \int |f_n(\mathbf{x}) - f_n(\boldsymbol{\mu})|^{1+\delta} dF_n(\mathbf{x}) \right]^{1/(1+\delta)} \left[ \int I_S(\mathbf{x}) dF_n(\mathbf{x}) \right]^{\delta/(1+\delta)} \end{aligned}$$

by Hölder's inequality. By Theorem 5.1.1 (Chebyshev's inequality),

$$\int I_S(\mathbf{x}) dF_n(\mathbf{x}) \leq M \int |\mathbf{x} - \boldsymbol{\mu}|^{\alpha s} dF_n(\mathbf{x}) = O(a_n^{\alpha s}) \quad \text{for some } M > 0.$$

Therefore, for  $\delta \in (0, \infty)$ ,

$$\int |R_n(\mathbf{x})| dF_n(\mathbf{x}) = O(a_n^s).$$

This result also holds for  $\alpha = 1$  and  $f_n(\mathbf{x})$  uniformly bounded (by  $K^\dagger$ , say), because then

$$\int I_S(\mathbf{x}) |f_n(\mathbf{x}) - f_n(\boldsymbol{\mu})| dF_n(\mathbf{x}) \leq (2K^\dagger) \int I_S(\mathbf{x}) dF_n(\mathbf{x}) = O(a_n^s).$$

We now have that, for  $s > 1$ ,

$$\begin{aligned} \int f_n(\mathbf{x}) dF_n(\mathbf{x}) &= f_n(\boldsymbol{\mu}) + \int \sum_{j=1}^{s-1} (j!)^{-1} D^j f_n(\boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^j dF_n(\mathbf{x}) \\ &\quad - \int_S \sum_{j=1}^{s-1} (j!)^{-1} D^j f_n(\boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^j dF_n(\mathbf{x}) + O(a_n^s). \end{aligned}$$

However,

$$\begin{aligned} \int_S \sum_{j=1}^{s-1} (j!)^{-1} D^j f_n(\boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^j dF_n(\mathbf{x}) &= \int_S \sum_{j=1}^{s-1} (j!)^{-1} \sum_{i_1=1}^k \cdots \sum_{i_j=1}^k f_n^{(i_1, \dots, i_j)}(\boldsymbol{\mu}) \prod_{r=1}^j (x_{i_r} - \mu_{i_r}) dF_n(\mathbf{x}) \end{aligned}$$

$$\begin{aligned} &\leq \int_S Kk^s[|\mathbf{x} - \boldsymbol{\mu}|^s + 1] dF_n(\mathbf{x}) \\ &= O(a_n^s). \end{aligned}$$

▲

We now give an extension of Theorem 5.4.3 to a product, where one factor is a function of the type studied in Theorem 5.4.3 and the second need not converge to zero.

**Corollary 5.4.3.** Let the assumptions of Theorem 5.4.3 hold. Let  $Z_n$  be a sequence of random variables defined on the same space as  $\mathbf{X}_n$ . Let  $F_n(\mathbf{x}, z)$  be the distribution function of  $(\mathbf{X}_n, Z_n)$ . Assume

- (vi)  $\int |z|^\alpha |(\mathbf{x} - \boldsymbol{\mu})|^{\alpha s} dF_n(\mathbf{x}, z) = O(a_n^{\alpha s}),$
- (vii)  $E\{|Z_n|\} = O(1).$

Then

$$\int z f_n(\mathbf{x}) dF_n(\mathbf{x}, z) = f_n(\boldsymbol{\mu}) E\{Z_n\} + \sum_{j=1}^{s-1} (j!)^{-1} \int D' f_n(\boldsymbol{\mu})(\mathbf{x} - \boldsymbol{\mu})^j z dF_n(\mathbf{x}, z) + O(a_n^s).$$

**Proof.** Omitted. ▲

Theorems 5.4.2 and 5.4.3 require that  $\int |f_n(\mathbf{x})|^{1+\delta} dF_n(\mathbf{x})$  be bounded for all  $n$  sufficiently large. The following theorem gives sufficient conditions for a sequence of integrals  $\int |f_n(\mathbf{x})| dF_n(\mathbf{x})$  to be  $O(1)$ .

**Theorem 5.4.4.** Let  $\{f_n(\mathbf{x})\}$  be a sequence of real valued (measurable) functions, and let  $\{\mathbf{X}_n\}$  be a sequence of  $k$ -dimensional random variables with corresponding distribution functions  $\{F_n(\mathbf{x})\}$ . Assume that:

- (i)  $|f_n(\mathbf{x})| \leq K_1$  for  $\mathbf{x} \in \bar{S}$ , where  $S$  is a bounded open set containing  $\boldsymbol{\mu}$ ,  $\bar{S}$  is the closure of  $S$ , and  $K_1$  is a finite constant.
- (ii)  $|f_n(\mathbf{x})| \leq Y(\mathbf{x})n^p$  for some  $p > 0$  and for a function  $Y(\cdot)$  such that  $\int |Y(\mathbf{x})|^\gamma dF_n(\mathbf{x}) = O(1)$  for some  $\gamma$ ,  $1 < \gamma < \infty$ .
- (iii)  $\int |\mathbf{x} - \boldsymbol{\mu}|^r dF_n(\mathbf{x}) = O(n^{-\eta p})$  for a positive integer  $r$  and an  $\eta$  such that  $1/\eta + 1/\gamma = 1$ .

Then

$$\int |f_n(\mathbf{x})| dF_n(\mathbf{x}) = O(1).$$

The result also holds for  $\eta = 1$  given that (i), (iii), and  $|f_n(\mathbf{x})| \leq K_2 n^p$  hold for all  $\mathbf{x}$ , some  $p > 0$ , and  $K_2$  a finite constant.

**Proof.** Let  $\delta > 0$  be such that

$$A = \{\mathbf{x}: |\mathbf{x} - \boldsymbol{\mu}| \leq \delta\} \subset S.$$

By Chebyshev's inequality,

$$\begin{aligned} P\{\mathbf{X}_n \in \tilde{A}\} &= P\{|\mathbf{X}_n - \boldsymbol{\mu}| > \delta\} \\ &\leq \delta^{-r} \int |\mathbf{x} - \boldsymbol{\mu}|^r dF_n(\mathbf{x}) \\ &= O(n^{-np}). \end{aligned}$$

For  $1 < \gamma < \infty$ ,

$$\begin{aligned} \int |f_n(\mathbf{x})| dF_n(\mathbf{x}) &= \int_A |f_n(\mathbf{x})| dF_n(\mathbf{x}) + \int_{\tilde{A}} |f_n(\mathbf{x})| dF_n(\mathbf{x}) \\ &\leq \int_A K_1 dF_n(\mathbf{x}) + n^p \int_{\tilde{A}} |Y(\mathbf{x})| dF_n(\mathbf{x}) \\ &\leq K_1 + n^p \int I_{\tilde{A}}(\mathbf{x}) |Y(\mathbf{x})| dF_n(\mathbf{x}) \\ &\leq K_1 + n_p \left[ \int I_{\tilde{A}}(\mathbf{x}) dF_n(\mathbf{x}) \right]^{1/\eta} \left[ \int |Y(\mathbf{x})|^\gamma dF_n(\mathbf{x}) \right]^{1/\gamma} \\ &= K_1 + n^p [P\{\mathbf{X}_n \in \tilde{A}\}]^{1/\eta} [O(1)] \\ &= O(1). \end{aligned}$$

For  $|f_n(\mathbf{x})| \leq K_2 n^p$  and  $\eta = 1$ , we have

$$\begin{aligned} \int |f_n(\mathbf{x})| dF_n(\mathbf{x}) &\leq K_1 + K_2 n^p \int I_{\tilde{A}}(\mathbf{x}) dF_n(\mathbf{x}) \\ &= O(1). \end{aligned}$$

▲

**Example 5.4.1.** To illustrate some of the ideas of this section, we consider the problem of estimating  $\log \mu$ , where  $\mu > 0$  is the mean of a random variable with finite sixth moment. Let  $\bar{x}_n$  be the mean of a simple random sample of  $n$  such observations. Since  $\log x$  is not defined for  $x \leq 0$ , we consider the estimator

$$f_n(\bar{x}_n) = \begin{cases} \log \bar{x}_n & \text{if } \bar{x}_n > 1/n, \\ -\log n & \text{if } \bar{x}_n \leq 1/n. \end{cases}$$

Suppose that we desire an approximation to the sequence of expectations of  $f_n(\bar{x}_n)$  to order  $n^{-1}$ . We first apply Theorem 5.4.2 with  $a_n = O(n^{-1/2})$  and  $s = 3$ . By Theorem 5.4.1, we have that

$$E\{(\bar{x}_n - \mu)^{2s}\} = O(n^{-s}), \quad s = 1, 2, 3, \tag{5.4.2}$$

so that condition i of Theorem 5.4.2 is met. To establish condition ii, we demonstrate that  $|f_n(x)|^2$  satisfies the conditions of Theorem 5.4.4. Since  $f_n(x)$  is continuous for  $x > 0$ , condition i of Theorem 5.4.4 is satisfied. For  $p = 1$ ,  $\gamma = \eta = 2$ , and

$$Y(x) = \begin{cases} 1, & x < 1, \\ x & \text{otherwise,} \end{cases}$$

we have that  $|f_n(x)|^2 < nY(x)$  and  $\int |Y(x)|^2 dF_n(x) = O(1)$ . This is condition ii of Theorem 5.4.4; from equation (5.4.2) we see that condition iii also holds for  $r = 4$ , 5, or 6. Therefore, the conditions of Theorem 5.4.4 are satisfied and the integral of  $|f_n(x)|^2$  is order one.

Since the third derivative of  $\log x$  is continuous for positive  $x$ , we may choose an  $N_0 > \mu^{-1}$  and an interval containing  $\mu$  such that  $f_n^{(3)}(x)$  is continuous on that interval for all  $n > N_0$ . To illustrate, let  $\mu = 0.1$ , and take the interval  $S$  to be  $[0.05, 0.15]$ . Then, for  $n > 20$ ,  $f_n^{(3)}(x)$  is continuous on  $S$ . Therefore, conditions iii, iv, and v of Theorem 5.4.2 are also met, and we may write

$$\begin{aligned} E\{f_n(\bar{x}_n)\} &= \log \mu + \frac{1}{\mu} E\{\bar{x}_n - \mu\} \\ &\quad - \frac{1}{2\mu^2} E\{(\bar{x}_n - \mu)^2\} + O(n^{-3/2}) \\ &= \log \mu - \frac{\sigma^2}{2n\mu^2} + O(n^{-3/2}), \end{aligned}$$

where  $\sigma^2$  is the variance of the observations. To order  $n^{-1}$ , the bias in  $f_n(\bar{x}_n)$  as an estimator of  $\log \mu$  is  $-(2n\mu^2)^{-1}\sigma^2$ .

Since  $\bar{x}_n$  possesses finite fifth moments, we can decrease the above remainder term to  $O(n^{-2})$  by carrying the expansion to one more term and using the more general form of Theorem 5.4.3. By equation (5.4.2), condition i of Theorem 5.4.3 holds for  $s = 4$ ,  $\alpha = \frac{3}{2}$ , and  $a_n = O(n^{-1/2})$ , so that  $\delta = 2$  by definition. Condition ii is established by demonstrating that  $|f_n(x)|^3$  satisfies the conditions of Theorem 5.4.4. Defining

$$Y(x) = \begin{cases} 1, & x < 1, \\ x^{3/2}, & x \geq 1, \end{cases}$$

and letting  $p = \frac{3}{2}$ , we see that  $|f_n(x)|^3 \leq Y(x)n^p$  and

$$\int |Y(x)|^2 dF_n(x) \leq \int |1 + x^{3/2}|^2 dF_n(x) = O(1),$$

as  $\bar{x}_n$  has finite third moments. Therefore, we are using  $\gamma = \eta = 2$  and  $\eta p = 3$ . Since  $\int |x - \mu|^5 dF_n(x) = O(n^{-3})$  and  $|f_n(x)|^3$  is continuous, the conditions of Theorem 5.4.4 are met. The sixth derivative of  $\log x$  is continuous for positive  $x$ ,

and we can find an  $N_0$  and a set  $S$  such that conditions iii, iv, and v of Theorem 5.4.3 are also met. Because  $f_n^{(3)}(\mu) = O(1)$  and  $E\{(\bar{x}_n - \mu)^3\} = O(n^{-2})$ , we have

$$\begin{aligned} E\{f_n(\bar{x}_n)\} &= \log \mu - (2n\mu^2)^{-1}\sigma^2 \\ &\quad + \frac{1}{3!} f_n^{(3)}(\mu)E\{(\bar{x}_n - \mu)^3\} + O(n^{-2}) \\ &= \log \mu - (2n\mu^2)^{-1}\sigma^2 + O(n^{-2}). \end{aligned} \quad \blacksquare$$

## 5.5. ESTIMATION FOR NONLINEAR MODELS

### 5.5.1. Estimators That Minimize an Objective Function

The expected value of a time series is sometimes expressible as a nonlinear function of unknown parameters and observable functions of time. For example, we might have for  $\{Y_t: t \in (0, 1, 2, \dots)\}$

$$E\{Y_t\} = \sum_{j=0}^t \lambda^j x_{t-j},$$

where  $\{x_t\}$  is a sequence of known constants and  $\lambda$  is unknown. The estimation of a parameter such as  $\lambda$  is considerably more complicated than the estimation of a parameter that enters the expected value function in a linear manner.

We assume the model

$$Y_t = f_t(\theta^0) + e_t, \quad (5.5.1)$$

for  $t = 1, 2, \dots$ , where the random variables are defined on a complete probability space  $(\Omega, \mathcal{A}, P)$ , the vector  $\theta^0 = (\theta_1^0, \theta_2^0, \dots, \theta_k^0)'$  of unknown parameters lies in a compact subset  $\Theta$  of  $k$ -dimensional Euclidean space  $\mathcal{R}^k$ , the function  $f_t$  is defined on  $\Omega \times \Theta$  with the form of  $f_t$  known, and  $f_t(\omega, \cdot)$  is continuous on  $\Theta$  almost surely, for all  $t$ . The expression  $f_t(\omega, \cdot)$  represents the function for a particular  $\omega$  in  $\Omega$ . The dependence of the function on  $\omega$  is generally suppressed for notational convenience. The function  $f_t$  may depend on an input vector  $x_t$ , and sometimes we may write the model as

$$Y_t = f(x_t; \theta^0) + e_t.$$

The dimension of  $x_t$  need not be fixed, and it may increase with  $t$ . Some elements of  $x_t$  may be fixed and some may be random.

The derivatives of the function  $f_t(\theta)$  are very important in the treatment of this problem, and we introduce a shorthand notation for them. Let  $f_t^{(j)}(\hat{\theta})$  denote the first partial derivative of  $f_t(\theta)$  with respect to the  $j$ th element of  $\theta$  evaluated at the point  $\theta = \hat{\theta}$ . Likewise, let  $f_t^{(js)}(\hat{\theta})$  denote the second partial derivative of  $f_t(\theta)$  with respect to the  $j$ th and  $s$ th elements of  $\theta$  evaluated at the point  $\theta = \hat{\theta}$ . The third

partial derivative,  $f_i^{(jrs)}(\hat{\theta})$ , is defined in a similar manner. Let

$$\mathbf{F}_i(\theta) = [f_i^{(1)}(\theta), f_i^{(2)}(\theta), \dots, f_i^{(k)}(\theta)]' \quad (5.5.2)$$

be the vector of first derivatives, and let

$$\mathbf{F}_{nk}(\theta) = [\mathbf{F}'_1(\theta), \mathbf{F}'_2(\theta), \dots, \mathbf{F}'_n(\theta)]'$$

be the  $n \times k$  matrix of first derivatives.

Let an estimator of the unknown  $\theta^0$  be the  $\theta$  that minimizes  $Q_n(\theta)$ , where  $Q_n(\theta)$  is a function of the observations and  $\theta$ . The least squares estimator of  $\theta^0$  is the  $\theta$  that minimizes

$$Q_{nl}(\theta) = n^{-1} \sum_{i=1}^n [Y_i - f_i(\theta)]^2. \quad (5.5.3)$$

For least squares, the function evaluated at  $\theta^0$  is

$$Q_{nl}(\theta^0) = n^{-1} \sum_{i=1}^n e_i^2.$$

The function  $Q_n(\theta)$  is the negative of the logarithm of the likelihood in the case of maximum likelihood estimation for the model (5.5.1) with normal  $e_i$ .

In order to obtain a limiting distribution for the estimator, the sequence  $\{x_i, e_i\}$ , the function  $f_i(\theta)$ , and the parameter space  $\Theta$  must satisfy certain conditions. Stronger conditions in one area will permit weaker conditions in another. The random variables and the function must be such that the estimator is uniquely defined, at least for large  $n$ . Furthermore, the function being minimized must have a unique minimum "close to" the true parameter value, and the distance must decrease as the sample size increases.

These ideas are made more precise in the following theorems. We begin with results for the general problem and then consider special cases. Let  $\hat{\theta}_n$  in  $\Theta$  be a measurable function that minimizes  $Q_n(\theta)$  on  $\Theta$  almost surely. Lemma 5.5.1 contains sufficient conditions for the consistency of  $\hat{\theta}_n$ . Lemma 5.5.1 is given in Wu (1981) and is based on the idea used by Wald (1949) in a proof of the strong consistency of the maximum likelihood estimator.

**Lemma 5.5.1.** Let  $\hat{\theta}_n$  in  $\Theta$  be a measurable function that minimizes an objective function  $Q_n(\theta)$  on  $\Theta$  almost surely.

(i) Suppose, for any  $\eta > 0$ ,

$$\lim_{n \rightarrow \infty} \inf_{|\theta - \theta^0| > \eta} [Q_n(\theta) - Q_n(\theta^0)] > 0 \quad (5.5.4)$$

almost surely. Then  $\hat{\theta}_n \rightarrow \theta^0$  almost surely as  $n \rightarrow \infty$ .

(ii) Suppose, for any  $\eta > 0$ ,

$$\lim_{n \rightarrow \infty} P \left\{ \inf_{|\theta - \theta^0| > \eta} [Q_n(\theta) - Q_n(\theta^0)] > 0 \right\} = 1. \quad (5.5.5)$$

Then  $\hat{\theta}_n \rightarrow \theta^0$  in probability as  $n \rightarrow \infty$ .

**Proof.** A proof of part i is given by Wu (1981). We prove part ii. Suppose  $\hat{\theta}_n$  does not converge to  $\theta^0$  in probability as  $n \rightarrow \infty$ . Then there exist  $\eta > 0$ ,  $1 \geq \epsilon_\eta > 0$ , and a subsequence  $\{n_k\}$  such that

$$P \{ |\hat{\theta}_{n_k} - \theta^0| \geq \eta \} > \epsilon_\eta \quad \text{for every } n_k,$$

implying

$$P \left\{ \inf_{|\theta - \theta^0| > \eta} [Q_{n_k}(\theta) - Q_{n_k}(\theta^0)] \leq 0 \right\} > \epsilon_\eta \quad \text{for every } n_k.$$

Therefore,

$$\limsup_{n \rightarrow \infty} P \left\{ \inf_{|\theta - \theta^0| > \eta} [Q_n(\theta) - Q_n(\theta^0)] \leq 0 \right\} > \epsilon_\eta.$$

That is,

$$\liminf_{n \rightarrow \infty} P \left\{ \inf_{|\theta - \theta^0| > \eta} [Q_n(\theta) - Q_n(\theta^0)] > 0 \right\} \leq 1 - \epsilon_\eta < 1,$$

which is a contradiction of the condition (5.5.5). Hence, if (5.5.5) holds, then  $\hat{\theta}_n$  converges to  $\theta^0$  in probability as  $n \rightarrow \infty$ .  $\blacktriangle$

The conditions in Lemma 5.5.1 are rather general. Gallant and White (1988, pp. 18–19) give a set of sufficient conditions for strong consistency. We give a form of their result in Lemma 5.5.2. The condition (5.5.6) of the lemma is a uniform law of large numbers for the function  $Q_n(\theta)$ . This condition guarantees that the sample function will be close to the limit function as the sample size increases. The condition (5.5.7) insures that the limit of  $Q_n(\theta)$  has a minimum at  $\theta^0$ . It is sometimes called the identification condition.

**Lemma 5.5.2.** Given  $(\Omega, \mathcal{A}, P)$  and a compact set  $\Theta$  that is a subset of  $\mathcal{R}^k$ , let  $Q_n : \Omega \times \Theta \rightarrow \mathcal{R}$  be a random function continuous on  $\Theta$  a.s., for  $n = 1, 2, \dots$ . Let  $\bar{Q}_n$  be as defined in Lemma 5.5.1. Suppose there exists a function  $\bar{Q}_n : \Theta \rightarrow \mathcal{R}$  such that

$$Q_n(\theta) - \bar{Q}_n(\theta) \rightarrow 0 \quad \text{a.s.} \quad (5.5.6)$$

uniformly on  $\Theta$ . Assume that for any  $\eta > 0$

$$\liminf_{n \rightarrow \infty} \left\{ \inf_{|\theta - \theta^0| \geq \eta} [\bar{Q}_n(\theta) - \bar{Q}_n(\theta^0)] \right\} > 0. \quad (5.5.7)$$

Then  $\hat{\theta}_n - \theta^0 \rightarrow 0$  a.s. as  $n \rightarrow \infty$ .

**Proof.** Now

$$Q_n(\theta) - Q_n(\theta^0) = Q_n(\theta) - \bar{Q}_n(\theta) + \bar{Q}_n(\theta) - \bar{Q}_n(\theta^0) + \bar{Q}_n(\theta^0) - Q_n(\theta^0).$$

Therefore,

$$\begin{aligned} \inf_{\theta} [Q_n(\theta) - Q_n(\theta^0)] \\ \geq \inf_{\theta} [Q_n(\theta) - \bar{Q}_n(\theta)] + \inf_{\theta} [\bar{Q}_n(\theta) - \bar{Q}_n(\theta^0)] + \bar{Q}_n(\theta^0) - Q_n(\theta^0). \end{aligned}$$

The first term on the right of the inequality converges to zero as  $n \rightarrow \infty$  because

$$\left| \inf_{\theta} [Q_n(\theta) - \bar{Q}_n(\theta)] \right| \leq \sup_{\theta} |Q_n(\theta) - \bar{Q}_n(\theta)|$$

and by the assumption (5.5.6), there exists a  $B$  in  $\mathcal{A}$  with  $P(B) = 1$  such that, given any  $\epsilon > 0$ , for each  $\omega$  in  $B$  there exists an integer  $N(\omega, \epsilon) < \infty$  such that for all  $n > N(\omega, \epsilon)$ ,

$$\sup_{\theta} |Q_n(\omega, \theta) - \bar{Q}_n(\theta)| < \epsilon.$$

The difference  $\bar{Q}_n(\theta^0) - Q_n(\theta^0)$  also converges to zero as  $n \rightarrow \infty$  by (5.5.6). The conclusion follows from the assumption (5.5.7) and Lemma 5.5.1.  $\blacktriangle$

A critical assumption of Lemma 5.5.2 is (5.5.6), which guarantees that the random functions  $Q_n(\theta)$  behave like nonrandom functions  $\bar{Q}_n(\theta)$  for large  $n$ . If  $Q_n(\theta)$  can be written as

$$Q_n(\theta) = n^{-1} \sum_{i=1}^n q_i(\theta), \quad (5.5.8)$$

then a natural choice for  $\bar{Q}_n(\theta)$  is

$$\bar{Q}_n(\theta) = n^{-1} \sum_{i=1}^n E\{q_i(\theta)\}. \quad (5.5.9)$$

Gallant and White (1988, p. 23) give a uniform law of large numbers for a function of the form (5.5.8) which is a slight modification of the result of Andrews (1986). We give the result as Lemma 5.5.3.

**Lemma 5.5.3.** Let  $q_t(\theta)$  be a function from  $\Omega \times \Theta$  to  $\mathcal{R}$ ,  $t = 1, 2, \dots$ , where  $(\Omega, \mathcal{A}, P)$  is the underlying probability space, and  $\Theta$  is a compact subset of  $\mathcal{R}^k$ . Assume  $q_t(\omega, \cdot)$  is continuous in  $\theta$  almost surely. Assume  $q_t(\cdot, \theta^0)$  is  $\mathcal{A}$ -measurable for each  $\theta^0$  in  $\Theta$  and for  $t = 1, 2, \dots$ . For given  $\theta^0$  in  $\Theta$  and  $\delta > 0$ , let

$$A(\delta) = \{\theta \in \Theta : |\theta - \theta^0| < \delta\}.$$

Assume that for each  $\theta^0$  in  $\Theta$  there exists a constant  $\delta^0 > 0$  and positive random variables  $L_i^0$  such that

$$n^{-1} \sum_{i=1}^n E\{L_i^0\} = O(1)$$

and

$$|q_t(\theta) - q_t(\theta^0)| \leq L_i^0 |\theta - \theta^0|, \quad t = 1, 2, \dots, \quad \text{a.s.}$$

for all  $\theta$  in the closure of  $A(\delta)$ . Also assume that for all  $0 < \delta < \delta^0$ , where  $\delta^0$  may depend on  $\theta^0$ ,

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n [l_i(\delta) - E\{l_i(\delta)\}] = 0 \quad \text{a.s.}$$

and

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n [s_i(\delta) - E\{s_i(\delta)\}] = 0 \quad \text{a.s.,}$$

where

$$l_i(\delta) = \sup_{A(\delta)} q_i(\theta),$$

and

$$s_i(\delta) = \inf_{A(\delta)} q_i(\theta).$$

Then  $\bar{Q}_n(\theta)$  defined in (5.5.9) is continuous on  $\Theta$  uniformly in  $n$ , and

$$\lim_{n \rightarrow \infty} \sup_{\Theta} |\bar{Q}_n(\theta) - \bar{Q}_n(\theta)| = 0.$$

**Proof.** See Gallant and White (1988, p. 38). ▲

It may be easier to obtain a uniform law of large numbers for the function  $Q_n(\theta)$  directly than for  $\{q_t(\theta), t = 1, 2, \dots, n\}$ . A uniform strong law of large numbers for  $Q_n(\theta)$  is given in Lemma 5.5.4, and a uniform weak law of large numbers is given in Lemma 5.5.5.

**Lemma 5.5.4.** Let  $\Theta$  be a convex, compact subset of  $k$ -dimensional Euclidean space,  $\mathcal{R}^k$ , and let  $\{Q_n(\theta), n = 1, 2, \dots\}$  be a sequence of random functions. Assume

(i) For each  $\theta_1$  in  $\Theta$

$$\lim_{n \rightarrow \infty} [Q_n(\theta_1) - E\{Q_n(\theta_1)\}] = 0 \quad \text{a.s.}$$

(ii) There exist random variables  $L_n$  such that

$$(a) \max_{1 \leq i \leq k} \left\{ \sup_{\theta} \left| \frac{\partial Q_n(\theta)}{\partial \theta_i} \right| \right\} \leq L_n,$$

$$(b) \limsup_{n \rightarrow \infty} L_n < \infty \text{ a.s.,}$$

$$(c) \sup_n E\{L_n\} < \infty.$$

Then

$$\lim_{n \rightarrow \infty} \left\{ \sup_{\theta} |Q_n(\theta) - E\{Q_n(\theta)\}| \right\} = 0 \quad \text{a.s.}$$

**Proof.** Omitted. ▲

**Lemma 5.5.5.** Let  $\Theta$  be a convex, compact subset of  $k$ -dimensional Euclidean space, and let  $\{Q_n(\theta): n = 1, 2, \dots\}$  be a sequence of random functions. Assume:

(i) For each  $\theta_1$  in  $\Theta$ ,

$$p\lim_{n \rightarrow \infty} [Q_n(\theta_1) - Q(\theta_1)] = 0$$

for some  $Q(\theta)$ .

- (ii) There exists a sequence of positive random variables  $\{L_n\}$  and an  $L$  such that for  $\theta_1$  and  $\theta_2$  in  $\Theta$ ,
- (a)  $|Q_n(\theta_1) - Q_n(\theta_2)| \leq L_n |\theta_1 - \theta_2|$ ,
  - (b)  $|Q(\theta_1) - Q(\theta_2)| \leq L |\theta_1 - \theta_2|$ ,
  - (c)  $L_n = O_p(1)$  and  $L = O_p(1)$ .

Then

$$p\lim_{n \rightarrow \infty} Q_n(\theta) - Q(\theta) = 0$$

uniformly in  $\theta$  in  $\Theta$ .

**Proof.** Given  $\eta > 0$ , consider the open set  $G_{\theta_1} = \{\theta: |\theta - \theta_1| < \eta\}$  defined by  $\theta_1$  and  $\eta$ . Since  $\Theta$  is compact, there exists a  $k = k(\eta)$  and  $\theta_1, \theta_2, \dots, \theta_k$  such that

$\Theta = \bigcup_{i=1}^k G_i$ , where  $G_i = G_{\theta_i}$ . Now,

$$\begin{aligned}\sup_{\theta \in \Theta} |Q_n(\theta) - Q(\theta)| &= \max_{1 \leq i \leq k} \sup_{\theta \in G_i} |Q_n(\theta) - Q(\theta)| \\ &\leq \max_{1 \leq i \leq k} \sup_{\theta \in G_i} |Q_n(\theta) - Q_n(\theta_i)| + \max_{1 \leq i \leq k} |Q_n(\theta_i) - Q(\theta_i)| \\ &\quad + \max_{1 \leq i \leq k} \sup_{\theta \in G_i} |Q(\theta) - Q(\theta_i)|.\end{aligned}$$

Given  $\epsilon > 0$  and  $\tau > 0$ , by (i) there exists an  $N_i = N_i(\epsilon, \tau)$  such that for  $n > N_i$ ,

$$P[|Q_n(\theta_i) - Q(\theta_i)| > 3^{-1}\epsilon] < \tau.$$

Thus, for  $n > N = \max(N_1, \dots, N_k)$ ,

$$P[\max_{1 \leq i \leq k} |Q_n(\theta_i) - Q(\theta_i)| \geq 3^{-1}\epsilon] \leq \sum_{i=1}^k P[|Q_n(\theta_i) - Q(\theta_i)| \geq 3^{-1}\epsilon] \leq kr.$$

By (ii)(c), for given  $\delta > 0$  there exists  $M > 0$  such that

$$\sup_n P[L_n > M] < 3^{-1}\delta$$

and  $P[L > M] < 3^{-1}\delta$ . Thus, given  $\epsilon > 0$  and  $\delta > 0$ , choose  $\eta = (3M)^{-1}\epsilon$ . Then,

$$P\left[\max_{1 \leq i \leq k} \sup_{\theta \in G_i} |Q_n(\theta) - Q_n(\theta_i)| > 3^{-1}\epsilon\right] \leq P[\eta L_n > 3^{-1}\epsilon] \leq P[L_n > M] \leq 3^{-1}\delta.$$

Similarly,

$$P\left[\max_{1 \leq i \leq k} \sup_{\theta \in G_i} |Q(\theta) - Q(\theta_i)| > 3^{-1}\epsilon\right] \leq 3^{-1}\delta.$$

Choosing  $\tau = (3k)^{-1}\delta$  gives the result. ▲

Given conditions for consistency of  $\hat{\theta}_n$ , we investigate the limiting distribution of the estimator. We begin by giving conditions under which the least squares estimator converges in law to a normal random variable. In the theorem statement we assume the estimator to be consistent. The other assumptions of the theorem are sufficient for the existence of a consistent sequence of estimators (see Theorem 5.5.3), but are not sufficient to rule out multiple local minima. The theorem is general enough to cover models that contain lagged values of  $Y_t$  in the explanatory vector  $x_t$ . To permit this, the sigma-field  $\mathcal{A}_{t-1}$  of the theorem is generated by  $x_1$  through  $x_t$  and  $e_1$  through  $e_{t-1}$ . For a vector  $a$ , let  $|a| = (a'a)^{1/2}$ .

**Theorem 5.5.1.** Let the model (5.5.1) hold, and let  $\hat{\theta}_n$  be a weakly consistent estimator of  $\theta^0$  that minimizes  $Q_{nl}(\theta)$  of (5.5.3). Let  $\mathcal{A}_{t-1}$  be the sigma-field

generated by

$$\{\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_t, e_1, e_2, \dots, e_{t-1}\}.$$

Assume:

- (i) There exists a convex, compact neighborhood  $S$  of the true parameter vector  $\theta^0$  such that  $f_t(\cdot, \theta)$  is measurable for each  $\theta$  in  $S$ ,  $f_t(\omega, \cdot)$  is almost surely twice continuously differentiable with respect to  $\theta$  on  $S$ ,  $S$  is in  $\Theta$  and  $\theta^0$  is an interior point of  $S$ .
- (ii) The vector sequence  $\{[\mathbf{F}_t(\theta^0)e_t, e_t]\}$  satisfies

$$\begin{aligned} E\{[\mathbf{F}_t(\theta^0)e_t, e_t, e_t^2]|\mathcal{A}_{t-1}\} &= [\mathbf{0}, 0, \sigma^2], \\ E\{[[\mathbf{F}_t(\theta^0)e_t, e_t]]^{2+\delta}|\mathcal{A}_{t-1}\} &< M_F < \infty, \\ E\{\mathbf{F}'_t(\theta^0)\mathbf{F}_t(\theta^0)e_t^2|\mathcal{A}_{t-1}\} &= \mathbf{F}'_t(\theta^0)\mathbf{F}_t(\theta^0)\sigma^2 \end{aligned}$$

a.s., for all  $t$ , where  $\delta > 0$ ,  $\mathbf{F}_t(\theta)$  is defined in (5.5.2), and  $M_F$  is a positive constant.

- (iii) There exists a  $k \times k$  fixed matrix  $\mathbf{B}(\theta)$  such that  $\mathbf{B}(\theta)$  is continuous at  $\theta^0$ ,

$$p\lim_{n \rightarrow \infty} \left[ 0.5 \frac{\partial^2 Q_{nt}(\theta)}{\partial \theta \partial \theta'} - \mathbf{B}(\theta) \right] = \mathbf{0} \quad \text{uniformly on } S,$$

$$p\lim_{n \rightarrow \infty} n^{-1} \mathbf{F}'_{nk}(\theta^0) \mathbf{F}_{nk}(\theta^0) = \lim_{n \rightarrow \infty} n^{-1} E\{\mathbf{F}'_{nk}(\theta^0) \mathbf{F}_{nk}(\theta^0)\} = \mathbf{B}(\theta^0),$$

and  $\mathbf{B}(\theta^0)$  is positive definite.

Let

$$s^2 = (n-k)^{-1} \sum_{i=1}^n [Y_i - f(\mathbf{x}_i; \hat{\theta}_n)]^2. \quad (5.5.10)$$

Then

$$(a) n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{\mathcal{L}} N[\mathbf{0}, \mathbf{B}^{-1}(\theta^0)\sigma^2] \quad (5.5.11)$$

and

$$(b) s^2 \text{ converges to } \sigma^2 \text{ in probability.}$$

**Proof.** Because the estimators  $\hat{\theta}_n$  lie inside the convex, compact neighborhood  $S$  of  $\theta^0$  with probability approaching one as  $n$  increases,  $\hat{\theta}_n$  satisfies

$$\frac{\partial Q_{nt}(\hat{\theta}_n)}{\partial \theta} = 0$$

with probability approaching one as  $n$  increases. We can expand the first derivative

in a Taylor series about  $\theta^0$  to obtain

$$0 = n^{1/2} \frac{\partial Q_{nl}(\theta^0)}{\partial \theta} + \frac{\partial^2 Q_{nl}(\hat{\theta}_n)}{\partial \theta \partial \theta'} n^{1/2} (\hat{\theta}_n - \theta^0) \quad (5.5.12)$$

for  $\hat{\theta}_n$  in  $S$ , where  $\hat{\theta}_n$  is on the line segment joining  $\hat{\theta}_n$  and  $\theta^0$ . By assumption iii and the consistency of  $\hat{\theta}_n$ ,

$$\frac{1}{2} \frac{\partial^2 Q_{nl}(\hat{\theta}_n)}{\partial \theta \partial \theta'} \xrightarrow{P} \mathbf{B}(\theta^0). \quad (5.5.13)$$

Now

$$\frac{\partial Q_{nl}(\theta^0)}{\partial \theta} = -2n^{-1} \mathbf{F}'_{nk}(\theta^0) \mathbf{e}_n, \quad (5.5.14)$$

where  $\mathbf{e}_n = (e_1, e_2, \dots, e_n)'$ . Therefore

$$\hat{\theta} - \theta^0 = \mathbf{B}^{-1}(\theta^0) n^{-1} \mathbf{F}'_{nk}(\theta^0) \mathbf{e}_n + \mathbf{r}_n, \quad (5.5.15)$$

where  $\mathbf{r}_n$  is of smaller order in probability than  $\hat{\theta} - \theta^0$ .

To prove that

$$n^{-1/2} \mathbf{F}'_{nk}(\theta^0) \mathbf{e}_n \xrightarrow{\mathcal{L}} N[\mathbf{0}, \mathbf{B}(\theta^0) \sigma^2]$$

it is enough to show that for any  $k$ -dimensional vector  $\lambda$ ,  $\lambda \neq 0$ ,

$$\lambda' [n^{-1/2} \mathbf{F}'_{nk}(\theta^0) \mathbf{e}_n] = \sum_{i=1}^n n^{-1/2} \left[ \sum_{j=1}^k \lambda_j f_i^{(j)}(\theta^0) e_i \right]$$

converges in law to a univariate normal distribution. We prove this by showing that the conditions of Theorem 5.3.4 hold. Let

$$Z_{tn} = n^{-1/2} \left[ \sum_{j=1}^k \lambda_j f_t^{(j)}(\theta^0) e_j \right],$$

and let  $\mathcal{A}_{t,n}$  be  $\mathcal{A}_t$ ,  $0 \leq t \leq n$ ,  $n \geq 1$ . Then by assumption ii

$$E[Z_{tn} | \mathcal{A}_{t-1,n}] = 0 \quad \text{a.s.}$$

and

$$V_{nn}^2 = \sum_{t=1}^n E[Z_{tn}^2 | \mathcal{A}_{t-1,n}] = \sigma^2 \lambda' [n^{-1} \mathbf{F}'_{nk}(\theta^0) \mathbf{F}_{nk}(\theta^0)] \lambda.$$

Note that

$$V\{n^{-1/2} \lambda' \mathbf{F}'_{nk}(\theta^0) \mathbf{e}\} = \sigma^2 \lambda' \{E[n^{-1} \mathbf{F}'_{nk}(\theta^0) \mathbf{F}_{nk}(\theta^0)]\} \lambda$$

and

$$s_{nn}^2 = E[V_{nn}^2] = \sigma^2 \lambda' [E\{n^{-1} \mathbf{F}'_{nk}(\boldsymbol{\theta}^0) \mathbf{F}_{nk}(\boldsymbol{\theta}^0)\}] \lambda.$$

By assumption iii

$$\begin{aligned} \sigma^{-2}(V_{nn}^2 - s_{nn}^2) &= \lambda' [n^{-1} \mathbf{F}'_{nk}(\boldsymbol{\theta}^0) \mathbf{F}_{nk}(\boldsymbol{\theta}^0)] \lambda \\ &\quad - \lambda' [E\{n^{-1} \mathbf{F}'_{nk}(\boldsymbol{\theta}^0) \mathbf{F}_{nk}(\boldsymbol{\theta}^0)\}] \lambda \end{aligned}$$

converges to zero in probability,

$$s_{nn}^2 \rightarrow \sigma^2 \lambda' \mathbf{B}(\boldsymbol{\theta}^0) \lambda > 0,$$

and

$$(V_{nn}^2 - s_{nn}^2)s_{nn}^{-2} \xrightarrow{P} 0 \quad \text{as } n \rightarrow \infty.$$

Hence, condition i of Theorem 5.3.4 is satisfied. Now for any arbitrary  $\epsilon > 0$ ,

$$\begin{aligned} s_{nn}^{-2} \sum_{j=1}^n E\{Z_{jn}^2 I(|Z_{jn}| \geq \epsilon s_{nn})\} &\leq s_{nn}^{-2} \sum_{j=1}^n (\epsilon s_{nn})^{-\delta} E\{|Z_{jn}|^{2+\delta} I(|Z_{jn}| \geq \epsilon s_{nn})\} \\ &\leq \epsilon^{-\delta} s_{nn}^{-(2+\delta)} n^{-(1+\delta/2)} \\ &\quad \times \sum_{t=1}^n E\left\{ \left| \sum_{j=1}^k \lambda_j f_t^{(j)}(\boldsymbol{\theta}^0) e_t \right|^{2+\delta} I(|Z_{tn}| \geq \epsilon s_{nn}) \right\} \\ &\leq \epsilon^{-\delta} s_{nn}^{-(2+\delta)} n^{-(1+\delta/2)} \sum_{t=1}^n E\left\{ \left| \sum_{j=1}^k \lambda_j f_t^{(j)}(\boldsymbol{\theta}^0) e_t \right|^{2+\delta} \right\} \\ &\leq \epsilon^{-\delta} s_{nn}^{-(2+\delta)} n^{-(\delta/2)} M_F, \end{aligned}$$

where  $I(A)$  is the indicator function for the set  $A$ . Hence, condition ii of Theorem 5.3.4 is satisfied and

$$n^{-1/2} \mathbf{F}'_{nk}(\boldsymbol{\theta}^0) e_n \xrightarrow{\mathcal{L}} N[\mathbf{0}, \mathbf{B}(\boldsymbol{\theta}^0) \sigma^2]. \quad (5.5.16)$$

By (5.5.12)–(5.5.14), using Corollary 5.2.6.2, we have

$$n^{1/2} (\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} N[\mathbf{0}, \mathbf{B}^{-1}(\boldsymbol{\theta}^0) \sigma^2].$$

To obtain the limit of  $s^2$ , we note that

$$\begin{aligned} Q_n(\hat{\boldsymbol{\theta}}) &= Q_n(\boldsymbol{\theta}^0) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)' \frac{\partial Q_n(\boldsymbol{\theta}^0)}{\partial \boldsymbol{\theta}} + \frac{1}{2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)' \frac{\partial^2 Q_n(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \\ &= Q_n(\boldsymbol{\theta}^0) + (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0)' \mathbf{B}(\boldsymbol{\theta}^0) (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) + O_p(n^{-1}) \\ &= Q_n(\boldsymbol{\theta}^0) + O_p(n^{-1}), \end{aligned}$$

where  $\hat{\theta}$  is on the line segment joining  $\hat{\theta}$  and  $\theta^0$  and we have used (5.5.15). Therefore

$$s^2 = n(n-k)^{-1} Q_n(\hat{\theta}) = (n-k)^{-1} \sum_{t=1}^n e_t^2 + O_p(n^{-1}). \quad (5.5.17)$$

By assumption ii, the  $e_t$  have bounded  $2 + \delta$  conditional moment. Therefore, by Corollary 5.3.5,

$$\lim_{n \rightarrow \infty} (n-k)^{-1} \sum_{t=1}^n e_t^2 = \sigma^2 \quad \text{almost surely.}$$

Conclusion b follows by (5.5.17).  $\blacktriangle$

Because  $s^2$  is converging to  $\sigma^2$  and  $n^{-1} \mathbf{F}'_{nk}(\hat{\theta}) \mathbf{F}_{nk}(\hat{\theta})$  is converging to  $\mathbf{B}(\theta^0)$ , we can use

$$\hat{V}\{\hat{\theta}\} = [\mathbf{F}'_{nk}(\hat{\theta}) \mathbf{F}_{nk}(\hat{\theta})]^{-1} s^2$$

as an estimator of the variance of  $\hat{\theta}$ . The statistics computed as the usual “*t*-statistics” of regression will be distributed as  $N(0, 1)$  random variables in the limit.

There are situations in which the estimator  $\hat{\theta}_n$  may have a limiting distribution that is not normal. Also, for some functions  $Q_n(\theta)$ , the covariance matrix of the limiting distribution may not take the simple form of (5.5.11). Some of these situations are covered by Theorem 5.5.2.

**Theorem 5.5.2.** Suppose that:

- (i) The assumptions (5.5.6) and (5.5.7) of Lemma 5.5.2 hold.
- (ii) The first and second derivatives of  $Q_n(\theta)$  with respect to  $\theta$  are continuous for all  $n$  on a convex, compact neighborhood  $S$  containing  $\theta^0$  as an interior point.
- (iii)  $n^{1/2} \partial Q_n(\theta^0) / \partial \theta \xrightarrow{\mathcal{L}} 2X$ , where  $X$  represents the limiting distribution.
- (iv) There exists a  $k \times k$  fixed matrix function  $\mathbf{B}(\theta)$  such that  $\mathbf{B}(\theta)$  is continuous on  $S$ ,  $\mathbf{B}(\theta^0)$  is a symmetric, positive definite matrix, and

$$\lim_{n \rightarrow \infty} \left[ \frac{\partial^2 Q_n(\theta)}{\partial \theta \partial \theta'} - 2\mathbf{B}(\theta) \right] = \mathbf{0}$$

almost surely, uniformly on  $S$ .

Let  $\hat{\theta}_n$  be a measurable function that minimizes the objective function  $Q_n(\theta)$ . Then

$$n^{1/2} (\hat{\theta}_n - \theta^0) \xrightarrow{\mathcal{L}} -\mathbf{B}^{-1}(\theta^0) \mathbf{X}.$$

**Proof.** By assumption i, the estimator  $\hat{\theta}_n$  converges to  $\theta^0$  almost surely, and  $\hat{\theta}_n$  lies inside the convex, compact neighborhood  $S$  of  $\theta^0$  for large  $n$ , almost surely. Therefore, we can expand the partial derivative of  $Q_n(\hat{\theta}_n)$  with respect to  $\theta$  in a Taylor series about  $\theta^0$ , for large  $n$ , almost surely. For  $\ddot{\theta}_n$  on the line segment joining  $\hat{\theta}_n$  and  $\theta^0$  we have

$$\begin{aligned} \mathbf{0} &= n^{1/2} \frac{\partial Q_n(\hat{\theta}_n)}{\partial \theta} \\ &= n^{1/2} \frac{\partial Q_n(\theta^0)}{\partial \theta} + \frac{\partial^2 Q_n(\ddot{\theta}_n)}{\partial \theta \partial \theta'} n^{1/2} (\hat{\theta}_n - \theta^0) \end{aligned} \quad (5.5.18)$$

for large  $n$ , almost surely. By assumption iv, the second partial derivative of  $Q_n(\theta)$  evaluated at  $\ddot{\theta}_n$  is converging to  $2\mathbf{B}(\theta^0)$  a.s. Therefore, using assumption iii, (5.5.18), and Corollary 5.2.6.2, we have the conclusion.  $\blacktriangle$

If, in Theorem 5.5.2,  $\mathbf{X} \sim N[\mathbf{0}, \Sigma_{xx}]$ , then

$$n^{1/2}(\hat{\theta}_n - \theta^0) \xrightarrow{\mathcal{D}} N[\mathbf{0}, \mathbf{B}^{-1}(\theta^0)\Sigma_{xx}\mathbf{B}^{-1}(\theta^0)].$$

If  $\Sigma_{xx} = \mathbf{B}(\theta^0)\sigma^2$ , we obtain the result of Theorem 5.5.1 that the limiting distribution is  $N[\mathbf{0}, \mathbf{B}^{-1}(\theta^0)\sigma^2]$ .

In Theorem 5.5.1 and Theorem 5.5.2, it is  $n^{1/2}(\hat{\theta} - \theta^0)$  that converges in distribution to a nondegenerate random variable. In assumption iii of Theorem 5.5.1, the matrix of derivatives divided by  $n$  is assumed to converge to a nonsingular matrix. The normalization of the sums of squares by  $n$  and the corresponding normalization of  $\hat{\theta} - \theta^0$  with  $n^{1/2}$  is not always appropriate.

In time series analyses, it sometimes happens that the sum of squares of the derivative with respect to one parameter increases at a faster rate than the derivative with respect to a second parameter. Consider the simple model

$$Y_t = \beta_0 + \beta_1 t + e_t,$$

with vector of derivatives  $\mathbf{F}(\mathbf{x}_t; \boldsymbol{\beta}) = [1, t]$ . It follows that the sum of squares of the derivative with respect to  $\beta_0$  increases at the rate  $n$ , while the sum of squares of the derivative with respect to  $\beta_1$  increases at the rate  $n^3$ . Therefore, Theorem 5.5.1 could not be used to obtain the limiting distribution of the least squares estimator of the parameter vector. Following Sarkar (1990), we now give a theorem, Theorem 5.5.3, that is applicable to situations in which the sums of squares of the first derivatives increase at different rates. Standard results are obtained by setting  $M_n$  and  $a_n$  of the theorem equal to  $n^{1/2}\mathbf{I}$  and  $n^{1/2}$ , respectively.

The assumptions of Theorem 5.5.3 are local in that assumption d is a local property of  $Q_n(\theta)$  in the vicinity of  $\theta^0$ . This assumption permits the existence of other local minima. Thus, the conclusion of Theorem 5.5.3 is local in that the sequence of estimators associated with the local region of assumption d are consistent for  $\theta^0$ .

We assume that an estimator of  $\theta$  is constructed by minimizing the function  $Q_n(\theta)$ . Let the vector of first partial derivatives be

$$\frac{\partial Q_n(\theta)}{\partial \theta} = 2U_n(\theta)$$

and the matrix of second partial derivatives be

$$\frac{\partial^2 Q_n(\theta)}{\partial \theta \partial \theta'} = 2B_n(\theta).$$

If the minimum of  $Q_n(\theta)$  occurs in the interior of the parameter space, then the  $\theta$  that minimizes  $Q_n(\theta)$  is a solution of the system

$$U_n(\theta) = 0. \quad (5.5.19)$$

**Theorem 5.5.3.** Suppose the true value  $\theta^0$  is an interior point of the parameter space  $\Theta$ . Assume that there is a sequence of square matrices  $\{M_n\}$  and a sequence of real numbers  $\{a_n\}$  such that

- (a)  $\lim_{n \rightarrow \infty} M_n^{-1} B_n(\theta^0) M_n^{-1'} = B$  a.s., where  $B$  is a  $k \times k$  symmetric, positive definite matrix, a.s.
- (b)  $\lim_{n \rightarrow \infty} a_n^{-1} M_n^{-1} U_n(\theta^0) = 0$  a.s.
- (c)  $\lim_{n \rightarrow \infty} a_n = \infty$ .
- (d) There exists a  $\delta > 0$  and random variables  $L_n$  such that for all  $n$ , all  $\theta$  in  $S_{n\delta_0}$ , and all  $\delta_0$ ,  $0 < \delta_0 < \delta$ ,

$$\|M_n^{-1} B_n(\theta) M_n^{-1'} - M_n^{-1} B_n(\theta^0) M_n^{-1'}\| \leq L_n \delta_0 \quad \text{a.s.},$$

and

$$\limsup_{n \rightarrow \infty} L_n < \infty \quad \text{a.s.},$$

where

$$S_{n\delta_0} = \{\theta \in \Theta : \|a_n^{-1} M_n'(\theta - \theta^0)\| < \delta_0\}$$

and for any  $k \times r$  matrix  $C$ ,  $\|C\|^2 = \text{tr } CC'$  denotes the sum of squares of the elements of  $C$ .

Then there exists a sequence of roots of (5.5.19), denoted by  $\{\tilde{\theta}_n\}$ , such that

$$\lim_{n \rightarrow \infty} \|a_n^{-1} M_n'(\tilde{\theta}_n - \theta^0)\| = 0 \quad \text{a.s.}$$

**Proof.** Given  $\epsilon > 0$  and  $0 < \delta_0 \leq \delta$ , it is possible to define a set  $A$ , an  $n_0$ , and positive constants  $K_1$ ,  $K_2$ , and  $K_3$  such that  $P\{A\} \geq 1 - \epsilon$  and for all  $\omega$  in  $A$  and all

$n > n_0$ ,

$$a_n > K_1, \quad (5.5.20)$$

$$\gamma_{\min}[\mathbf{M}_n^{-1}\mathbf{B}_n(\boldsymbol{\theta}^0)\mathbf{M}_n^{-1'}] \geq 3K_2, \quad (5.5.21)$$

$$\begin{aligned} \|\mathbf{M}_n^{-1}\mathbf{B}_n(\boldsymbol{\theta}^0)\mathbf{M}_n^{-1'} - \mathbf{B}\|^2 &< K_2, \\ \|\mathbf{M}_n^{-1}[\mathbf{B}_n(\boldsymbol{\theta}) - \mathbf{B}_n(\boldsymbol{\theta}^0)]\mathbf{M}_n^{-1'}\| &< K_3\delta_0, \end{aligned} \quad (5.5.22)$$

and  $L_n(\omega) \leq K_3$  for all  $\boldsymbol{\theta}$  in  $S_{n\delta_0}$ , by assumptions c, a, a, b, and d, respectively, where  $\gamma_{\min}(\mathbf{C})$  is the smallest root of the matrix  $\mathbf{C}$ . Let  $\delta_0^* = \min(\delta_0, K_3^{-1}K_2)$ . Then, for  $\omega$  in  $A$ ,  $\boldsymbol{\theta}$  in  $S_{n\delta_0^*}$ , and  $n > n_0$ , we have

$$\|\mathbf{d}_n\| < K_2\delta_0^* \quad (5.5.23)$$

and

$$\|\mathbf{M}_n^{-1}[\mathbf{B}_n(\boldsymbol{\theta}) - \mathbf{B}_n(\boldsymbol{\theta}^0)]\mathbf{M}_n^{-1'}\| < K_2, \quad (5.5.24)$$

where  $\|\mathbf{d}_n\| = \|0.5a_n^{-1}\mathbf{M}_n^{-1}\mathbf{U}_n(\boldsymbol{\theta}^0)\|$ .

Let the Taylor expansion of  $Q_n(\boldsymbol{\theta})$  about  $\boldsymbol{\theta}^0$  be

$$\begin{aligned} Q_n(\boldsymbol{\theta}) &= Q_n(\boldsymbol{\theta}^0) + 0.5(\boldsymbol{\theta} - \boldsymbol{\theta}^0)' \mathbf{U}_n(\boldsymbol{\theta}^0) + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)' \mathbf{B}_n(\hat{\boldsymbol{\theta}}_n)(\boldsymbol{\theta} - \boldsymbol{\theta}^0) \\ &= Q_n(\boldsymbol{\theta}^0) + 0.5(\boldsymbol{\theta} - \boldsymbol{\theta}^0)' \mathbf{U}_n(\boldsymbol{\theta}^0) + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)' [\mathbf{B}_n(\hat{\boldsymbol{\theta}}_n) - \mathbf{B}_n(\boldsymbol{\theta}^0)](\boldsymbol{\theta} - \boldsymbol{\theta}^0) \\ &\quad + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)' \mathbf{B}_n(\boldsymbol{\theta}^0)(\boldsymbol{\theta} - \boldsymbol{\theta}^0), \end{aligned} \quad (5.5.25)$$

where  $\hat{\boldsymbol{\theta}}_n$  lies on the line segment joining  $\boldsymbol{\theta}$  and  $\boldsymbol{\theta}^0$ . Using (5.5.24),

$$(\boldsymbol{\theta} - \boldsymbol{\theta}^0)' \mathbf{M}_n^{-1} [\mathbf{B}_n(\hat{\boldsymbol{\theta}}_n) - \mathbf{B}_n(\boldsymbol{\theta}^0)] \mathbf{M}_n^{-1'} \mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0) \leq \|\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^2 K_2 \quad (5.5.26)$$

for all  $\boldsymbol{\theta}$  in  $S_{n\delta_0^*}$ . By (5.5.21),

$$(\boldsymbol{\theta} - \boldsymbol{\theta}^0)' \mathbf{B}_n(\boldsymbol{\theta}^0)(\boldsymbol{\theta} - \boldsymbol{\theta}^0) \geq 3K_2 \|\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^2. \quad (5.5.27)$$

Let  $R(S_{n\delta_0^*})$  be the boundary of  $S_{n\delta_0^*}$ ,

$$R(S_{n\delta_0^*}) = \{\boldsymbol{\theta} \in \Theta : \|a_n^{-1}\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^2 = \delta_0^{*2}\}.$$

From the expansion (5.5.25), using (5.5.26), (5.5.27), and (5.5.20),

$$Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}^0) \geq \|\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^2 [2K_2 + (\boldsymbol{\theta} - \boldsymbol{\theta}^0)' \|\mathbf{M}_n'(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^{-2} a_n \mathbf{M}_n \mathbf{d}_n] \quad (5.5.28)$$

and

$$\|\mathbf{M}'_n(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^2 = \|a_n^{-1}\mathbf{M}'_n(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^2 a_n^2 \geq \delta_0^{*2} K_1^2 \quad (5.5.29)$$

for all  $\omega$  in  $A$ , all  $n > n_0$ , and all  $\boldsymbol{\theta}$  in  $R(S_{n\delta_0^*})$ . Then, by (5.5.23) and (5.5.29),

$$\|\mathbf{M}'_n(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^{-2} \|a_n(\boldsymbol{\theta} - \boldsymbol{\theta}^0)' \mathbf{M}_n \mathbf{d}_n\| \leq a_n \|\mathbf{M}'_n(\boldsymbol{\theta} - \boldsymbol{\theta}^0)\|^{-1} \|\mathbf{d}_n\| \leq \delta_0^{*-1} (K_2 \delta_0^*)$$

and

$$Q_n(\boldsymbol{\theta}) - Q_n(\boldsymbol{\theta}^0) \geq \delta_0^{*2} K_1 K_2$$

for all  $\omega$  in  $A$ , all  $n > n_0$ , and all  $\boldsymbol{\theta}$  in  $R(S_{n\delta_0^*})$ . It follows that  $Q_n(\boldsymbol{\theta})$  must attain a minimum in the interior of  $S_{n\delta_0^*}$ , at which point the system (5.5.19) is satisfied. Therefore,

$$P\{\|a_n^{-1}\mathbf{M}'_n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)\| < \delta_0^*\} \geq 1 - \epsilon$$

for  $n > n_0$ .

Now, let  $A_j$ ,  $n_j$ , and  $\delta_j^*$  denote  $A$ ,  $n_0$ , and  $\delta_0^*$  associated with  $\epsilon_j = 2^{-j}$ . Since  $\delta_0$  is arbitrary, we can choose the  $\delta_j^*$  to be a sequence decreasing to zero and take  $n_j$  to be an increasing sequence. Define  $A = \liminf A_j$ . Then  $P(A) = 1$ . Given  $\eta > 0$  and a fixed  $\omega$  in  $A$ , there exists a  $j_0$  such that  $\omega$  is in  $A_j$  for all  $j > j_0$ . Choose  $j_1 > j_0$  such that  $\delta_{j_1} < \eta$ . Then, for  $j > j_1$  and  $n > n_j$ , we have

$$\|a_n^{-1}\mathbf{M}'_n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)\| < \delta_j < \eta. \quad \blacktriangle$$

**Corollary 5.5.3.1.** Let assumptions a through d of Theorem 5.5.3 hold in probability instead of almost surely. Then there exists a sequence of roots of (5.5.19), denoted by  $\{\tilde{\boldsymbol{\theta}}_n\}$ , such that

$$\plim_{n \rightarrow \infty} \|a_n^{-1}\mathbf{M}'_n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0)\| = 0.$$

**Proof.** Omitted. ▲

If the roots of  $a_n^{-2}\mathbf{M}_n \mathbf{M}'_n$  are bounded away from zero, then  $\tilde{\boldsymbol{\theta}}_n$  is consistent for  $\boldsymbol{\theta}^0$ . If  $\tilde{\boldsymbol{\theta}}_n$  is consistent, and if the properly normalized first and second derivatives converge in distribution, then the estimator of  $\boldsymbol{\theta}$  converges in distribution.

**Corollary 5.5.3.2.** Let assumptions a through d of Theorem 5.5.3 hold in probability. In addition, assume  $\tilde{\boldsymbol{\theta}}_n$  is consistent for  $\boldsymbol{\theta}^0$  and

$$(e) \{\mathbf{M}_n^{-1} \mathbf{U}_n(\boldsymbol{\theta}^0), \mathbf{M}_n^{-1} \mathbf{B}_n(\boldsymbol{\theta}^0) \mathbf{M}_n^{-1}\}' \xrightarrow{\mathcal{L}} (\mathbf{X}, \mathbf{B}) \text{ as } n \rightarrow \infty.$$

Then

$$\mathbf{M}'_n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} -\mathbf{B}^{-1} \mathbf{X} \quad \text{as } n \rightarrow \infty,$$

where  $\mathbf{B}$  is defined in assumption a and  $\mathbf{U}_n(\boldsymbol{\theta}^0)$  is the vector of first partial derivatives evaluated at  $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ .

**Proof.** Because  $\tilde{\boldsymbol{\theta}}_n$  is converging to  $\boldsymbol{\theta}^0$  in probability, there is a compact set  $S$  containing  $\boldsymbol{\theta}^0$  as an interior point such that, given  $\epsilon > 0$ , the probability that  $\tilde{\boldsymbol{\theta}}_n$  is in  $S$  is greater than  $1 - \epsilon$  for  $n$  greater than some  $N_\epsilon$ . For  $\tilde{\boldsymbol{\theta}}_n$  in  $S$ , we expand  $\mathbf{U}_n(\boldsymbol{\theta})$  in a first order Taylor series about  $\boldsymbol{\theta}^0$  and multiply the Taylor series expansion by  $\mathbf{M}_n^{-1}$  to obtain

$$\mathbf{0} = \mathbf{M}_n^{-1} \mathbf{U}_n(\tilde{\boldsymbol{\theta}}_n) = \mathbf{M}_n^{-1} \mathbf{U}_n(\boldsymbol{\theta}^0) + \mathbf{M}_n^{-1} \mathbf{B}_n(\tilde{\boldsymbol{\theta}}_n) \mathbf{M}_n^{-1} \mathbf{M}'_n(\tilde{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^0),$$

where  $\ddot{\boldsymbol{\theta}}$  is an intermediate point between  $\tilde{\boldsymbol{\theta}}_n$  and  $\boldsymbol{\theta}^0$ . Note that

$$\mathbf{M}_n^{-1} \mathbf{B}_n(\tilde{\boldsymbol{\theta}}_n) \mathbf{M}_n^{-1} = \mathbf{M}_n^{-1} \mathbf{B}_n(\boldsymbol{\theta}^0) \mathbf{M}_n^{-1} + \mathbf{M}_n^{-1} [\mathbf{B}_n(\tilde{\boldsymbol{\theta}}_n) - \mathbf{B}_n(\boldsymbol{\theta}^0)] \mathbf{M}_n^{-1},$$

and the conclusion follows by assumptions a, d, and e.  $\blacktriangleleft$

By Theorem 5.5.3, there is a sequence of roots of (5.5.19) that converges to  $\boldsymbol{\theta}^0$ . Hence, the assumption of the corollary that  $\tilde{\boldsymbol{\theta}}$  is consistent is equivalent to assuming that we are considering the consistent sequence of roots.

**Example 5.5.1.** Consider the model

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 x_{t1} + \beta_0 \beta_1 t + e_t \\ &= f[(1, x_{t1}, t), (\beta_0, \beta_1)] + e_t, \end{aligned} \quad (5.5.30)$$

where  $e_t \sim NI(0, \sigma^2)$ ,  $\beta_i \neq 0$ , and

$$x_{t1} = \begin{cases} 1 & \text{if } t \text{ is odd,} \\ -1 & \text{if } t \text{ is even.} \end{cases}$$

Assume that a sample of size  $n$  is available and that an estimator of  $(\beta_0, \beta_1)' = \boldsymbol{\beta}$  is constructed as the  $(\hat{\beta}_0, \hat{\beta}_1)$  that minimizes

$$Q_n(\boldsymbol{\beta}) = n^{-1} \sum_{t=1}^n (Y_t - \beta_0 - \beta_1 x_{t1} - \beta_0 \beta_1 t)^2.$$

The vector of first derivatives of  $f[(1, x_{t1}, t), (\beta_0, \beta_1)]$  is

$$\mathbf{F}_t(\boldsymbol{\beta}) = (1 + \beta_1 t, x_{t1} + \beta_0 t).$$

We note that  $n^{-1} \sum_{t=1}^n t^2$  increases at the rate  $n^2$ . Therefore, if  $\beta_0^0 \beta_1^0 \neq 0$ ,  $Q_n(\boldsymbol{\beta})$  will not converge for any  $(\beta_0, \beta_1)$  for which  $\beta_0 \beta_1 \neq \beta_0^0 \beta_1^0$ . Also,

$$\lim_{n \rightarrow \infty} n^{-3} \sum_{t=1}^n \mathbf{F}'_t(\boldsymbol{\beta}) \mathbf{F}_t(\boldsymbol{\beta})$$

is singular if  $\beta_0 \neq 0$  and  $\beta_1 \neq 0$ .

We now show how we can apply Theorem 5.5.3 to this problem. Let

$$\mathbf{M}_n^{-1} = \begin{pmatrix} 1 & -\beta_0^0 \\ 0 & \beta_1^0 \end{pmatrix} \begin{pmatrix} n^2 & 0 \\ 0 & 1 \end{pmatrix}^{-1/2}.$$

Then

$$n^{-1/2}(1 + \beta_1^0 t, x_{t1} + \beta_0^0 t) \mathbf{M}_n^{-1} = \{n^{-3/2}(1 + \beta_1^0 t), n^{-1/2}(\beta_1^0 x_{t1} - \beta_0^0)\}$$

and

$$\lim_{n \rightarrow \infty} \mathbf{M}_n^{-1} \mathbf{B}_n(\boldsymbol{\theta}^0) \mathbf{M}_n^{-1} = \begin{pmatrix} 3^{-1}(\beta_1^0)^2 & -2^{-1}\beta_0^0\beta_1^0 \\ -2^{-1}\beta_0^0\beta_1^0 & (\beta_0^0)^2 + (\beta_1^0)^2 \end{pmatrix}. \quad (5.5.31)$$

Setting  $a_n = n^{1/2}$ , we have

$$n^{-1/2} \sum_{t=1}^n [n^{-3/2}(1 + \beta_1^0 t)e_t, n^{-1/2}(\beta_1^0 x_{t1} - \beta_0^0)e_t] \rightarrow (0, 0) \quad \text{a.s.}$$

The second derivatives of  $Q_n(\boldsymbol{\beta})$  are continuous with continuous derivatives, and hence condition d of Theorem 5.5.3 also holds. It follows that the estimator of  $\boldsymbol{\beta}$  is consistent for  $\boldsymbol{\beta}$ . Using Theorem 5.3.4, one can show that

$$\mathbf{M}_n^{-1} \mathbf{U}_n(\boldsymbol{\beta}^0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{V}_{\boldsymbol{\beta}\boldsymbol{\beta}}),$$

where  $\mathbf{V}_{\boldsymbol{\beta}\boldsymbol{\beta}}$  is the matrix of (5.5.31) multiplied by  $\sigma^2$ .

An alternative method of defining a sequence of estimators with a nonsingular covariance matrix is to reparametrize the model. Assume that  $\beta_1 \neq 0$ . Let

$$(\alpha_0, \alpha_1) = (\beta_0 \beta_1, \beta_1),$$

from which

$$(\beta_0, \beta_1) = (\alpha_1^{-1} \alpha_0, \alpha_1),$$

and

$$Y_t = \alpha_1^{-1} \alpha_0 + \alpha_1 x_{t1} + \alpha_0 t + e_t.$$

The vector of first derivatives of the  $\boldsymbol{\alpha}$ -form of the model is

$$\mathbf{G}_t(\boldsymbol{\alpha}) = (\alpha_1^{-1} + t, -\alpha_1^{-2} \alpha_0 + x_{t1}).$$

It follows that

$$\lim_{n \rightarrow \infty} \mathbf{H}_n^{-1/2} \sum_{t=1}^n \mathbf{G}_t'(\boldsymbol{\alpha}) \mathbf{G}_t(\boldsymbol{\alpha}) \mathbf{H}_n^{-1/2} = \mathbf{A},$$

where

$$\mathbf{A} = \begin{bmatrix} \frac{1}{3} & -0.5\alpha_1^{-2}\alpha_0 \\ -0.5\alpha_1^{-2}\alpha_0 & 1 + \alpha_0^2\alpha_1^{-4} \end{bmatrix}$$

and  $\mathbf{H}_n = \text{diag}(n^3, n)$ . It can be shown that

$$\mathbf{H}_n^{-1/2} \sum_{i=1}^n \mathbf{G}_i(\boldsymbol{\alpha}) e_i \xrightarrow{\mathcal{L}} N[\mathbf{0}, \mathbf{A}\sigma^2].$$

Thus, the least squares estimator of  $\boldsymbol{\beta}$  is consistent and there is a normalizer that gives a limiting normal distribution. However, the two normalizers that we used are functions of the true, unknown parameters. The practitioner prefers a distribution based on a normalizer that is a function of the sample.

It is natural to consider a normalizer derived from the derivatives evaluated at the least squares estimator. For the model in the  $\boldsymbol{\alpha}$ -form, let

$$\mathbf{K}_n = \sum_{i=1}^n \mathbf{G}'_i(\hat{\boldsymbol{\alpha}}) \mathbf{G}_i(\hat{\boldsymbol{\alpha}}).$$

One can show that

$$\mathbf{H}_n^{-1/2} \mathbf{K}_n \mathbf{H}_n^{-1/2} \xrightarrow{P} \mathbf{A}$$

and hence

$$\mathbf{K}_n^{-1/2} (\hat{\boldsymbol{\alpha}} - \boldsymbol{\alpha}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}\sigma^2).$$

It follows that one can use the derivatives output by most nonlinear least squares programs to construct approximate tests and confidence intervals for the parameters of the model.  $\blacktriangleleft \blacktriangleleft$

**Example 5.5.2.** As an example where Theorem 5.5.3 does not hold, consider the model

$$Y_i = \beta + \beta^2 t + e_i,$$

where  $F(x_i; \beta) = (1 + 2\beta t)$  and  $e_i \sim NI(0, \sigma^2)$ . Assume that the model is to be estimated by least squares by choosing  $\beta$  to minimize

$$Q_n = n^{-1} \sum_{i=1}^n (Y_i - \beta - \beta^2 t)^2.$$

Then

$$U_n(\beta^0) = -n^{-1} \sum_{i=1}^n e_i (1 + 2\beta^0 t)$$

and

$$B_n(\beta^0) = n^{-1} \sum_{t=1}^n [(1 + 2\beta^0 t)^2 + 2te_t].$$

If  $\beta^0 = 0$ ,

$$B_n(\beta^0) = 1 + 2n^{-1} \sum_{t=1}^n te_t.$$

The quantity  $n^{-1} \sum_{t=1}^n te_t$  does not converge to zero in probability. If one divides  $B_n(\beta^0)$  by  $n^{1/2}$ , the normalized quantity converges in distribution, but not in probability. If one divides  $B_n(\beta^0)$  by a quantity increasing at a rate faster than  $n^{1/2}$ , the limit is zero. Therefore, the model with  $\beta^0 = 0$  does not satisfy assumption a of Theorem 5.5.3. However, one can use Lemma 5.5.1 to demonstrate that the least squares estimator is a consistent estimator of  $\beta^0$  when  $\beta^0 = 0$ .  $\blacktriangle\blacktriangle$

### 5.5.2. One-Step Estimation

An important special case of nonlinear estimation occurs when one is able to obtain an initial consistent estimator of the parameter vector  $\theta^0$ . In such cases, asymptotic results can be obtained for a single step or for a finite number of steps of a sequential minimization procedure.

We retain the model (5.5.1), which we write as

$$Y_t = f(\mathbf{x}_t; \theta^0) + e_t, \quad t = 1, 2, \dots, \quad (5.5.32)$$

where the  $e_t$  are iid( $0, \sigma^2$ ) random variables or are  $(0, \sigma^2)$  martingale differences satisfying condition ii of Theorem 5.5.1. Let  $\hat{\theta}$  be an initial consistent estimator of the unknown true value  $\theta^0$ . A Taylor's series expansion of  $f(\mathbf{x}_t; \theta^0)$  about  $\hat{\theta}$  gives

$$f(\mathbf{x}_t; \theta^0) = f(\mathbf{x}_t; \hat{\theta}) + \sum_{j=1}^k f^{(j)}(\mathbf{x}_t; \hat{\theta})(\theta_j^0 - \hat{\theta}_j) + d_t(\mathbf{x}_t; \hat{\theta}), \quad (5.5.33)$$

where  $\theta_j^0$  and  $\hat{\theta}_j$  are the  $j$ th elements of  $\theta^0$  and  $\hat{\theta}$ , respectively,

$$d_t(\mathbf{x}_t; \hat{\theta}) = \frac{1}{2} \sum_{r=1}^k \sum_{j=1}^k f^{(jr)}(\mathbf{x}_t; \hat{\theta}_r)(\theta_j^0 - \hat{\theta}_j)(\theta_r^0 - \hat{\theta}_r),$$

and  $\hat{\theta}_r$  is on the line segment joining  $\hat{\theta}$  and  $\theta^0$ . On the basis of equation (5.5.33), we consider the modified sum of squares

$$\begin{aligned} \hat{Q}(\delta) &= n^{-1} \sum_{t=1}^n \left\{ Y_t - f(\mathbf{x}_t; \hat{\theta}) - \sum_{j=1}^k f^{(j)}(\mathbf{x}_t; \hat{\theta})[\theta_j - \hat{\theta}_j] \right\}^2 \\ &= n^{-1} [\mathbf{w} - \mathbf{F}(\hat{\theta})\delta]' [\mathbf{w} - \mathbf{F}(\hat{\theta})\delta], \end{aligned}$$

where  $\delta = \theta - \hat{\theta}$ ,  $w$  is the  $n \times 1$  vector with  $t$ th element given by

$$w_t = Y_t - f(\mathbf{x}_t; \hat{\theta}),$$

and  $\mathbf{F}(\theta)$  is the  $n \times k$  matrix with  $t$ th row  $F_t(\theta)$  defined in (5.5.2). Minimizing  $\hat{Q}(\delta)$  with respect to  $\delta$  leads to

$$\tilde{\theta} = \hat{\theta} + \delta \quad (5.5.34)$$

as an estimator of  $\theta^0$ , where

$$\delta = [\mathbf{F}'(\hat{\theta})\mathbf{F}(\hat{\theta})]^{-1}\mathbf{F}'(\hat{\theta})w.$$

We call  $\tilde{\theta}$  the one-step Gauss–Newton estimator.

**Theorem 5.5.4.** Assume that the nonlinear model (5.5.32) holds. Let  $\tilde{\theta}$  be the one-step Gauss–Newton estimator defined in (5.5.34). Assume:

- (1) There is an open set  $S$  such that  $S$  is in  $\Theta$ ,  $\theta^0 \in S$ , and

$$\operatorname{plim}_{n \rightarrow \infty} n^{-1}\mathbf{F}'(\theta)\mathbf{F}(\theta) = \mathbf{B}(\theta)$$

is nonsingular for all  $\theta$  in  $S$ , where  $\mathbf{F}(\theta)$  is the  $n \times k$  matrix with  $t$ th element given by  $f^{(t)}(\mathbf{x}_t; \theta)$ .

- (2)  $\operatorname{plim}_{n \rightarrow \infty} n^{-1}\mathbf{G}'(\theta)\mathbf{G}(\theta) = \mathbf{L}(\theta)$  uniformly in  $\theta$  on the closure  $\bar{S}$  of  $S$ , where the elements of  $\mathbf{L}(\theta)$  are continuous functions of  $\theta$  on  $\bar{S}$ , and  $\mathbf{G}(\theta)$  is an  $n \times (1 + k + k^2 + k^3)$  matrix with  $t$ th row given by

$$\begin{aligned} & [f(\mathbf{x}_t; \theta), f^{(1)}(\mathbf{x}_t; \theta), f^{(2)}(\mathbf{x}_t; \theta), \dots, f^{(k)}(\mathbf{x}_t; \theta), \\ & f^{(11)}(\mathbf{x}_t; \theta), f^{(12)}(\mathbf{x}_t; \theta), \dots, f^{(1k)}(\mathbf{x}_t; \theta), f^{(21)}(\mathbf{x}_t; \theta), \dots, f^{(2k)}(\mathbf{x}_t; \theta), \\ & f^{(111)}(\mathbf{x}_t; \theta), \dots, f^{(kkk)}(\mathbf{x}_t; \theta)]. \end{aligned}$$

- (3) The initial estimator of  $\theta^0$ , say  $\hat{\theta}$ , satisfies  $(\hat{\theta} - \theta^0) = O_p(a_n)$ , where  $\lim_{n \rightarrow \infty} a_n = 0$ .

Then

$$\tilde{\theta} - \theta^0 = [\mathbf{F}'(\theta^0)\mathbf{F}(\theta^0)]^{-1}\mathbf{F}'(\theta^0)\mathbf{e} + O_p(\max\{a_n^2, a_n n^{-1/2}\}).$$

Furthermore, if  $n^{-1/2} \sum_{t=1}^n \mathbf{F}'_t(\theta^0)\mathbf{e}_t \xrightarrow{\mathcal{L}} N[\mathbf{0}, \mathbf{B}(\theta^0)\sigma^2]$  and  $a_n^2 = o(n^{-1/2})$ , then

$$n^{1/2}(\tilde{\theta} - \theta^0) \xrightarrow{\mathcal{L}} N[\mathbf{0}, \mathbf{B}^{-1}(\theta^0)\sigma^2].$$

**Proof.** Because  $\tilde{\theta} - \theta^0$  is  $O_p(a_n)$ , given  $\epsilon_0 > 0$ , one can choose  $N_0$  such that

the probability is greater than  $1 - \epsilon_0$  that  $\hat{\theta}$  is in  $S$  for all  $n > N_0$ . On the basis of a Taylor expansion, we can write, for  $\hat{\theta}$  in  $S$ ,

$$\begin{aligned}\delta &= [\mathbf{F}'(\hat{\theta})\mathbf{F}(\hat{\theta})]^{-1}\mathbf{F}'(\hat{\theta})[\mathbf{f}(\theta^0) - \mathbf{f}(\hat{\theta}) + \mathbf{e}] \\ &= [\mathbf{F}'(\hat{\theta})\mathbf{F}(\hat{\theta})]^{-1}\mathbf{F}'(\hat{\theta})[\mathbf{F}(\hat{\theta})\delta^0 + \mathbf{e}] \\ &\quad + [\mathbf{F}'(\hat{\theta})\mathbf{F}(\hat{\theta})]^{-1}\mathbf{R}(\hat{\theta}),\end{aligned}$$

where  $\mathbf{f}(\theta)$  is the  $n \times 1$  vector with  $r$ th element  $f(\mathbf{x}_r; \theta)$ ,  $\mathbf{e}$  is the  $n \times 1$  vector with  $r$ th element  $e_r = \theta^0_r - \hat{\theta}_r$ , and the  $j$ th element of  $n^{-1}\mathbf{R}(\hat{\theta})$  is

$$(2n)^{-1} \sum_{r=1}^k \sum_{s=1}^k \sum_{t=1}^n f^{(j)}(\mathbf{x}_r; \hat{\theta}) f^{(rs)}(\mathbf{x}_r; \hat{\theta}) [\theta_s^0 - \hat{\theta}_s] [\theta_t^0 - \hat{\theta}_t],$$

in which  $\hat{\theta}$  is on the line segment joining  $\hat{\theta}$  and  $\theta^0$ . The elements of  $\mathbf{L}(\theta)$  are bounded on  $\bar{S}$ , and, given  $\epsilon_1 > 0$ , there exists an  $N_1$  such that the probability is greater than  $1 - \epsilon_1$  that the elements of

$$n^{-1} \left\{ \sum_{r=1}^n [f^{(j)}(\mathbf{x}_r; \theta)]^2, \sum_{r=1}^n [f^{(rs)}(\mathbf{x}_r; \theta)]^2 \right\}$$

differ from the respective elements of  $\mathbf{L}(\theta)$  by less than  $\epsilon_1$  for all  $\theta \in \bar{S}$  and all  $n > N_1$ . Therefore,

$$(2n)^{-1} \sum_{r=1}^n f^{(j)}(\mathbf{x}_r; \hat{\theta}) f^{(rs)}(\mathbf{x}_r; \hat{\theta}) = O_p(1), \quad j, r, s = 1, 2, \dots, k,$$

and

$$n^{-1}\mathbf{R}(\hat{\theta}) = O_p(a_n^2). \quad (5.5.35)$$

Because  $\mathbf{B}(\theta)$  is nonsingular on  $S$ , there is an  $N_2$  such that  $\mathbf{F}'(\theta)\mathbf{F}(\theta)$  is nonsingular with probability greater than  $1 - \epsilon_2$  for all  $\theta$  in  $S$  and all  $n > N_2$ . It follows that

$$n^{-1}\mathbf{F}'(\hat{\theta})\mathbf{F}(\hat{\theta}) = n^{-1}\mathbf{F}'(\theta^0)\mathbf{F}(\theta^0) + O_p(a_n)$$

and

$$p\lim[n^{-1}\mathbf{F}'(\hat{\theta})\mathbf{F}(\hat{\theta})]^{-1} = \mathbf{B}^{-1}(\theta^0),$$

where the  $j$ th element of  $\mathbf{B}(\theta^0)$  is

$$\{\mathbf{B}(\theta^0)\}_{jj} = p\lim_{n \rightarrow \infty} \left[ n^{-1} \sum_{r=1}^n f^{(j)}(\mathbf{x}_r; \theta^0) f^{(s)}(\mathbf{x}_r; \theta^0) \right].$$

Therefore,

$$[\mathbf{F}'(\hat{\theta})\mathbf{F}(\hat{\theta})]^{-1}\mathbf{R}(\hat{\theta}) = O_p(a_n^2).$$

The  $j$ th element of  $n^{-1}\mathbf{F}'(\hat{\theta})\mathbf{e}$  is

$$\begin{aligned} n^{-1} \sum_{t=1}^n f^{(j)}(\mathbf{x}_t; \hat{\theta}) e_t &= n^{-1} \sum_{t=1}^n f^{(j)}(\mathbf{x}_t; \theta^0) e_t \\ &\quad + n^{-1} \sum_{t=1}^n \sum_{s=1}^k f^{(js)}(\mathbf{x}_t; \theta^0) [\hat{\theta}_s - \theta_s^0] e_t \\ &\quad + (2n)^{-1} \sum_{t=1}^n \sum_{s=1}^k \sum_{r=1}^k f^{(jsr)}(\mathbf{x}_t; \theta^0) [\hat{\theta}_s - \theta_s^0][\hat{\theta}_r - \theta_r^0] e_t, \end{aligned}$$

where  $\theta^0$  is on the line segment joining  $\hat{\theta}$  and  $\theta^0$ . By assumption 2,

$$n^{-1} \sum_{t=1}^n [f^{(jsr)}(\mathbf{x}_t; \theta^0) - f^{(jsr)}(\mathbf{x}_t; \theta^0)]^2 \xrightarrow{P} 0$$

as  $n \rightarrow \infty$ . Therefore,

$$n^{-1} \sum_{t=1}^n [f^{(jsr)}(\mathbf{x}_t; \theta^0) - f^{(jsr)}(\mathbf{x}_t; \theta^0)] e_t = o_p(1)$$

and

$$n^{-1} \sum_{t=1}^n f^{(j)}(\mathbf{x}_t; \hat{\theta}) e_t = n^{-1} \sum_{t=1}^n f^{(j)}(\mathbf{x}_t; \theta^0) e_t + O_p(\max\{a_n n^{-1/2}\}). \quad (5.5.36)$$

The results follow from (5.5.35) and (5.5.36). ▲

To estimate the variance of the limiting distribution of  $n^{1/2}(\hat{\theta} - \theta^0)$ , we must estimate  $\mathbf{B}^{-1}(\theta^0)$  and  $\sigma^2$ . By the assumptions,  $[n^{-1}\mathbf{F}'(\hat{\theta})\mathbf{F}(\hat{\theta})]^{-1}$  and  $[n^{-1}\mathbf{F}'(\hat{\theta})\mathbf{F}(\hat{\theta})]^{-1}$  are consistent estimators for  $\mathbf{B}^{-1}(\theta^0)$ . Also

$$s^2 = (n-k)^{-1} \sum_{t=1}^n [Y_t - f(\mathbf{x}_t; \hat{\theta})]^2 \quad (5.5.37)$$

is a consistent estimator of  $\sigma^2$ . Hence, using the matrix  $[\mathbf{F}'(\hat{\theta})\mathbf{F}(\hat{\theta})]^{-1}$  and the mean square  $s^2$ , all of the standard linear regression theory holds approximately for  $\hat{\theta}$ .

The theorem demonstrates that the one-step estimator is asymptotically unchanged by additional iteration if  $\hat{\theta} - \theta^0 = o_p(n^{-1/4})$ . For small samples, we may choose to iterate the procedure. For a particular sample we are not guaranteed that iteration of (5.5.34), using  $\tilde{\theta}$  of the previous step as the initial estimator, will

converge. Therefore, if one iterates, the estimator  $\hat{\theta}$  should be replaced at each step by

$$\tilde{\theta}_v = \hat{\theta} + v\tilde{\delta},$$

where  $\hat{\theta}$  is the estimator of the previous step and  $v \in (0, 1]$  is chosen so that  $n^{-1} \sum_{t=1}^n [Y_t - f(\mathbf{x}_t; \tilde{\theta}_v)]^2$  is less than  $n^{-1} \sum_{t=1}^n [Y_t - f(\mathbf{x}_t; \hat{\theta})]^2$ , and so that  $\tilde{\theta}_v \in \Theta$ . This iteration furnishes a method of obtaining the least squares estimator that minimizes (5.5.3).

**Example 5.5.3.** To illustrate the Gauss–Newton procedure, consider the model

$$Y_t = \theta_0 + \theta_1 x_{t1} + \theta_1^2 x_{t2} + e_t, \quad (5.5.38)$$

where the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables. While the superscript 0 was used on  $\theta$  in our derivation to identify the true parameter value, it is not a common practice to use that notation when discussing a particular application of the procedure. Observations generated by this model are given in Table 5.5.1. To obtain initial estimators of  $\theta_0$  and  $\theta_1$ , we ignore the nonlinear restriction and regress  $Y_t$  on  $x_{t0} = 1$ ,  $x_{t1}$ , and  $x_{t2}$ . This gives the regression equation

$$\hat{Y}_t = 0.877 + 1.262x_{t1} + 1.150x_{t2}.$$

Assuming that  $n^{-1} \sum x_{t1}^2$ ,  $n^{-1} \sum x_{t2}^2$ , and  $n^{-1} \sum x_{t1}x_{t2}$  converge to form a positive definite matrix, the coefficients for  $x_{t0}$  and  $x_{t1}$  are estimators for  $\theta_0$  and  $\theta_1$ , respectively, with errors  $O_p(n^{-1/2})$ . We note that  $(1.150)^{1/2}$  is also a consistent estimate of  $\theta_1$ . Using 0.877 and 1.262 as initial estimators, we compute

$$w_t = Y_t - 0.877 - 1.262x_{t1} - 1.593x_{t2}.$$

The rows of  $\mathbf{F}(\hat{\theta})$  are given by

$$\mathbf{F}_t(\hat{\theta}) = (1, x_{t1} + 2\hat{\theta}_1 x_{t2}), \quad t = 1, 2, \dots, 6.$$

**Table 5.5.1. Data and Regression Variables Used in the Estimation of the Parameters of the Model (5.5.38)**

$t$	$Y_t$	$x_{t0}$	$x_{t1}$	$x_{t2}$	$w_t$	$x_{t1} + 2\hat{\theta}_1 x_{t2}$
1	9	1	2	4	-0.777	12.100
2	19	1	7	8	-3.464	27.199
3	11	1	1	9	-5.484	23.724
4	14	1	3	7	-1.821	20.674
5	9	1	7	0	-0.714	7.000
6	3	1	0	2	-1.065	5.050

Regressing  $w_i$  on  $x_{i0}$  and  $x_{i1} + 2\hat{\theta}_1 x_{i2}$ , we obtain

$$\delta = \begin{pmatrix} 0.386 \\ -0.163 \end{pmatrix},$$

$$\tilde{\theta} = \begin{pmatrix} 0.877 \\ 1.262 \end{pmatrix} + \begin{pmatrix} 0.386 \\ -0.163 \end{pmatrix} = \begin{pmatrix} 1.263 \\ 1.099 \end{pmatrix}$$

as the one-step Gauss-Newton estimate.

Our computations illustrate the nature of the derivatives that enter the covariance matrix of the asymptotic distribution. In practice, we would use one of the several nonlinear least squares programs to estimate the parameters. These programs often have the option of specifying the initial values for the iteration or permitting the program to use a search technique. In this example, we have excellent start values. If we use the start values  $(0.877, 1.262)$ , the nonlinear least squares estimate is  $\hat{\theta} = (1.2413, 1.0913)'$ , the estimated covariance matrix is

$$\hat{V}\{\hat{\theta}\} = \begin{pmatrix} 1.3756 & -0.0763 \\ -0.0763 & 0.00054 \end{pmatrix},$$

and the residual mean square is 1.7346. We used procedure NLIN of SAS® [SAS (1989)] for the computations. ▲▲

## 5.6. INSTRUMENTAL VARIABLES

In many applications, estimators of the parameters of an equation of the regression type are desired, but the classical assumption that the matrix of explanatory variables is fixed is violated. Some of the columns of the matrix of explanatory variables may be measured with error and (or) may be generated by a stochastic mechanism such that the assumption that the error in the equation is independent of the explanatory variables becomes suspect.

Let us assume that we have the model

$$y = \Phi\beta + X\lambda + z, \quad (5.6.1)$$

where  $\beta$  is a  $k_1 \times 1$  vector of unknown parameters,  $\lambda$  is a  $k_2 \times 1$  vector of unknown parameters,  $y$  is an  $n \times 1$  vector,  $\Phi$  is an  $n \times k_1$  matrix,  $X$  is an  $n \times k_2$  matrix, and  $z$  is an  $n \times 1$  vector of unknown random variables with zero mean. The matrix  $\Phi$  is fixed, but the elements of  $X$  may contain a random component that is correlated with  $z$ .

Estimators obtained by ordinary least squares may be seriously biased because of the correlation between  $z$  and  $X$ . If information is available on variables that do not enter the equation, it may be possible to use these variables to obtain consistent estimators of the parameters. Such variables are called *instrumental variables*. The instrumental variables must be correlated with the variables entering the matrix  $X$  but not with the random components of the model. Assume that we have observations on  $k_3$  instrumental variables,  $k_3 \geq k_2$ . We denote the  $n \times k_3$  matrix of

observations on the instrumental variables by  $\psi$  and assume the elements of  $\psi$  are fixed. We express  $X$  as a linear combination of  $\Phi$  and  $\psi$ :

$$\begin{aligned} X &= \Phi\delta_1 + \psi\delta_2 + w \\ &= (\Phi : \psi)\delta + w, \end{aligned} \quad (5.6.2)$$

where

$$\delta = \begin{pmatrix} \delta_1 \\ \delta_2 \end{pmatrix} = \begin{pmatrix} \Phi' \Phi & \Phi' \psi \\ \psi' \Phi & \psi' \psi \end{pmatrix}^{-1} \begin{pmatrix} E\{\Phi' X\} \\ E\{\psi' X\} \end{pmatrix}.$$

Note that the residuals  $w$  follow from the definition of  $\delta$ . Therefore,  $w$  may be a sum of random and fixed components, but the fixed component is orthogonal to  $\Phi$  and  $\psi$  by construction.

The instrumental variable estimators we consider are obtained by regressing  $X$  on  $\Phi$  and  $\psi$ , computing the estimated values  $\hat{X}$  from this regression, and then replacing  $X$  by  $\hat{X}$  in the regression equation

$$y = \Phi\beta + X\lambda + z.$$

The instrumental variable estimators of  $\beta$  and  $\lambda$  are given by the regression of  $y$  on  $\Phi$  and  $\hat{X}$ :

$$\hat{\theta} = \begin{pmatrix} \hat{\beta} \\ \hat{\lambda} \end{pmatrix} = \begin{pmatrix} \Phi' \Phi & \Phi' \hat{X} \\ \hat{X}' \Phi & \hat{X}' \hat{X} \end{pmatrix}^{-1} \begin{pmatrix} \Phi' y \\ \hat{X}' y \end{pmatrix}, \quad (5.6.3)$$

where

$$\hat{X} = (\Phi : \psi)[(\Phi : \psi)'(\Phi : \psi)]^{-1}(\Phi : \psi)'X.$$

These estimators are called two-stage least squares estimators in the econometrics literature. An alternative instrumental variable estimator is the limited information maximum likelihood estimator. See Johnston (1984) and Fuller (1987).

To investigate the properties of estimator (5.6.3), we assume:

1.  $\mathbf{Q}_n = \mathbf{D}_{1,n}^{-1}(\Phi : \psi)'(\Phi : \psi)\mathbf{D}_{1,n}^{-1}$  is a nonsingular matrix for all  $n > k_1 + k_3$ , and  $\lim_{n \rightarrow \infty} \mathbf{Q}_n = \mathbf{Q}$ , where  $\mathbf{Q}$  is nonsingular and  $\mathbf{D}_{1,n}$  is a diagonal matrix whose elements are the square roots of the diagonal elements of  $(\Phi : \psi)'(\Phi : \psi)$ .
2.  $\mathbf{M}_n$  is nonsingular for all  $n > k_1 + k_3$ , and

$$\lim_{n \rightarrow \infty} \mathbf{M}_n = \mathbf{M},$$

where  $\mathbf{M}$  is a positive definite matrix,

$$\begin{aligned}\mathbf{M}_n &= \mathbf{D}_{2n}^{-1}(\Phi : \bar{\mathbf{X}})'(\Phi : \bar{\mathbf{X}})\mathbf{D}_{2n}^{-1}, \\ \bar{\mathbf{X}} &= (\Phi : \psi)\hat{\boldsymbol{\delta}},\end{aligned}$$

and  $\mathbf{D}_{2n}$  is a diagonal matrix whose elements are the square roots of the diagonal elements of  $(\Phi : \bar{\mathbf{X}})'(\Phi : \bar{\mathbf{X}})$ .

3.  $\lim_{n \rightarrow \infty} \mathbf{R}_n = \mathbf{R}$ , where  $\mathbf{R}$  is finite and

$$\mathbf{R}_n = E\{\mathbf{D}_{2n}^{-1}[\Phi : \bar{\mathbf{X}}]'zz'[\Phi : \bar{\mathbf{X}}]\mathbf{D}_{2n}^{-1}\}.$$

4. (a)  $\lim_{n \rightarrow \infty} \mathbf{B}_n = \mathbf{B}$ , where  $\mathbf{B}$  is finite and

$$\mathbf{B}_n = E\{\mathbf{D}_{1n}^{-1}(\Phi : \psi)'zz'(\Phi : \psi)\mathbf{D}_{1n}^{-1}\}.$$

- (b)  $\lim_{n \rightarrow \infty} \mathbf{G}_{nij} = \mathbf{G}_{ij}$ ,  $i, j = 1, 2, \dots, k_1 + k_2$ , where  $\mathbf{G}_{ij}$  is finite,

$$\mathbf{G}_{nij} = E\{\mathbf{D}_{1n}^{-1}[\Phi : \psi]'w_{.i}w_{.j}'[\Phi : \psi]\mathbf{D}_{1n}^{-1}\},$$

and  $w_{.i}$  is the  $i$ th column of the matrix  $\mathbf{w}$ .

5. (a)  $\lim_{n \rightarrow \infty} d_{jnn} = \infty$ ,  $j = 1, 2, i = 1, 2, \dots, k_1 + k_{4-j}$ , where  $d_{jnn}$  is the  $i$ th diagonal element of  $\mathbf{D}_{jn}$ .

$$(b) \lim_{n \rightarrow \infty} \left( \sum_{t=1}^n \varphi_{it}^2 \right)^{-1} \varphi_{ni}^2 = 0, \quad i = 1, 2, \dots, k_1,$$

$$\lim_{n \rightarrow \infty} \left( \sum_{t=1}^n \psi_{it}^2 \right)^{-1} \psi_{nj}^2 = 0, \quad j = 1, 2, \dots, k_3,$$

where  $\varphi_{it}$  is the  $t$ th element of  $\Phi$  and  $\psi_{ij}$  is the  $tj$ th element of  $\psi$ .

**Theorem 5.6.1.** Let the model of (5.6.1) to (5.6.2) and assumptions 1 through 4 and 5(a) hold. Then

$$\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{M}_n^{-1}\mathbf{D}_{2n}^{-1}(\Phi : \bar{\mathbf{X}})'z + o_p(1),$$

where

$$\begin{aligned}\hat{\boldsymbol{\theta}} &= [(\Phi : \hat{\mathbf{X}})'(\Phi : \hat{\mathbf{X}})]^{-1}(\Phi : \hat{\mathbf{X}})'y, \\ \hat{\mathbf{X}} &= (\Phi : \psi)\hat{\boldsymbol{\delta}},\end{aligned}$$

and

$$\hat{\boldsymbol{\delta}} = [(\Phi : \psi)'(\Phi : \psi)]^{-1}(\Phi : \psi)'X.$$

**Proof.** Define  $\hat{\mathbf{M}}_n$  to be the matrix  $\mathbf{M}_n$  with  $\bar{\mathbf{X}}$  replaced by  $\hat{\mathbf{X}}$ . By assumption 4, the variance of  $\mathbf{D}_{1n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$  is of order one, and by assumption 3, the variance of  $\mathbf{D}_{2n}^{-1}(\Phi : \bar{\mathbf{X}})' \mathbf{z}$  is of order one. Thus,

$$\begin{aligned}\hat{\mathbf{M}}_n - \mathbf{M}_n &= \mathbf{D}_{2n}^{-1}(\Phi : \hat{\mathbf{X}})'(\Phi : \hat{\mathbf{X}})\mathbf{D}_{2n}^{-1} - \mathbf{D}_{2n}^{-1}(\Phi : \bar{\mathbf{X}})'(\Phi : \bar{\mathbf{X}})\mathbf{D}_{2n}^{-1} \\ &= \mathbf{D}_{2n}^{-1}(\Phi : \hat{\mathbf{X}})'[\mathbf{0} : (\Phi : \psi)\mathbf{D}_{1n}^{-1}\mathbf{D}_{1n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]\mathbf{D}_{2n}^{-1} \\ &\quad + \mathbf{D}_{2n}^{-1}[\mathbf{0} : (\Phi : \psi)\mathbf{D}_{1n}^{-1}\mathbf{D}_{1n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]'(\Phi : \hat{\mathbf{X}})\mathbf{D}_{2n}^{-1} \\ &\quad + \mathbf{D}_{2n}^{-1}[\mathbf{0} : (\Phi : \psi)\mathbf{D}_{1n}^{-1}\mathbf{D}_{1n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]'[\mathbf{0} : (\Phi : \psi)\mathbf{D}_{1n}^{-1}\mathbf{D}_{1n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})]\mathbf{D}_{2n}^{-1} \\ &= o_p(1).\end{aligned}$$

Similarly,

$$\mathbf{D}_{2n}^{-1}(\Phi : \hat{\mathbf{X}})' \mathbf{z} - \mathbf{D}_{2n}^{-1}(\Phi : \bar{\mathbf{X}})' \mathbf{z} = o_p(1).$$

Using

$$\begin{aligned}\mathbf{y} &= (\Phi : \mathbf{X})\boldsymbol{\theta} + \mathbf{z} \\ &= \Phi\boldsymbol{\beta} + (\hat{\mathbf{X}} + \hat{\mathbf{w}})\boldsymbol{\lambda} + \mathbf{z}\end{aligned}$$

and

$$(\Phi : \hat{\mathbf{X}})' \hat{\mathbf{w}} = \mathbf{0},$$

we have

$$\begin{aligned}\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) &= \hat{\mathbf{M}}_n^{-1}\mathbf{D}_{2n}^{-1}(\Phi : \hat{\mathbf{X}})' \mathbf{z} = [\mathbf{M}_n^{-1} + o_p(1)]\mathbf{D}_{2n}^{-1}(\Phi : \bar{\mathbf{X}})' \mathbf{z} + o_p(1) \\ &= \mathbf{M}_n^{-1}\mathbf{D}_{2n}^{-1}(\Phi : \bar{\mathbf{X}})' \mathbf{z} + o_p(1).\end{aligned}$$
▲

Since  $\lim_{n \rightarrow \infty} \mathbf{M}_n = \mathbf{M}$ , we can also write

$$\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \mathbf{M}^{-1}\mathbf{D}_{2n}^{-1}(\Phi : \bar{\mathbf{X}})' \mathbf{z} + o_p(1). \quad (5.6.4)$$

In many applications  $n^{-1/2}\mathbf{D}_{1n}$  and  $n^{-1/2}\mathbf{D}_{2n}$  have finite nonsingular limits. In these cases  $n^{1/2}$  can be used as the normalizing factor, and the remainder in (5.6.4) is  $O_p(n^{-1/2})$ .

It follows from Theorem 5.6.1 that the sampling behavior of  $\mathbf{D}_{2n}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$  is approximately that of  $\mathbf{M}_n^{-1}\mathbf{D}_{2n}^{-1}(\Phi : \bar{\mathbf{X}})' \mathbf{z}$ , which has variance  $\mathbf{M}_n^{-1}\mathbf{R}_n\mathbf{M}_n^{-1}$ , where  $\mathbf{R}_n$  was defined in assumption 3. In some situations it is reasonable to assume the elements of  $\mathbf{z}$  are independent  $(0, \sigma_z^2)$  random variables.

**Corollary 5.6.1.** Let the model (5.6.1) to (5.6.2) and assumptions 1 to 5 hold with the elements of  $\mathbf{z}$  independently distributed  $(0, \sigma_z^2)$  random variables with

$E\{Z_i^4\} = \eta\sigma_z^4$ . Then

$$\mathbf{D}_{2n}(\hat{\theta} - \theta) \xrightarrow{d} N(\mathbf{0}, \mathbf{M}\sigma_z^2).$$

Furthermore, a consistent estimator for  $\sigma_z^2$  is

$$s_z^2 = \frac{1}{n - k_1 - k_2} \hat{\mathbf{z}}' \hat{\mathbf{z}}, \quad (5.6.5)$$

where  $\hat{\mathbf{z}} = \mathbf{y} - (\Phi : \mathbf{X})\hat{\theta}$ .

**Proof.** A proof is not presented here. The normality result is a special case of Theorem 6.3.4 of Chapter 6. That  $s_z^2$  is a consistent estimator of  $\sigma_z^2$  can be demonstrated by the arguments of Theorem 9.8.3.  $\blacktriangle$

Our analysis treated  $\Phi$  and  $\psi$  as fixed. Theorem 5.6.1 will hold for  $\Phi$  and (or)  $\psi$  random, provided the second moments exist, the probability limits analogous to the limits of assumptions 1, 2, and 5 exist, and the error terms in  $\mathbf{y}$  and  $\mathbf{X}$  are independent of  $\Phi$  and  $\psi$ .

A discussion of instrumental variables particularly applicable when  $k_3$  is larger than  $k_2$  is given in Sargan (1958). A model where the method of instrumental variables is appropriate will be discussed in Chapter 9.

**Example 5.6.1.** To illustrate the method of instrumental variables, we use some data collected in an animal feeding experiment. Twenty-four lots of pigs were fed six different rations characterized by the percentage of protein in the ration. The remainder of the ration was primarily carbohydrate from corn. We simplify by calling this remainder corn. Twelve of the lots were weighed after two weeks, and twelve after four weeks. The logarithms of the gain and of the feed consumed are given in Table 5.6.1. We consider the model

$$G_i = \beta_0 + \beta_1 P_i + \beta_2 C_i + Z_i, \quad i = 1, 2, \dots, 24,$$

where  $G_i$  is the logarithm of gain,  $C_i$  is the logarithm of corn consumed,  $P_i$  is the logarithm of protein consumed, and  $Z_i$  is the random error for the  $i$ th lot.

It is clear that neither corn nor protein, but instead their ratio, is fixed by the experimental design. The observations on corn and protein for a particular ration are constrained to lie on a ray through the origin with slope corresponding to the ratio of the percentages of the two items in the ration. The logarithms of these observations will lie on parallel lines. Since  $C_i - P_i$  is fixed, we rewrite the model as

$$G_i = \beta_0 + (\beta_1 + \beta_2)P_i + \beta_2(C_i - P_i) + Z_i.$$

In terms of the notation of (5.6.1),  $G_i = Y_i$  and  $P_i = X_i$ . Candidates for  $\psi_i$  are functions of  $C_i - P_i$  and of time on feed. As one simple model for the protein

Table 5.6.1. Gain and Feed Consumed by 24 Lots of Pigs

Lot <i>i</i>	Time on Feed Weeks	Log Gain <i>G</i>	Log Corn <i>C</i>	Log Protein <i>P</i>	$\hat{P}$	$\hat{Z}$
1	2	4.477	5.366	3.465	3.601	-0.008
2	2	4.564	5.488	3.587	3.601	-0.042
3	2	4.673	5.462	3.647	3.682	0.035
4	2	4.736	5.598	3.783	3.682	-0.036
5	2	4.718	5.521	3.787	3.757	-0.033
6	2	4.868	5.580	3.846	3.757	0.059
7	2	4.754	5.516	3.858	3.828	-0.043
8	2	4.844	5.556	3.898	3.828	0.007
9	2	4.836	5.470	3.884	3.895	0.035
10	2	4.828	5.463	3.877	3.895	0.034
11	2	4.745	5.392	3.876	3.961	-0.026
12	2	4.852	5.457	3.941	3.961	0.017
13	4	5.384	6.300	4.399	4.453	-0.021
14	4	5.493	6.386	4.485	4.453	0.003
15	4	5.513	6.350	4.535	4.537	0.001
16	4	5.583	6.380	4.565	4.537	0.041
17	4	5.545	6.314	4.580	4.616	0.013
18	4	5.613	6.368	4.634	4.616	0.028
19	4	5.687	6.391	4.733	4.690	0.028
20	4	5.591	6.356	4.698	4.690	-0.033
21	4	5.591	6.288	4.702	4.760	-0.015
22	4	5.700	6.368	4.782	4.760	0.015
23	4	5.700	6.355	4.839	4.828	-0.019
24	4	5.656	6.332	4.816	4.828	-0.041

Source: Data courtesy of Research Department, Moorman Manufacturing Company. The data are a portion of a larger experiment conducted by the Moorman Manufacturing Company in 1974.

consumption we suggest

$$P_i = \delta_0 + \delta_1(C_i - P_i) + \delta_2 t_i + \delta_3(t_i - 3)(C_i - P_i) + W_i,$$

where  $t_i$  is the time on feed of the  $i$ th lot. The ordinary least squares estimate of the equation is

$$\hat{P}_i = 4.45 - 0.95(C_i - P_i) + 0.46 t_i - 0.02(t_i - 3)(C_i - P_i),$$

(0.48)	(0.09)	(0.15)	(0.09)
--------	--------	--------	--------

where the numbers in parentheses are the estimated standard errors of the regression coefficients. If the  $W_i$  are independent  $(0, \sigma^2)$  random variables, the usual regression assumptions are satisfied. The interaction term contributes very little to the regression, but the time coefficient is highly significant. This supports

assumption 2 because it suggests that the partial correlation between  $P_i$  and  $t_i$  after adjusting for  $C_i - P_i$  is not zero. This, in turn, implies that the matrix  $M_n$  is nonsingular. The  $\hat{P}$ -values for this regression are given in Table 5.6.1. Regressing  $G_i$  on  $\hat{P}_i$  and  $C_i - P_i$ , we obtain

$$\hat{G}_i = 0.49 + 0.98P_i + 0.31(C_i - P_i).$$

In this problem it is reasonable to treat the  $Z$ 's as independent random variables. We also assume that they have common variance. The estimated residuals are shown in the last column of Table 5.6.1. These must be computed directly as

$$G_i - 0.49 - 0.98P_i - 0.31(C_i - P_i).$$

The residuals obtained in the second round regression computations are  $G_i - 0.49 - 0.98\hat{P}_i - 0.31(C_i - P_i)$  and are inappropriate for the construction of variance estimates. From the  $\hat{Z}$ 's we obtain

$$s_z^2 = (21)^{-1} \sum_{i=1}^{24} \hat{Z}_i^2 = 0.0010.$$

The inverse of the matrix used in computing the estimates is

$$\begin{pmatrix} 14.73 & -1.32 & -5.37 \\ -1.32 & 0.23 & 0.21 \\ -5.37 & 0.21 & 2.62 \end{pmatrix},$$

and it follows that the estimated standard errors of the estimates are (0.121), (0.015), and (0.051), respectively.  $\blacktriangleleft \blacktriangleright$

## 5.7. ESTIMATED GENERALIZED LEAST SQUARES

In this section, we investigate estimation for models in which the covariance matrix of the error is estimated. Our treatment is restricted to linear models, but many of the results extend to nonlinear models.

Consider the linear model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}, \quad (5.7.1)$$

where  $\mathbf{Y}$  is an  $n \times 1$  vector,  $\mathbf{X}$  is an  $n \times k$  matrix, and  $\boldsymbol{\beta}$  is the  $k \times 1$  vector of unknown parameters. We assume

$$E\{(\mathbf{u}, \mathbf{u}\mathbf{u}') | \mathbf{X}\} = (\mathbf{0}, \mathbf{V}_{uu}), \quad (5.7.2)$$

where  $\mathbf{V}_{uu}$  is positive definite. In many of our applications,  $\mathbf{u}' = (u_1, u_2, \dots, u_n)$  will be a portion of a realization of a time series. For example, the time series may be a  $p$ th order autoregressive process.

For known  $\mathbf{V}_{uu}$ , the generalized least squares estimator of  $\beta$  is

$$\hat{\beta} = (\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X})^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{Y}, \quad (5.7.3)$$

where we have assumed  $\mathbf{V}_{uu}$  and  $\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X}$  to be nonsingular. The conditional variance of the estimator is

$$\mathbf{V}\{\hat{\beta} | \mathbf{X}\} = (\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X})^{-1}. \quad (5.7.4)$$

Often  $\mathbf{V}_{uu}$  is not known. We are interested in the situation where an estimator of  $\mathbf{V}_{uu}$ , denoted by  $\hat{\mathbf{V}}_{uu}$ , is used to construct an estimator of  $\beta$ . We define the estimated generalized least squares estimator by

$$\check{\beta} = (\mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{Y}, \quad (5.7.5)$$

where we assume  $\hat{\mathbf{V}}_{uu}$  and  $\mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{X}$  are nonsingular. The estimated generalized least squares estimator of  $\beta$  is consistent for  $\beta$  under mild conditions. In the theorem, we use a general normalizer matrix,  $\mathbf{M}'_n$ . A natural choice for  $\mathbf{M}'_n$  is  $\mathbf{G}_n^{1/2} = (\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X})^{1/2}$ .

**Theorem 5.7.1.** Let the model (5.7.1) and (5.7.2) hold. Assume there exists a sequence of estimators  $\hat{\mathbf{V}}_{uu}$  and a sequence of nonsingular matrices  $\{\mathbf{M}_n\}$  such that

$$\mathbf{M}'_n(\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X})^{-1}\mathbf{M}_n = O_p(1), \quad (5.7.6)$$

$$\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{u} = O_p(1), \quad (5.7.7)$$

$$\mathbf{M}_n^{-1}\mathbf{X}'(\hat{\mathbf{V}}_{uu}^{-1} - \mathbf{V}_{uu}^{-1})\mathbf{X}\mathbf{M}_n^{-1} = O_p(\xi_n), \quad (5.7.8)$$

$$\mathbf{M}_n^{-1}\mathbf{X}'(\hat{\mathbf{V}}_{uu}^{-1} - \mathbf{V}_{uu}^{-1})\mathbf{u} = O_p(\xi_n), \quad (5.7.9)$$

where  $\xi_n \rightarrow 0$  as  $n \rightarrow \infty$ . Then

$$\mathbf{M}'_n(\check{\beta} - \hat{\beta}) = O_p(\xi_n).$$

**Proof.** By the assumption (5.7.8),

$$\mathbf{M}'_n\hat{\mathbf{G}}_n^{-1}\mathbf{M}_n = \mathbf{M}'_n\mathbf{G}_n^{-1}\mathbf{M}_n + O_p(\xi_n), \quad (5.7.10)$$

where

$$(\mathbf{G}_n, \hat{\mathbf{G}}_n) = (\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X}, \mathbf{X}'\hat{\mathbf{V}}_{uu}^{-1}\mathbf{X}). \quad (5.7.11)$$

Therefore

$$\begin{aligned}\mathbf{M}'_n(\tilde{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}) &= \mathbf{M}'_n(\hat{\mathbf{G}}_n^{-1} - \mathbf{G}_n^{-1})\mathbf{M}_n\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{u} \\ &\quad + \mathbf{M}'_n\hat{\mathbf{G}}_n^{-1}\mathbf{M}_n\mathbf{M}_n^{-1}\mathbf{X}'(\hat{\mathbf{V}}_{uu}^{-1} - \mathbf{V}_{uu}^{-1})\mathbf{u} \\ &= O_p(\xi_n).\end{aligned}$$

▲

In working with models such as (5.7.1)–(5.7.2), it is often assumed that

$$\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}^{-1}\mathbf{X}\mathbf{M}_n^{-1} \xrightarrow{P} \mathbf{A}_0, \quad (5.7.12)$$

where  $\mathbf{A}_0$  is a fixed positive definite matrix. This assumption is sufficient for the condition (5.7.6). If there exists a sequence  $\{\mathbf{M}_n\}$  that satisfies the assumptions of Theorem 5.7.1, then any nonsingular  $\mathbf{H}_n$  such that  $\mathbf{G}_n = \mathbf{H}_n\mathbf{H}_n'$  also will satisfy the assumptions because (5.7.6) implies that  $\mathbf{H}_n^{-1}\mathbf{M}_n = O_p(1)$ .

Theorem 5.7.1 is given for a general normalizing matrix  $\mathbf{M}_n$ . If  $\mathbf{M}_n$  is chosen to be  $\mathbf{G}_n^{1/2}$ , where  $\mathbf{G}_n$  is defined in (5.7.11), then (5.7.6) and (5.7.7) follow directly.

Under the assumptions of Theorem 5.7.1, the limiting distribution of the normalized estimated generalized least squares estimator is the same as the limiting distribution of the normalized generalized least squares estimator constructed with known  $\mathbf{V}_{uu}$ , provided the limiting distribution exists. Note that the estimator  $\hat{\mathbf{V}}_{uu}$  can converge to  $\mathbf{V}_{uu}$  rather slowly.

**Corollary 5.7.1.1.** Let the assumptions (5.7.10)–(5.7.12) of Theorem 5.7.2 hold. In addition assume

$$\mathbf{M}_n^{-1}\mathbf{G}_n\mathbf{M}_n^{-1} \xrightarrow{P} \mathbf{A}_0, \quad (5.7.13)$$

where  $\mathbf{A}_0$  is a fixed positive definite matrix, and

$$\mathbf{M}'_n(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{A}_0^{-1}). \quad (5.7.14)$$

Then

$$\mathbf{M}'_n(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{A}_0^{-1}), \quad (5.7.15)$$

and

$$(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta})'\hat{\mathbf{G}}_n(\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} \chi^2(k), \quad (5.7.16)$$

where  $\hat{\mathbf{G}}_n$  is defined in (5.7.11) and  $\chi^2(k)$  is a chi-square random variable with  $k$  degrees of freedom.

**Proof.** The assumption (5.7.13) implies the assumption (5.7.6) of Theorem 5.7.1. Hence, (5.7.15) follows from the variance of the generalized least squares

estimator and (5.7.14). Also, (5.7.16) follows from

$$\begin{aligned} (\tilde{\beta} - \beta)' \hat{G}_n (\tilde{\beta} - \beta) &= (\tilde{\beta} - \beta)' M_n [M_n^{-1} \hat{G}_n M_n^{-1}]' M_n' (\tilde{\beta} - \beta) \\ &= [M_n' (\tilde{\beta} - \beta)]' A_0 [M_n' (\tilde{\beta} - \beta)] + o_p(1), \end{aligned}$$

where we have used (5.7.13), (5.7.4), and the assumption (5.7.8).  $\blacktriangle$

The conditions (5.7.6) and (5.7.7) of Theorem 5.7.1 hold if  $M_n'$  is chosen equal to  $G_n^{1/2}$ . However, in practice, one wishes a matrix  $M_n'$  that is a function of the data. In some situations, there is a known transformation such that the estimated parameters of the transformed model can be normalized with a diagonal matrix. The transformation can be a function of  $n$ . In Corollary 5.7.1.2, we demonstrate how the estimator in such a case can be normalized using sample statistics.

**Corollary 5.7.1.2.** Let the assumptions of Theorem 5.7.1 hold with  $M_n = D_n^{1/2}$ , where  $D_n = \text{diag}\{X' V_{uu}^{-1} X\}$ . Also assume (5.7.14) holds with  $M_n = D_n^{1/2}$ . Then

$$\hat{D}_n^{1/2} (\tilde{\beta} - \beta) \xrightarrow{\mathcal{L}} N(\mathbf{0}, A_0^{-1}),$$

where  $\hat{D}_n = \text{diag}\{X' \hat{V}_{uu}^{-1} X\}$ , and

$$\hat{A}^{1/2} \hat{D}_n^{1/2} (\tilde{\beta} - \beta) \xrightarrow{\mathcal{L}} N(\mathbf{0}, I), \quad (5.7.17)$$

where  $\hat{A} = \hat{D}_n^{-1/2} X' \hat{V}_{uu}^{-1} X \hat{D}_n^{-1/2}$ .

**Proof.** By assumption (5.7.8) of Theorem 5.7.1,

$$\text{diag}\{D_n^{-1/2} X' \hat{V}_{uu}^{-1} X D_n^{-1/2}\} = D_n^{-1} \hat{D}_n = I + O_p(\xi_n). \quad (5.7.18)$$

Thus,

$$\hat{D}_n^{1/2} (\tilde{\beta} - \beta) = D_n^{1/2} (\hat{\beta} - \beta) + O_p(\xi_n)$$

and the distribution result is established. By (5.7.18) and the assumptions (5.7.6) and (5.7.8) of Theorem 5.7.1,

$$\hat{A} = \hat{D}_n^{-1/2} D_n^{1/2} D_n^{-1/2} X' \hat{V}_{uu}^{-1} X D_n^{-1/2} D_n^{1/2} \hat{D}_n^{-1/2} \xrightarrow{P} A_0$$

and  $\hat{A}^{1/2} \xrightarrow{P} A_0^{1/2}$ .  $\blacktriangle$

Given the conditions of Theorem 5.7.1 and Corollary 5.7.1.1, the estimated variance of the estimated generalized least squares estimator can be used to construct pivotal statistics. We state the result in Corollary 5.7.1.3.

**Corollary 5.7.1.3.** Let the assumptions of Theorem 5.7.1 and Corollary 5.7.1.1 hold, where  $\{\mathbf{M}_n\}$  is a sequence of fixed matrices. Then

$$\hat{\sigma}_{n\lambda}^{-1} \boldsymbol{\lambda}'_n (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(0, 1),$$

where  $\hat{\sigma}_{n\lambda}^2 = \boldsymbol{\lambda}'_n \hat{\mathbf{G}}_n^{-1} \boldsymbol{\lambda}_n$ ,  $\hat{\mathbf{G}}_n$  is defined in (5.7.12), and  $\boldsymbol{\lambda}_n$  is a fixed nonzero vector that can depend on  $n$ .

**Proof.** Under the assumptions, we show that

$$\sigma_{n\lambda}^{-1} \boldsymbol{\lambda}'_n (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(0, 1),$$

where

$$\sigma_{n\lambda}^2 = \boldsymbol{\lambda}'_n \mathbf{M}_n^{-1} \mathbf{A}_0^{-1} \mathbf{M}_n^{-1} \boldsymbol{\lambda}_n.$$

By Corollary 5.7.1.1, we have

$$\mathbf{L}_n = \mathbf{M}'_n (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{A}_0^{-1}).$$

By Skorohod's theorem [see Theorem 2.9.6 of Billingsley (1979)], there exist random vectors  $\mathbf{Z}_n$  and  $\mathbf{Z}$  on a common probability space such that  $\mathbf{Z}_n$  has the same distribution as  $\mathbf{L}_n$ ,  $\mathbf{Z} \sim N(\mathbf{0}, \mathbf{A}_0^{-1})$ , and  $\lim \mathbf{Z}_n = \mathbf{Z}$  a.s. Therefore,

$$\sigma_{n\lambda}^{-1} \boldsymbol{\lambda}'_n \mathbf{M}_n^{-1} \mathbf{Z}_n = \sigma_{n\lambda}^{-1} \boldsymbol{\lambda}'_n \mathbf{M}_n^{-1} \mathbf{Z} + \sigma_{n\lambda}^{-1} \boldsymbol{\lambda}'_n \mathbf{M}_n^{-1} (\mathbf{Z}_n - \mathbf{Z}).$$

Since  $\sigma_{n\lambda}^{-2} \boldsymbol{\lambda}'_n \mathbf{M}_n^{-1} \mathbf{A}_0^{-1} \mathbf{M}_n^{-1} \boldsymbol{\lambda}_n = 1$ , we have

$$\sigma_{n\lambda}^{-1} \boldsymbol{\lambda}'_n \mathbf{M}_n^{-1} \mathbf{Z} \sim N(0, 1)$$

and

$$\begin{aligned} |\sigma_{n\lambda}^{-1} \boldsymbol{\lambda}'_n \mathbf{M}_n^{-1} \mathbf{A}_0^{-1/2} \mathbf{A}_0^{1/2} (\mathbf{Z}_n - \mathbf{Z})|^2 &\leq (\mathbf{Z}_n - \mathbf{Z})' \mathbf{A}_0 (\mathbf{Z}_n - \mathbf{Z}) \\ &\leq \zeta_{\max}(\mathbf{A}_0) (\mathbf{Z}_n - \mathbf{Z})' (\mathbf{Z}_n - \mathbf{Z}) \\ &\rightarrow 0 \quad \text{a.s.}, \end{aligned}$$

where  $\zeta_{\max}(\mathbf{A}_0)$  is the maximum eigenvalue of  $\mathbf{A}_0$ . Therefore,  $\sigma_{n\lambda}^{-1} \boldsymbol{\lambda}'_n \mathbf{M}_n^{-1} \mathbf{Z}_n$  and hence  $\sigma_{n\lambda}^{-1} \boldsymbol{\lambda}'_n \mathbf{M}_n^{-1} \mathbf{L}_n$  converges in distribution to the standard normal distribution. See Sanger (1992, Section 7.1.3).

From the assumption (5.7.8) of Theorem 5.7.1 and the assumption (5.7.13) of Corollary 5.7.1.1,

$$\begin{aligned} \mathbf{M}_n^{-1} \hat{\mathbf{G}}_n \mathbf{M}_n^{-1} &= \mathbf{M}_n^{-1} \mathbf{G}_n \mathbf{M}_n^{-1} + O_p(\xi_n) \\ &= \mathbf{A}_0 + o_p(1), \end{aligned}$$

and  $\mathbf{M}'_n \hat{\mathbf{G}}_n^{-1} \mathbf{M}_n = \mathbf{A}_0^{-1} + \Delta_n$ , where  $\Delta_n = o_p(1)$ . Let  $\delta_n = \mathbf{M}_n^{-1} \lambda_n$ . Then

$$\begin{aligned} \left| \frac{\hat{\sigma}_{n\lambda}^2}{\sigma_{n\lambda}^2} - 1 \right| &= \left| \frac{\lambda_n' \mathbf{M}_n^{-1} \mathbf{M}'_n \hat{\mathbf{G}}_n^{-1} \mathbf{M}_n \mathbf{M}_n^{-1} \lambda_n}{\lambda_n' \mathbf{M}_n^{-1} \mathbf{A}_0^{-1} \mathbf{M}_n^{-1} \lambda_n} - 1 \right| \\ &= \left| (\delta_n' \mathbf{A}_0^{-1} \delta_n)^{-1} \sum_{i=1}^k \sum_{j=1}^k \delta_{ni} \delta_{nj} \Delta_{nij} \right| \\ &\leq [\delta_n' \delta_n \zeta_{\max}^{-1}(\mathbf{A}_0)]^{-1} \delta_{n,\max}^2 \sum_{i=1}^k \sum_{j=1}^k |\Delta_{nij}| \\ &= o_p(1), \end{aligned}$$

where  $\delta_{n,\max}$  is the maximum element of  $\delta_n$ . Therefore  $\hat{\sigma}_{n\lambda}^{-1} \sigma_{n\lambda}$  converges to one in probability, and the result follows.  $\blacktriangle$

Under the conditions of Theorem 5.7.1, the asymptotic distribution of the estimated generalized least squares estimator is the same as that of the generalized least squares estimator, provided the limiting distribution of the generalized least squares estimator exists. The following theorem gives conditions such that the generalized least squares estimator is asymptotically normal. To obtain a limiting normal distribution, the transformed errors must satisfy a central limit theorem. In Theorem 5.7.2, we assume the transformed errors to be independent. For example, if  $\{u_t\}$  is an autoregressive process with independent increments, then the assumption is satisfied.

**Theorem 5.7.2.** Assume the model (5.7.1) and (5.7.2) with fixed  $\mathbf{X}$ . Assume there exists a sequence of fixed, nonsingular matrices  $\{\mathbf{M}_n\}$  such that

$$\lim_{n \rightarrow \infty} \mathbf{M}_n^{-1} \mathbf{X}' \mathbf{V}_{uu}^{-1} \mathbf{X} \mathbf{M}_n^{-1} = \mathbf{A}_0, \quad (5.7.19)$$

where  $\mathbf{A}_0$  is positive definite. Assume there exists a sequence of nonsingular transformations  $\{\mathbf{T}_n\}$  such that the elements of  $\mathbf{e} = \mathbf{T}_n \mathbf{u}$  are independent with zero expectation, variance one, and

$$E\{|e_m|^{2+\delta}\} < K$$

for some  $\delta > 0$  and finite  $K$ . Also assume that

$$\lim_{n \rightarrow \infty} \sup_{\substack{1 \leq j \leq n \\ 1 \leq i \leq p}} |(\mathbf{M}_n^{-1} \mathbf{X}' \mathbf{T}_n)'_{ij}| = 0. \quad (5.7.20)$$

Then

$$\mathbf{M}'_n (\hat{\beta} - \beta) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{A}_0^{-1}),$$

where  $\hat{\beta}$  is the generalized least squares estimator defined in (5.7.3).

**Proof.** Consider the linear combination

$$\lambda' \mathbf{M}_n^{-1} \mathbf{X}' \mathbf{T}_n' \mathbf{e} = \sum_{t=1}^n c_{tn} e_{tn},$$

where

$$c_{tn} = \lambda' \mathbf{M}_n^{-1} \mathbf{X}' \mathbf{T}_n'.$$

$\lambda$  is an arbitrary  $k$ -dimensional vector with  $0 < |\lambda| < \infty$ , and  $\mathbf{T}_n'$  is the  $t$ th row of  $\mathbf{T}_n$ . By construction, the variance of  $\sum_{t=1}^n c_{tn} e_{tn}$ , denoted by  $V_n$ , is

$$V_n = \sum_{t=1}^n c_{tn}^2 V\{e_{tn}\} = \sum_{t=1}^n c_{tn}^2.$$

Hence,  $V_n^{-1} \sum_{t=1}^n c_{tn}^2$  is bounded. By the assumption (5.7.19),

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{t=1}^n c_{tn}^2 &= \lim_{n \rightarrow \infty} \lambda' \mathbf{M}_n^{-1} \mathbf{X}' \mathbf{T}_n' \mathbf{T}_n \mathbf{X} \mathbf{M}_n^{-1} \lambda \\ &= \lambda' \mathbf{A}_0 \lambda > 0. \end{aligned} \quad (5.7.21)$$

Let  $b_{ij}$  be the  $ij$ th element of  $\mathbf{M}_n^{-1} \mathbf{X}' \mathbf{T}_n'$ . Thus,

$$c_{tn}^2 = \sum_{i=1}^k \sum_{j=1}^k \lambda_i \lambda_j b_{ii} b_{jj} \leq k^2 \max_{1 \leq i \leq k} \lambda_i^2 \max_{1 \leq i \leq k} b_{ii}^2,$$

which implies that

$$\sup_{1 \leq i \leq n} c_{tn}^2 \leq k^2 \max_{1 \leq i \leq k} \lambda_i^2 \sup_{\substack{1 \leq j \leq n \\ 1 \leq i \leq k}} b_{ij}^2.$$

By the assumption (5.7.20) and (5.7.21),

$$\lim_{n \rightarrow \infty} V_n^{-1} \sup_{1 \leq i \leq n} c_{tn}^2 = 0.$$

Hence, by Corollary 5.3.4,

$$V_n^{-1/2} \sum_{t=1}^n c_{tn} e_{tn} \xrightarrow{\mathcal{D}} N(0, 1).$$

Since  $\lambda$  is arbitrary, by the assumption (5.7.19) we have

$$\mathbf{M}_n^{-1} \mathbf{X}' \mathbf{V}_{uu}^{-1} \mathbf{u} \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{A}_0),$$

and we obtain the conclusion. ▲

To this point, no structure has been imposed on  $\mathbf{V}_{uu}$  other than that it must be

positive definite. The number of unknown parameters in  $\mathbf{V}_{uu}$  could conceivably grow with the sample size  $n$ . A model of practical importance is

$$\begin{aligned} E\{\mathbf{u} | \mathbf{X}\} &= \mathbf{0} \\ E\{\mathbf{u}\mathbf{u}' | \mathbf{X}\} &= \mathbf{V}_{uu} = \mathbf{V}_{uu}(\boldsymbol{\theta}^0), \end{aligned} \quad (5.7.22)$$

where  $\boldsymbol{\theta}$  is an  $l \times 1$  vector of unknown parameters,  $l$  is fixed, and  $\boldsymbol{\theta}^0$  is the true value. The parameter space for  $\boldsymbol{\theta}$  is  $\Theta$ . It is assumed that the form of the function  $\mathbf{V}_{uu}(\boldsymbol{\theta})$  is known and that  $\mathbf{V}_{uu}(\boldsymbol{\theta})$  is a continuous function of  $\boldsymbol{\theta}$ . If, for example, the time series is known to be a  $p$ th order autoregressive process, the vector  $\boldsymbol{\theta}$  will contain the parameters of the autoregressive process. We are interested in the situation in which an estimator of  $\boldsymbol{\theta}$ , denoted by  $\hat{\boldsymbol{\theta}}$ , is used to construct an estimator of  $\mathbf{V}_{uu}(\boldsymbol{\theta})$ , denoted by  $\hat{\mathbf{V}}_{uu} = \mathbf{V}_{uu}(\hat{\boldsymbol{\theta}})$ .

The following theorem gives sufficient conditions for the estimated generalized least squares estimator to have the same asymptotic distribution as the generalized least squares estimator for general normalizing matrices under the model (5.7.1) and (5.7.22).

**Theorem 5.7.3.** Let the model (5.7.1) and (5.7.22) hold. Let  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_l)'$  and let

$$\mathbf{B}_{ni}(\boldsymbol{\theta}) = \frac{\partial \mathbf{V}_{uu}^{-1}(\boldsymbol{\theta})}{\partial \theta_i}, \quad i = 1, 2, \dots, l,$$

be continuous in  $\boldsymbol{\theta}$ . Let  $\tilde{\boldsymbol{\beta}}$  be defined by (5.7.5). Assume there exists a sequence of nonsingular matrices  $\{\mathbf{M}_n\}$  such that

$$\mathbf{M}_n' \mathbf{G}_n^{-1} \mathbf{M}_n = O_p(1), \quad (5.7.23)$$

$$\mathbf{M}_n^{-1} \mathbf{X}' \mathbf{V}_{uu}^{-1}(\boldsymbol{\theta}^0) \mathbf{u} = O_p(1), \quad (5.7.24)$$

where  $\mathbf{G}_n = \mathbf{X}' \mathbf{V}_{uu}^{-1}(\boldsymbol{\theta}^0) \mathbf{X}$ . Also assume

$$\mathbf{M}_n^{-1} \mathbf{X}' \mathbf{B}_{ni}(\boldsymbol{\theta}) \mathbf{X} \mathbf{M}_n^{-1} = O_p(1), \quad (5.7.25)$$

$$\mathbf{M}_n^{-1} \mathbf{X}' \mathbf{B}_{ni}(\boldsymbol{\theta}) \mathbf{u} = O_p(1), \quad (5.7.26)$$

for  $i = 1, 2, \dots, l$ , uniformly in an open neighborhood of  $\boldsymbol{\theta}^0$ , denoted by  $C(\boldsymbol{\theta}^0)$ . Let an estimator of  $\boldsymbol{\theta}^0$ , denoted by  $\hat{\boldsymbol{\theta}}$ , be available, and assume

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O_p(\xi_n), \quad (5.7.27)$$

where  $\xi_n \rightarrow 0$  as  $n \rightarrow \infty$ . Then

$$\mathbf{M}_n' (\tilde{\boldsymbol{\beta}} - \boldsymbol{\beta}) = \mathbf{M}_n' (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) + O_p(\xi_n).$$

**Proof.** By a Taylor expansion

$$\begin{aligned} \mathbf{M}_n^{-1} \mathbf{X}' \mathbf{V}_{uu}^{-1}(\hat{\theta}) \mathbf{X} \mathbf{M}_n^{-1'} &= \mathbf{M}_n^{-1} \mathbf{X}' \mathbf{V}_{uu}^{-1}(\theta^0) \mathbf{X} \mathbf{M}_n^{-1'} \\ &\quad + \sum_{i=1}^l \mathbf{M}_n^{-1} \mathbf{X}' \hat{\mathbf{B}}_{ni}(\theta^0, \hat{\theta}) \mathbf{X} \mathbf{M}_n^{-1'} (\hat{\theta}_i - \theta_i^0), \end{aligned}$$

where  $\theta^0$  is the true value of  $\theta$ , and  $\hat{\mathbf{B}}_{ni}(\theta^0, \hat{\theta})$  is a matrix whose  $js$  element is the  $js$  element of  $\mathbf{B}_{ni}(\theta)$  evaluated at a point  $\hat{\theta}_{js}$  on the line segment joining  $\hat{\theta}$  and  $\theta^0$ . Let the probability space be  $(\Omega, \mathcal{A}, P)$ . Let  $\epsilon > 0$  be given. Choose  $\delta$  such that  $\{\theta : |\theta^0 - \theta| < \delta\} \subset C(\theta^0)$ , where for any vector  $v$ ,  $|v|^2 = v'v$ . Since  $\hat{\theta}$  converges in probability to  $\theta^0$ , there exists an  $N_1$  and a set  $D_{1,n} \in \mathcal{A}$  such that  $P(D_{1,n}) > 1 - \epsilon/2$  and

$$|\hat{\theta}_{js} - \theta^0| < \delta$$

for  $j = 1, 2, \dots, n$ , for  $s = 1, 2, \dots, n$ , for all  $n > N_1$ , and for all  $\omega \in D_{1,n}$ . By the assumption (5.7.25), there exists a  $K$ , an  $N_2$ , and a set  $D_{2,n} \in \mathcal{A}$  with  $P(D_{2,n}) > 1 - \epsilon/2$  such that

$$\|\mathbf{M}_n^{-1} \mathbf{X}' \mathbf{B}_{ni}(\theta) \mathbf{X} \mathbf{M}_n^{-1'}\| < K$$

for all  $n > N_2$ , all  $\omega \in D_{2,n}$ , and all  $\theta \in C(\theta^0)$ , where  $\|\mathbf{H}\| = [\text{tr}(\mathbf{H}'\mathbf{H})]^{1/2}$ . Let  $D_n = D_{1,n} \cap D_{2,n}$ , and observe that  $P(D_n) > 1 - \epsilon$ . Let  $N = \max(N_1, N_2)$ . Therefore, for all  $n > N$  and for all  $\omega \in D_n$ ,

$$\|\mathbf{M}_n^{-1} \mathbf{X}' \hat{\mathbf{B}}_{ni}(\theta^0, \hat{\theta}) \mathbf{X} \mathbf{M}_n^{-1'}\| < K.$$

Hence,

$$\mathbf{M}_n^{-1} \mathbf{X}' \hat{\mathbf{B}}_{ni}(\theta^0, \hat{\theta}) \mathbf{X} \mathbf{M}_n^{-1'} = O_p(1), \quad (5.7.28)$$

which implies the assumption (5.7.28) of Theorem 5.7.1. By a similar expansion,

$$\begin{aligned} \mathbf{M}_n^{-1} \mathbf{X}' \mathbf{V}_{uu}^{-1}(\hat{\theta}) \mathbf{u} &= \mathbf{M}_n^{-1} \mathbf{X}' \mathbf{V}_{uu}^{-1}(\theta^0) \mathbf{u} \\ &\quad + \sum_{i=1}^l \mathbf{M}_n^{-1} \mathbf{X}' \hat{\mathbf{B}}_{ni}(\theta^0, \hat{\theta}) \mathbf{u} (\hat{\theta}_i - \theta_i^0), \end{aligned}$$

where  $\hat{\mathbf{B}}_{ni}(\theta^0, \hat{\theta})$  is analogous to  $\hat{\mathbf{B}}_{ni}(\theta^0, \hat{\theta})$ . Using the assumption (5.7.26) and an argument similar to the one used to show (5.7.28),

$$\mathbf{M}_n^{-1} \mathbf{X}' \hat{\mathbf{B}}_{ni}(\theta^0, \hat{\theta}) \mathbf{u} = O_p(1).$$

Thus the assumption (5.7.9) of Theorem 5.7.1 is satisfied. Because the assump-

tions (5.7.23) and (5.7.24) are the same as the assumptions (5.7.6) and (5.7.7) of Theorem 5.7.1, the result follows.  $\blacktriangle$

A sufficient condition for (5.7.23) is

$$\mathbf{M}_n^{-1} \mathbf{G}_n \mathbf{M}_n^{-1\prime} \xrightarrow{P} \mathbf{A}_0, \quad (5.7.29)$$

where  $\mathbf{A}_0$  is a fixed positive definite matrix, and a sufficient condition for (5.7.25) is

$$\mathbf{M}_n^{-1} \mathbf{X}' \mathbf{B}_{ni}(\boldsymbol{\theta}) \mathbf{X} \mathbf{M}_n^{-1\prime} \xrightarrow{P} \mathbf{A}_i(\boldsymbol{\theta}), \quad i = 1, 2, \dots, l,$$

uniformly in a neighborhood of  $\boldsymbol{\theta}^0$  as  $n \rightarrow \infty$ , where the  $\mathbf{A}_i(\boldsymbol{\theta})$  are continuous in  $\boldsymbol{\theta}$ .

Under the assumptions of Theorem 5.7.3, the limiting distribution of the normalized estimated generalized least squares estimator is the same as the limiting distribution of the normalized generalized least squares estimator constructed with known  $\mathbf{V}_{uu}$ , provided the limiting distribution exists. Note that the estimator of  $\boldsymbol{\theta}$  can converge to  $\boldsymbol{\theta}^0$  at a relatively slow rate.

A common procedure is to estimate  $\boldsymbol{\beta}$  by the ordinary least squares estimator,

$$\hat{\boldsymbol{\beta}} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{Y}, \quad (5.7.30)$$

compute residuals

$$\hat{\mathbf{u}} = \mathbf{Y} - \mathbf{X} \hat{\boldsymbol{\beta}}, \quad (5.7.31)$$

and use these residuals to estimate  $\boldsymbol{\theta}$ . Then the estimator of  $\boldsymbol{\theta}$  is used in (5.7.5) to estimate  $\boldsymbol{\beta}$ . In order for this procedure to be effective, the estimator of  $\boldsymbol{\theta}$  based on  $\hat{\mathbf{u}}$  must be a consistent estimator of  $\boldsymbol{\theta}^0$ , and this, in turn, requires  $\hat{\boldsymbol{\beta}}$  of (5.7.30) to be a consistent estimator of  $\boldsymbol{\beta}$ . The following lemma gives sufficient conditions for the estimator of  $\boldsymbol{\theta}^0$  based on the ordinary least squares residuals to be consistent.

**Lemma 5.7.1.** Consider the model (5.7.1). Let  $\hat{\boldsymbol{\theta}} = \hat{\boldsymbol{\theta}}(\mathbf{u})$  be an estimator of  $\boldsymbol{\theta}^0$  based upon the true  $\mathbf{u}$ , and let

$$\hat{\mathbf{u}}^* = [\mathbf{I} - c \mathbf{X} (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}'] \mathbf{u},$$

where  $0 < c < 1$ . Assume that  $\hat{\boldsymbol{\theta}}(\mathbf{u})$  is a continuous function of  $\mathbf{u}$  with a continuous first derivative, and that

$$\hat{\boldsymbol{\theta}}(\mathbf{u}) - \boldsymbol{\theta}^0 = O_p(\xi_n), \quad (5.7.32)$$

where  $\xi_n \rightarrow 0$  as  $n \rightarrow \infty$ . Also assume

$$\sum_{i=1}^n \left( \frac{\partial \hat{\boldsymbol{\theta}}_j(\hat{\mathbf{u}})}{\partial u_i} \right) \mathbf{X}_i (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) = o_p(\xi_n), \quad (5.7.33)$$

for  $j = 1, 2, \dots, l$ , uniformly in  $c$  for  $0 \leq c \leq 1$ , where  $\hat{\beta}$  is defined in (5.7.30). Let  $\hat{\theta}(\bar{u})$  be an estimator of  $\theta$  constructed with the residuals defined in (5.7.31). Then

$$\hat{\theta}(\bar{u}) = \hat{\theta}(u) + o_p(\xi_n) = \theta^0 + O_p(\xi_n).$$

**Proof.** By a Taylor expansion about  $u$ , we have

$$\hat{\theta}_j(\bar{u}) = \hat{\theta}_j(u) - \sum_{i=1}^n \frac{\partial \hat{\theta}_j(\bar{u})}{\partial u_i} X_i(\hat{\beta} - \beta), \quad (5.7.34)$$

where  $\bar{u}$  is on the line segment joining  $u$  and  $\bar{u}$ ,  $X_i$  is the  $i$ th row of  $X$ , and  $\bar{u}_i - u_i = -X_i(\hat{\beta} - \beta)$ . Because  $\bar{u}$  is of the form given in the assumptions, it follows from (5.7.33) that

$$\hat{\theta}_j(\bar{u}) = \hat{\theta}_j(u) + o_p(\xi_n) = \theta_j^0 + O_p(\xi_n). \quad \blacktriangle$$

Given that  $\hat{\theta}$  is a consistent estimator of  $\theta^0$ , the difference between the estimator of  $\beta$  based upon  $\hat{\theta}$  and the estimator of  $\beta$  based on  $\theta^0$  converges to zero.

**Theorem 5.7.4.** Let the assumptions (5.7.23)–(5.7.26) of Theorem 5.7.3 hold. In addition, assume (5.7.32),

$$M'_n(X'X)^{-1}M_n = O_p(1), \quad (5.7.35)$$

$$E\{M_n^{-1}X'V_{uu}(\theta^0)XM_n^{-1}\} = O(1), \quad (5.7.36)$$

and

$$\sum_{i=1}^n \frac{\partial \hat{\theta}_j(\bar{u})}{\partial u_i} X_i M_n^{-1} = o_p(\xi_n) \quad (5.7.37)$$

uniformly in  $c$  for  $0 \leq c \leq 1$  and  $j = 1, 2, \dots, l$ , where  $X_i$  is the  $i$ th row of  $X$ , and  $\bar{u}$  is defined in Lemma 5.7.1. Then

$$M'_n[\hat{\beta}(\hat{\theta}(\bar{u})) - \beta] = M'_n[\hat{\beta}(\theta^0) - \beta] + O_p(\xi_n),$$

where

$$\hat{\beta}(\theta^0) = [X'V_{uu}^{-1}(\theta^0)X]^{-1}X'V_{uu}^{-1}(\theta^0)Y.$$

If it is further assumed that

$$M'_n[\hat{\beta}(\theta^0) - \beta] \xrightarrow{\mathcal{L}} N(\mathbf{0}, A_0^{-1}), \quad (5.7.38)$$

then

$$\mathbf{M}'_n[\hat{\beta}(\hat{\theta}(\bar{u})) - \beta] \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{A}_0^{-1}), \quad (5.7.39)$$

where  $\hat{\theta}(\bar{u})$  is defined in Lemma 5.7.1.

**Proof.** Under the assumptions (5.7.36) and (5.7.22),

$$E\{E[\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{u}\mathbf{u}'\mathbf{X}\mathbf{M}_n^{-1}' | \mathbf{X}]\} = E\{\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{V}_{uu}(\theta^0)\mathbf{X}\mathbf{M}_n^{-1}'\} = O(1),$$

which implies that  $\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{u} = O_p(1)$  and

$$\mathbf{M}'_n(\ddot{\beta} - \beta) = \mathbf{M}'_n(\mathbf{X}'\mathbf{X})^{-1}\mathbf{M}_n\mathbf{M}_n^{-1}\mathbf{X}'\mathbf{u} = O_p(1). \quad (5.7.40)$$

By (5.7.40) and the assumption (5.7.37),

$$\sum_{i=1}^n \frac{\partial \hat{\theta}_i(\bar{u})}{\partial u_i} \mathbf{X}_i(\ddot{\beta} - \beta) = o_p(\xi_n). \quad (5.7.41)$$

By (5.7.41), the assumption (5.7.32), and Lemma 5.7.1, we have  $\hat{\theta}(\bar{u}) = \theta^0 + O_p(\xi_n)$ , and the assumption (5.7.27) of Theorem 5.7.3 holds. Also, the other assumptions of Theorem 5.7.3 hold for  $\hat{\theta}(\bar{u})$ . Therefore,

$$\mathbf{M}'_n[\hat{\beta}(\hat{\theta}(\bar{u})) - \beta] = \mathbf{M}'_n[\hat{\beta}(\hat{\theta}^0) - \beta] + O_p(\xi_n),$$

and (5.7.39) follows from the assumption (5.7.38). ▲

Under the conditions of Theorem 5.7.4, the procedure of using the ordinary least squares residuals to estimate the covariance matrix and then using the estimated covariance matrix in the estimated generalized least squares estimator produces an estimator of  $\beta$  with the same large sample properties as the estimator constructed with known  $\theta$ . Also

$$\hat{\mathbf{V}}\{\hat{\beta}(\hat{\theta}(\bar{u}))\} = [\mathbf{X}'\mathbf{V}_{uu}^{-1}(\hat{\theta})\mathbf{X}]^{-1} \quad (5.7.42)$$

can be used as an estimator of the covariance matrix of the approximate distribution of  $\hat{\beta}$ .

## 5.8. SEQUENCES OF ROOTS OF POLYNOMIALS

In Chapter 2, we saw that the roots of the characteristic polynomial were important in defining the behavior of autoregressive and autoregressive moving average time series. Hence, the relationships between the coefficients of a polynomial and the

roots are of interest. We write the  $p$ th order polynomial as

$$g(m) = m^p + a_1 m^{p-1} + \cdots + a_p = (m - m_1)(m - m_2) \cdots (m - m_p), \quad (5.8.1)$$

where the  $m_i$ ,  $i = 1, 2, \dots, p$ , are the roots of  $g(m) = 0$ . Let  $\bar{m}_1, \bar{m}_2, \dots, \bar{m}_q$  denote the distinct roots with multiplicities  $r_1, r_2, \dots, r_q$ , where  $\sum_{i=1}^q r_i = p$ . If  $r_i = 1$ , we say that  $\bar{m}_i = m_i$  is a simple root. The coefficients can be expressed in terms of the roots by

$$\begin{aligned} a_1 &= - \sum_{j=1}^p m_j, & a_2 &= \sum_{i=1}^p \sum_{j=i+1}^p m_i m_j, \\ a_3 &= - \sum_{i=1}^p \sum_{j=i+1}^p \sum_{l=j+1}^p m_i m_j m_l, \dots, & a_p &= (-1)^p \prod_{i=1}^p m_i. \end{aligned} \quad (5.8.2)$$

Thus, the coefficients are continuous differentiable functions of the roots. The functions (5.8.2) define a mapping from a  $p$ -dimensional space to a  $p$ -dimensional space. The differentials are

$$\begin{aligned} da_1 &= - \sum_{j=1}^p dm_j, \\ da_2 &= -a_1 \sum_{j=1}^p dm_j - \sum_{j=1}^p m_j dm_j, \\ da_3 &= -a_2 \sum_{j=1}^p dm_j - a_1 \sum_{j=1}^p m_j dm_j - \sum_{j=1}^p m_j^2 dm_j, \\ &\vdots \\ da_p &= -a_{p-1} \sum_{j=1}^p dm_j - a_{p-2} \sum_{j=1}^p m_j dm_j - \cdots - \sum_{j=1}^p m_j^{p-1} dm_j. \end{aligned} \quad (5.8.3)$$

Thus

$$d\mathbf{a} = \mathbf{B}\mathbf{H} d\mathbf{m}, \quad (5.8.4)$$

where  $d\mathbf{a} = (da_1, da_2, \dots, da_p)', d\mathbf{m} = (dm_1, dm_2, \dots, dm_p)'$ ,

$$\mathbf{B} = \begin{pmatrix} -1 & 0 & 0 & \cdots & 0 \\ -a_1 & -1 & 0 & \cdots & 0 \\ -a_2 & -a_1 & -1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -a_{p-1} & -a_{p-2} & -a_{p-3} & \cdots & -1 \end{pmatrix}, \quad (5.8.5)$$

and

$$\mathbf{H} = \begin{pmatrix} 1 & 1 & \cdots & 1 \\ m_1 & m_2 & \cdots & m_p \\ m_1^2 & m_2^2 & \cdots & m_p^2 \\ \vdots & \vdots & & \vdots \\ m_1^{p-1} & m_2^{p-1} & \cdots & m_p^{p-1} \end{pmatrix}. \quad (5.8.6)$$

The matrix  $\mathbf{H}$  is nonsingular if all roots are simple. Because  $\mathbf{B}$  is nonsingular, it is possible to solve for the differentials of  $m_i$  as functions of the differentials of  $a_i$ , when all roots are simple.

If the coefficients of the polynomial are real, the complex roots are complex conjugate pairs. In that case, we can use a transformation to define real quantities

$$c_1 = m_1 + m_2, \quad (5.8.7)$$

$$c_2^2 = (m_1 - m_2)^2 \quad (5.8.8)$$

for every complex conjugate pair. The inverse transformation is

$$m_1 = 0.5(c_1 + c_2),$$

$$m_2 = 0.5(c_1 - c_2), \quad (5.8.9)$$

where we adopt the convention that  $m_1$  is defined with the positive coefficient on the radical and  $m_2$  is defined with the negative coefficient. The inverse transformation has continuous derivatives except at  $c_2 = 0$ , the point where  $m_1 = m_2$ . The derivatives of the coefficients with respect to  $c_1$  and with respect to  $c_2^2$  are derivatives of real valued functions with respect to real valued arguments.

From (5.8.6) we see that repeated roots will produce a singularity in the Jacobian of the transformation. However, even if there are repeated roots, the roots of a polynomial are continuous functions of the coefficients of the polynomial.

**Lemma 5.8.1.** Let  $g(m)$  of (5.8.1) have distinct roots  $\bar{m}_1, \bar{m}_2, \dots, \bar{m}_q$  with multiplicities  $r_1, r_2, \dots, r_q$ , where  $a_1, a_2, \dots, a_p$  are complex numbers, and  $\sum_{j=1}^q r_j = p$ . Let

$$G(m) = m^p + (a_1 + \delta_1)m^{p-1} + \cdots + (a_p + \delta_p),$$

where  $\delta_1, \dots, \delta_p$  are complex numbers. For  $i = 1, 2, \dots, q$ , let  $\epsilon_i$  be an arbitrary real number satisfying

$$0 < \epsilon_i < \min_{j \neq i} |\bar{m}_j - \bar{m}_i|.$$

Then there exists a  $\delta > 0$  such that, if  $|\delta_i| \leq \delta$  for  $i = 1, 2, \dots, p$ , then  $G(m)$  has precisely  $r_i$  roots in the circle  $\phi_i$  with center at  $\bar{m}_i$  and radius  $\epsilon_i$ , for  $i = 1, 2, \dots, q$ .

**Proof.** Omitted. See Marden (1966, p. 3). ▲

To extend the results of Lemma 5.8.1 to sequences of polynomials, let

$$g_n(m) = m^p + a_{1n}m^{p-1} + \dots + a_{pn}, \quad (5.8.10)$$

where  $a_{in}$ ,  $i = 1, 2, \dots, p$ , are sequences of real numbers such that  $a_{in} \rightarrow a_i$ ,  $i = 1, 2, \dots, p$ , as  $n \rightarrow \infty$ , where the  $a_i$  are the coefficients of (5.8.1). Let  $\bar{m}_1, \bar{m}_2, \dots, \bar{m}_q$  be the distinct roots of (5.8.1) with multiplicities  $r_1, r_2, \dots, r_q$ , and let  $\epsilon$  be the arbitrary real number defined in Lemma 5.8.1. Then, by Lemma 5.8.1, there exists an  $N$  such that, for  $n \geq N$ ,  $g_n(m)$  has precisely  $r_i$  roots in the sphere  $\phi_i$  with center  $\bar{m}_i$  and radius  $\epsilon$ , for  $i = 1, \dots, q$ .

We now investigate the special case where a root has multiplicity two. To study the relation between the coefficients and the roots in the neighborhood of  $m_1 = m_2$ , we use the transformation introduced in (5.8.7) and (5.8.8) and assume  $m_3, m_4, \dots, m_p$  are simple roots. Then

$$\begin{aligned} a_1 &= -c_1 - \sum_{j=3}^p m_j, \\ a_2 &= m_1 m_2 + c_1 \sum_{j=3}^p m_j + \sum_{j=3}^p \sum_{i=j+1}^p m_i m_j \\ &= 0.25(c_1^2 - c_2^2) + c_1 \sum_{j=3}^p m_j + \sum_{j=3}^p \sum_{i=j+1}^p m_i m_j, \\ a_3 &= -0.25(c_1^2 - c_2^2) \sum_{j=3}^p m_j - c_1 \sum_{j=3}^p \sum_{i=j+1}^p m_i m_j \\ &\quad + \sum_{i=3}^p \sum_{j=i+1}^p \sum_{k=j+1}^p m_i m_j m_k \\ &\vdots \\ a_p &= 0.25(c_1^2 - c_2^2)(-1)^p \prod_{j=3}^p m_j. \end{aligned} \quad (5.8.11)$$

We observe that the partial derivatives of the coefficients with respect to  $c_1$  evaluated at  $m_1 = m_2$ , are equal to the partial derivatives of the functions (5.8.2) with respect to  $m_1$ . Also,  $c_2$  enters the expressions for the coefficients only as a square. In the proof of Lemma 5.8.2, we show that  $c_1, c_3, m_3, \dots, m_p$ , where  $c_3 = c_2^2$  and  $m_3, \dots, m_p$  are simple roots, can be expressed as locally continuous differentiable functions of the coefficients. Thus, Lemma 5.8.2 extends the result of Lemma 5.8.1 to the case in which some of the roots are of multiplicity two. The

bound on the difference between two equal roots is the square root of the bound on the error in a simple root.

**Lemma 5.8.2.** Let  $\mathbf{a}_n = (a_{1n}, a_{2n}, \dots, a_{pn})$ ,  $n = 1, 2, \dots$ , be a sequence of real vectors such that

$$|\mathbf{a}_n - \mathbf{a}| \leq M_n,$$

where  $M_n \rightarrow 0$  as  $n \rightarrow \infty$  and  $\mathbf{a} = (a_1, a_2, \dots, a_p)$  is the vector of real coefficients of (5.8.1). Let  $m_{in}$ ,  $i = 1, 2, \dots, p$ , be the roots of

$$m^p + a_{1n}m^{p-1} + a_{2n}m^{p-2} + \dots + a_{pn} = 0,$$

and let  $m_i$ ,  $i = 1, 2, \dots, p$ , be the roots of (5.8.1). Assume the roots of (5.8.1) are either simple roots or roots of multiplicity two. If  $m_i$  is a simple root of (5.8.1),

$$|m_{in} - m_i| \leq KM_n$$

for some  $K < \infty$ , where it is understood that for  $n$  sufficiently large,  $K$  can be chosen such that there is exactly one  $m_{in}$  satisfying the inequality for each simple  $m_i$ .

If  $m_1 = m_2$ ,

$$\begin{aligned} |m_{1n} + m_{2n} - 2m_1| &\leq KM_n, \\ |m_{1n} - m_{2n}|^2 &\leq KM_n \end{aligned}$$

for some  $K < \infty$ , where, for  $n$  sufficiently large,  $K$  can be chosen such that exactly one pair of roots satisfies the inequalities.

**Proof.** Let the roots, denoted by  $\mathbf{m} = (m_1, m_2, \dots, m_p)$ , be arranged so that the repeated roots occur as pairs in the first part of the vector. Let  $(m_1, m_2)$ ,  $(m_3, m_4), \dots, (m_{2k-1}, m_{2k})$  be the pairs of equal roots. Let

$$c_{1i} = m_{2i-1} + m_{2i} \quad \text{and} \quad c_{3i} = (m_{2i-1} - m_{2i})^2$$

and

$$m_{2i-1} = 0.5(c_{1i} + c_{3i}^{1/2}) \quad \text{and} \quad m_{2i} = 0.5(c_{1i} - c_{3i}^{1/2})$$

for  $i = 1, 2, \dots, k$ . The total differential of  $\mathbf{a}$  with respect to

$$\mathbf{u} = (c_{11}, c_{12}, c_{31}, c_{32}, \dots, c_{1k}, c_{3k}, m_{2k+1}, \dots, m_p),$$

analogous to (5.8.4), is

$$d\mathbf{a} = \mathbf{B}\mathbf{H}_1 d\mathbf{u}, \tag{5.8.12}$$

where  $\mathbf{H}_1 = \mathbf{H}\mathbf{J}$ ,

$$\mathbf{J} = \text{block diag}(\mathbf{J}_{11}, \mathbf{J}_{22}, \dots, \mathbf{J}_{kk}, \mathbf{I}),$$

$\mathbf{B}$  is defined in (5.8.5),  $\mathbf{H}$  is defined in (5.8.6),  $\mathbf{J}$  is the Jacobian of the transformation of  $\mathbf{u}$  into  $\mathbf{m}$ , and

$$\mathbf{J}_{ii} = \begin{pmatrix} 0.5 & 0.25c_{3i}^{-1/2} \\ 0.5 & -0.25c_{3i}^{-1/2} \end{pmatrix}.$$

The matrix  $\mathbf{H}$  is a Vandermonde matrix with determinant

$$|\mathbf{H}| = \prod_{1 \leq i < j \leq p} (m_j - m_i),$$

and the determinant of  $\mathbf{H}_1$  is

$$|\mathbf{H}_1| = (0.25)^k \prod_{i=1}^k (m_{2i-1} - m_{2i})^{-1} |\mathbf{H}|.$$

The determinant  $|\mathbf{H}_1|$  is not zero at  $m_{2i-1} = m_{2i}$  because every zero product in the determinant of  $\mathbf{H}$  is removed by the transformation. Hence, the differentials of  $\mathbf{u}$  are locally continuous differentiable functions of  $\mathbf{a}$ , and the results follow.  $\blacktriangle$

It is important that the bound given for the difference in Lemma 5.8.2 can be achieved. Consider, for example, the sequence of polynomials

$$g_n(m) = m^2 - (2 + 2n^{-1})m + 1 + n^{-1} + n^{-2}.$$

The roots of  $g_n(m) = 0$  are  $1 + n^{-1} + n^{-1/2}$  and  $1 + n^{-1} - n^{-1/2}$ . Thus, the difference between the two roots is  $O(n^{-1/2})$ . The conclusion of Lemma 5.8.2 can also be obtained from the following theorem.

**Theorem 5.8.1.** Let  $\mathbf{a}_n = (a_{1n}, a_{2n}, \dots, a_{pn})$ ,  $n = 1, 2, \dots$ , be a sequence of (possibly complex valued) vectors, and let  $\mathbf{a} = (a_1, a_2, \dots, a_p)$  be the vector of (possibly complex valued) coefficients of (5.8.1). Assume that

$$|\mathbf{a}_n - \mathbf{a}| = O(\kappa_n), \quad (5.8.13)$$

where  $\kappa_n \rightarrow 0$  as  $n \rightarrow \infty$ . Let  $\bar{m}_1, \bar{m}_2, \dots, \bar{m}_q$  denote the distinct roots of  $g(m) = 0$  with multiplicities  $r_1, r_2, \dots, r_q$ , where  $\sum_{i=1}^q r_i = p$ . Let  $\epsilon$  be an arbitrary real number satisfying

$$0 < \epsilon < 0.5 \min_{2 \leq j \leq q} |\bar{m}_1 - \bar{m}_j|,$$

and let  $r_1 = s$ . Then there exists an  $N$  such that if  $n \geq N$ , the equation

$$m^p + a_{n1}m^{p-1} + \dots + a_{np} = 0 \quad (5.8.14)$$

has exactly  $s$  roots in the circle  $\phi_1$  with center  $\bar{m}_1$  and radius  $\epsilon$  and

$$\begin{aligned} \sum_{i=1}^s \xi_{ni} &= O(\kappa_n) \\ \sum_{i=1}^s \sum_{\substack{j=1 \\ i \neq j}}^s \xi_{ni} \xi_{nj} &= O(\kappa_n) \\ &\quad \vdots \\ \prod_{i=1}^s \xi_{ni} &= O(\kappa_n), \end{aligned} \tag{5.8.15}$$

where  $\xi_{ni} = m_{ni} - \bar{m}_1$ ,  $i = 1, 2, \dots, s$ , in which  $m_{ni}$ ,  $i = 1, 2, \dots, s$ , are the  $s$  roots in  $\phi_1$ .

**Proof.** That  $s$  roots will fall in  $\phi_1$  when  $n \geq N$  follows from Lemma 5.8.1. Therefore, we only demonstrate the order results.

If  $\bar{m}_1 \neq 0$ , we can replace  $m$  in (5.8.1) with  $m - \bar{m}_1$  to obtain a polynomial with  $s$  zero roots. Therefore, without loss of generality, we assume  $\bar{m}_1 = 0$ . Then

$$a_p = a_{p-1} = \cdots = a_{p-s+1} = 0$$

and

$$a_{np} = O(\kappa_n), \quad a_{n,p-1} = O(\kappa_n), \dots, \quad a_{n,p-s+1} = O(\kappa_n).$$

The polynomial (5.8.1) can be written as

$$(m - \xi_{n1})(m - \xi_{n2}) \cdots (m - \xi_{ns})(m^{p-s} + b_{n1}m^{p-s-1} + \cdots + b_{n,p-s}), \tag{5.8.16}$$

where

$$\begin{aligned} (-1)^s \prod_{i=1}^s \xi_{ni} b_{n,p-s} &= a_{np} = O(\kappa_n), \\ (-1)^{s-1} \sum_{j=1}^s \prod_{\substack{i=1 \\ i \neq j}}^s \xi_{ni} b_{n,p-s} + (-1)^s \prod_{i=1}^s \xi_{ni} b_{n,p-s-1} &= a_{n,p-1} = O(\kappa_n), \\ (-1)^{s-2} \sum_{j=1}^s \sum_{k=1}^s \prod_{\substack{i=1 \\ i \neq j \\ i \neq k}}^s \xi_{ni} b_{n,p-s} + (-1)^{s-1} \sum_{j=1}^s \prod_{\substack{i=1 \\ i \neq j}}^s \xi_{ni} b_{n,p-s-1} \\ &\quad + (-1)^s \prod_{i=1}^n \xi_{ni} b_{n,p-s-2} = a_{n,p-2} = O(\kappa_n), \end{aligned}$$

(5.8.17)

⋮

$$\begin{aligned} & (-1) \sum_{i=1}^s \xi_{ni} b_{n,p-s} + (-1)^2 \sum_{i=1}^s \sum_{\substack{j=1 \\ i \neq j}}^s \xi_{ni} \xi_{nj} b_{n,p-s-1} \\ & + \cdots + (-1)^s \prod_{i=1}^s \xi_{ni} b_{n,p-2s+1} = a_{n,p-s+1} = O(\kappa_n). \end{aligned}$$

Because the  $a_{ni}$ ,  $i = 1, 2, \dots, p$ , converge, the  $b_{nj}$ ,  $j = 1, 2, \dots, p-s$ , also converge, and we write

$$(b_{n1}, b_{n2}, \dots, b_{n,p-s}) \rightarrow (b_1, b_2, \dots, b_{p-s}).$$

Since the multiplicity of  $\bar{m}_1 = 0$  is  $s$ , we have  $b_{p-s} \neq 0$ . Thus, for example,

$$\prod_{i=1}^s \xi_{ni} = O(\kappa_n)$$

from the first equation of (5.8.17). ▲

**Corollary 5.8.1.1.** If the assumption (5.8.13) is replaced with

$$|\mathbf{a}_n - \mathbf{a}| = O_p(\kappa_n),$$

then  $O(\kappa_n)$  of (5.8.15) is replaced with  $O_p(\kappa_n)$ .

**Proof.** Omitted. ▲

**Corollary 5.8.1.2.** Let the assumptions of Theorem 5.8.1 hold with  $s = 2$ . Then, for  $n \geq N$ ,

$$(m_{n1} - m_{n2})^2 = O(\kappa_n),$$

where  $m_{n1}$  and  $m_{n2}$  are the two roots in  $\phi_1$ .

**Proof.** We have, for  $m_{n1}$  and  $m_{n2}$  in  $\phi_1$ ,

$$\begin{aligned} (m_{n1} - m_{n2})^2 &= m_{n1}^2 + m_{n2}^2 - 2m_{n1}m_{n2} \\ &= O(\kappa_n) \end{aligned}$$

by Lemma 5.8.1 and by Theorem 5.8.1. ▲

We now give a result on the roots and vectors of square matrices. Let  $\mathbf{A}$  be any  $p \times p$  matrix with possibly complex elements. The roots of the determinantal

equation

$$|\mathbf{A} - m\mathbf{I}| = 0 \quad (5.8.18)$$

are called the characteristic roots of  $\mathbf{A}$ , the eigenvalues of  $\mathbf{A}$ , the eigenroots of  $\mathbf{A}$ , or simply the roots of  $\mathbf{A}$ . The vector  $\mathbf{u}_i$  that satisfies the equation

$$(\mathbf{A} - m_i \mathbf{I})\mathbf{u}_i = \mathbf{0}, \quad (5.8.19)$$

where  $m_i$  is a root of (5.8.18), is called a characteristic vector of  $\mathbf{A}$  or an eigenvector of  $\mathbf{A}$ . The following theorem, taken from Magnus and Neudecker (1988), demonstrates that the simple roots of  $\mathbf{A}$  are continuous differentiable functions of the elements of  $\mathbf{A}$ . The theorem gives the differentials of the roots and of the vectors.

**Theorem 5.8.2.** Let  $m_1$  be a simple root of (5.8.18) for  $\mathbf{A}_1$ , where  $\mathbf{A}_1$  is any (possibly complex)  $p \times p$  matrix. Let  $\mathbf{u}_1$  be the associated eigenvector. Then a complex valued function  $m(\mathbf{A})$  and a complex vector valued function  $\mathbf{u}(\mathbf{A})$  are defined for all  $\mathbf{A}$  in some neighborhood  $S$  of  $\mathbf{A}_1$ , such that

$$m(\mathbf{A}_1) = m_1, \quad \mathbf{u}(\mathbf{A}_1) = \mathbf{u}_1, \quad (5.8.20)$$

$$\mathbf{A}\mathbf{u}(\mathbf{A}) = m(\mathbf{A})\mathbf{u}(\mathbf{A}), \quad \text{and} \quad \mathbf{u}_1^*(\mathbf{A})\mathbf{u}(\mathbf{A}) = 1 \quad (5.8.21)$$

for all  $\mathbf{A}$  in  $S$ , where  $\mathbf{u}_1^*(\mathbf{A})$  is the complex conjugate of  $\mathbf{u}_1(\mathbf{A})$ . Furthermore, the functions are differentiable any number of times on  $S$ , and the differentials are

$$dm = (\mathbf{v}_1^*\mathbf{u}_1)^{-1} \mathbf{v}_1^*(d\mathbf{A})\mathbf{u}_1 \quad (5.8.22)$$

and

$$d\mathbf{u} = (\mathbf{A}_1 - m_1 \mathbf{I})^\dagger [\mathbf{I} - (\mathbf{v}_1^*\mathbf{u}_1)^{-1} \mathbf{u}_1 \mathbf{v}_1^*](d\mathbf{A})\mathbf{u}_1, \quad (5.8.23)$$

where  $\mathbf{B}^\dagger$  is the Moore-Penrose generalized inverse of  $\mathbf{B}$ ,  $\mathbf{v}_1$  is the eigenvector associated with the eigenvalue  $m_1^*$  of  $\mathbf{A}_1^*$ , satisfying

$$\mathbf{A}_1^*\mathbf{v}_1 = m_1^*\mathbf{v}_1,$$

and  $\mathbf{B}^*$  is the complex conjugate of  $\mathbf{B}$ .

**Proof.** Omitted. See Magnus and Neudecker (1988, p. 161). ▲

If  $\mathbf{A}_1$  is a real symmetric matrix, then the roots are real, and the expressions (5.8.20) and (5.8.21) simplify to

$$dm = \mathbf{u}_1'(d\mathbf{A})\mathbf{u}_1, \quad (5.8.24)$$

$$d\mathbf{u} = (\mathbf{A}_1 - m_1 \mathbf{I})^\dagger(d\mathbf{A})\mathbf{u}_1. \quad (5.8.25)$$

It follows from Theorem 5.8.1 that if the coefficients of a matrix are converging to limiting values at the rate  $M_n$ , then a simple root  $m_n$  is converging to a limiting value at the same rate.

The results for the roots of a matrix are applicable to the roots of a polynomial written in the form (5.8.1). Let the  $p \times p$  matrix  $\mathbf{A}$  be defined by

$$\mathbf{A} = \begin{pmatrix} -a_1 & -a_2 & -a_3 & \cdots & -a_{p-1} & a_p \\ 1 & 0 & 0 & \cdots & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 \end{pmatrix}, \quad (5.8.26)$$

where the  $a_i$  are the coefficients of (5.8.1). Then the roots of the determinantal equation

$$|\mathbf{A} - m\mathbf{I}| = 0 \quad (5.8.27)$$

are the same as the roots of (5.8.1).

## REFERENCES

- Section 5.1.** Chernoff (1956), Chung (1968), Mann and Wald (1943b), Pratt (1959), Tucker (1967).
- Sections 5.2, 5.3.** Billingsley (1968), Brown (1971), Chan and Wei (1988), Chung (1968), Cramér (1946), Dvoretzky (1972), Gnedenko (1967), Hall and Heyde (1980), Loeve (1963), Pollard (1984), Rao (1973), Scott (1973), Tucker (1967), Varadarajan (1958), Wilks (1962).
- Section 5.4.** Cramér (1946), DeGracie and Fuller (1972), Hansen, Hurwitz, and Madow (1953), Hatanaka (1973).
- Section 5.5.** Andrews (1987), Gallant (1971, 1987), Gallant and White (1988), Gumpertz and Pantula (1992), Hartley (1961), Hartley and Booker (1965), Jennrich (1969), Klimko and Nelson (1978), Malinvaud (1970b), Potscher and Prucha (1991a,b), Sarkar (1990), Wald (1949), Wu (1981).
- Section 5.6.** Basmann (1957), Fuller (1987), Sargan (1958, 1959), Theil (1961).
- Section 5.7.** Carroll and Ruppert (1988), Jobson and Fuller (1980), Rothenberg (1984), Sanger (1992), Toyooka (1982, 1990), Van Der Genugten (1983).
- Section 5.8.** Amemiya (1990), Magnus and Neudecker (1988), Marden (1966).

## EXERCISES

- Let  $\{a_n\}$  and  $\{b_n\}$  be sequences of real numbers, and let  $\{f_n\}$  and  $\{g_n\}$  be sequences of positive real numbers such that  $a_n = O(f_n)$  and  $b_n = O(g_n)$ . Show that:
  - $|a_n|^s = O(f_n^s)$ ,  $s > 0$ .

- (b)  $a_n b_n = O(f_n g_n)$ .
2. Let  $\{f_n\}$ ,  $\{g_n\}$ , and  $\{r_n\}$  be sequences of positive real numbers, and let  $\{X_n\}$ ,  $\{Y_n\}$ , and  $\{Z_n\}$  be sequences of random variables such that  $X_n = O_p(f_n)$ ,  $Y_n = O_p(g_n)$ , and  $Z_n = o_p(r_n)$ . Without recourse to Theorems 5.1.5 and 5.1.6, show that:
- $|X_n|^s = O_p(f_n^s)$ ,  $s > 0$ .
  - $X_n Y_n = O_p(f_n g_n)$ .
  - $X_n + Y_n = O_p(\max\{f_n, g_n\})$ .
  - $X_n Z_n = o_p(f_n r_n)$ .
  - If  $g_n/r_n = o(1)$ , then  $Y_n + Z_n = o_p(r_n)$ .
3. Let  $X_1, \dots, X_n$  and  $Y_1, \dots, Y_n$  be two independent random samples, each of size  $n$ . Let the  $X_i$  be  $N(0, 4)$  and the  $Y_i$  be  $N(2, 9)$ . Find the order in probability of the following statistics.
- $\bar{X}_n$ .
  - $\bar{Y}_n^2$ .
  - $\bar{Y}_n$ .
  - $\bar{X}_n \bar{Y}_n$ .
  - $\bar{X}_n^3$ .
  - $(\bar{X}_n + 1)$ .
- Find the most meaningful expression (i.e., the smallest quantity for which the order in probability statement is true).
4. Prove that  $\text{plim}_{n \rightarrow \infty} \hat{\theta} = \theta^0$  implies that there exists a sequence of positive real numbers  $\{a_n\}$  such that  $\lim_{n \rightarrow \infty} a_n = 0$  and  $\hat{\theta} - \theta^0 = O_p(a_n)$ .
5. Let  $\{a_t\}$  be a sequence of constants satisfying  $\sum_{t=1}^{\infty} |a_t| < \infty$ ; also let  $\{X_{tn}: t = 1, 2, \dots, n; n = 1, 2, \dots\}$  be a triangular array of random variables such that  $E\{X_{tn}^2\} = O(b_n^2)$ ,  $t = 1, 2, \dots, n$ , where  $\lim_{n \rightarrow \infty} b_n = 0$ . Prove
- $$\sum_{t=1}^n a_t X_{tn} = O_p(b_n).$$
6. Let  $\bar{x}_n$  be distributed as a normal  $(\mu, \sigma^2/n)$  random variable with  $\mu \neq 0$ , and define  $Y_n = \bar{x}_n^3 - \bar{x}_n^{-2}$ . Expand  $Y_n$  in a Taylor's series through terms of  $O_p(n^{-1})$ . Find the expectation of these terms.
7. Let  $F_1, F_2, \dots, F_M$  be a finite collection of distribution functions with finite variances and common mean  $\mu$ . A sequence of random variables  $\{X_t: t \in (1, 2, \dots)\}$  is created by randomly choosing, for each  $t$ , one of the distribution functions and then making a random selection from the chosen distribution. Show that, as  $n \rightarrow \infty$ ,

$$\left( \sum_{t=1}^n \sigma_t^2 \right)^{-1/2} \sum_{t=1}^n (X_t - \mu) \xrightarrow{d} N(0, 1),$$

where  $\sigma_t^2$  is the variance of the distribution chosen for index  $t$ . Does

$$n^{1/2} \left( M^{-1} \sum_{j=1}^M \sigma_j^2 \right)^{-1/2} \left( n^{-1} \sum_{t=1}^n X_t - \mu \right) \xrightarrow{\mathcal{L}} N(0, 1),$$

where  $\sigma_j^2$  is the variance of the  $j$ th distribution,  $j = 1, 2, \dots, M$ ?

8. Let  $X_i$  be normal independent  $(\mu, \sigma^2)$  random variables, and let  $\bar{x}_n = n^{-1} \sum_{i=1}^n X_i$ .
- It is known that  $\mu \neq 0$ . How would you approximate the distribution of  $\bar{x}_n^2$  in large samples?
  - It is known that  $\mu = 0$ . How would you approximate the distribution of  $\bar{x}_n^2$  in large samples?

Explain and justify your answers, giving the normalizing constants necessary to produce nondegenerate limiting distributions.

9. Let  $\bar{x}_n$  be distributed as a normal  $(\mu, \sigma^2/n)$  random variable,  $\mu > 0$ .
- Find  $E\{Y_n\}$  to order  $(1/n)$ , where

$$Y_n = \operatorname{sgn} \bar{x}_n |\bar{x}_n|^{1/2}$$

and  $\operatorname{sgn} \bar{x}_n$  denotes the sign of  $\bar{x}_n$ .

- Find  $E\{(Y_n - \mu^{1/2})^2\}$  to order  $(1/n)$ .
- What is the limiting distribution of  $n^{1/2}(Y_n - \mu^{1/2})$ ?

10. Let  $\bar{x}_n$  be distributed as a normal  $(\mu, \sigma^2/n)$  random variable,  $\mu \neq 0$ , and define

$$Z_n = \frac{n\bar{x}_n}{n\bar{x}_n^2 + \sigma^2}.$$

Show that  $E\{\bar{x}_n^{-1}\}$  is not defined but that  $E\{Z_n\}$  is defined. Show further that  $Z_n$  satisfies the conditions of Theorem 5.4.3, and find the expectation of  $Z_n$  through terms of  $O(1/n)$ .

11. Let  $(X, Y)'$  be distributed as a bivariate normal random variable with mean  $(\mu_x, \mu_y)'$  and covariance matrix

$$\begin{pmatrix} \sigma_x^2 & \sigma_{xy} \\ \sigma_{xy} & \sigma_y^2 \end{pmatrix}.$$

Given a sample of size  $n$  from this bivariate population, derive an estimator for the product  $\mu_x \mu_y$  that is unbiased to  $O(1/n)$ .

**12. Assume the model**

$$Y_t = \alpha + \lambda e^{-\beta x_t} + u_t,$$

where the  $u_t$  are distributed as normal independent  $(0, \sigma^2)$  random variables. We have available the following data:

$t$	$Y_t$	$x_t$	$t$	$Y_t$	$x_t$
1	47.3	0.0	7	136.5	2.0
2	87.0	0.4	8	132.0	4.0
3	120.1	0.8	9	68.8	0.0
4	130.4	1.6	10	138.1	1.5
5	58.8	0.0	11	145.7	3.0
6	111.9	1.0	12	143.0	5.9

where  $Y$  is yield of corn and  $x$  is applied nitrogen. Given the initial values  $\hat{\alpha} = 143$ ,  $\hat{\lambda} = -85$ ,  $\hat{\beta} = 1.20$ , carry out two iterations of the Gauss–Newton procedure. Using the estimates obtained at the second iteration, estimate the covariance matrix of the estimator.

- 13.** In the illustration of Section 5.5 only the coefficient of  $x_{t1}$  was used in constructing an initial estimate of  $\theta_1$ . Identifying the original equation as

$$Y_t = \theta_0 + \theta_1 x_{t1} + \alpha^2 x_{t2} + e_t,$$

construct an estimator for the covariance matrix of  $(\hat{\theta}_0, \hat{\theta}_1, \hat{\alpha})$ , where the coefficients  $\hat{\theta}_0$ ,  $\hat{\theta}_1$ , and  $\hat{\alpha}^2$  are the ordinary least squares estimates. Using this covariance matrix, find the  $\lambda$  that minimizes the estimated variance of

$$\lambda \hat{\theta}_1 + (1 - \lambda) \hat{\alpha}$$

as an estimator of  $\theta_1$ . Use the estimated linear combination as an initial estimate in the Gauss–Newton procedure.

- 14.** Assuming that the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables, obtain the likelihood function associated with the model (5.5.1). By evaluating the expectations of the second partial derivatives of the likelihood function with respect to the parameters, demonstrate that the asymptotic covariance matrix of the maximum likelihood estimator is the same as that given in Theorem 5.5.1.

- 15.** Let  $Y_t$  satisfy the model

$$Y_t = \theta_0 + \theta_1 e^{-\theta_2 t} + e_t, \quad t = 1, 2, \dots,$$

where  $\theta_2 > 0$  and  $e_t$  is a sequence of normal independent  $(0, \sigma^2)$  random variables. Does this model satisfy assumptions 1 and 2 of Theorem 5.4.4? Would the model with  $t$  in the exponent replaced by  $x_t$ , where  $\{x_t\} = \{1, 2, 3, 4, 1, 2, 3, 4, \dots\}$ , satisfy the three assumptions?

16. An experiment is conducted to study the relationship between the phosphate content of the leaf and the yield of grain for the corn plant. In the experiment different levels of phosphate fertilizer were applied to the soil of 20 experimental plots. The leaf phosphate and the grain yield of the corn were recorded for each plot. Denoting yield by  $Y$ , applied phosphate by  $A$ , and leaf phosphate by  $P$ , the sums of squares and cross products matrix for  $A$ ,  $P$ , and  $Y$  were computed as

$$\begin{pmatrix} \sum A^2 & \sum AP & \sum AY \\ \sum PA & \sum P^2 & \sum PY \\ \sum YA & \sum YP & \sum Y^2 \end{pmatrix} = \begin{pmatrix} 69,600 & 16,120 & 3,948 \\ 16,120 & 8,519 & 1,491 \\ 3,948 & 1,491 & 739 \end{pmatrix}.$$

The model

$$Y_t = \beta P_t + v_t, \quad t = 1, 2, \dots, 20,$$

where the  $v_t$  are normal independent  $(0, \sigma^2)$  random variables, is postulated. Estimate  $\beta$  by the method of instrumental variables, using  $A_t$  as the instrumental variable. Estimate the standard error of your coefficients. Compare your estimate with the least squares estimate.

17. Show that if  $X_n = O_p(a_n)$  and  $a_n = o(b_n)$ , then  $X_n = o_p(b_n)$ .
18. Prove the following corollary to Theorem 5.5.1.

**Corollary 5.5.1.** Let the assumptions of Theorem 5.5.1 hold. Also assume

$$E\{(e_t^2 - \sigma^2)^2 | \mathcal{A}_{t-1}\} = \kappa$$

and

$$E\{|e_t^2 - \sigma^2|^{2+\delta}\} < M_\epsilon.$$

Then

$$n^{1/2}(s^2 - \sigma^2) \xrightarrow{\mathcal{L}} N(0, \kappa).$$

19. Let  $s^2$  be defined by (5.5.57). Show that

$$s^2 = (n-k)^{-1} \mathbf{e}' \{ \mathbf{I} - \mathbf{F}(\boldsymbol{\theta}^0) [\mathbf{F}'(\boldsymbol{\theta}^0) \mathbf{F}(\boldsymbol{\theta}^0)]^{-1} \mathbf{F}'(\boldsymbol{\theta}^0) \} \mathbf{e} + O_p(\max\{a_n^3, n^{-1/2} a_n^2\})$$

under the assumptions of Theorem 5.5.4.

20. Show that if the assumptions (5.7.6) through (5.7.9) hold and if  $M_n^{-1} = o_p(1)$ , then  $\hat{\beta}$  converges to  $\beta$  in probability, where  $\hat{\beta}$  is defined in (5.7.5).

21. Let  $X_{tn}$ ,  $t = 1, 2, \dots, n = 1, 2, \dots$ , be a triangular array of random variables. Let  $\{g_t\}_{t=1}^\infty$  and  $\{h_n\}_{n=1}^\infty$  be sequences of positive real numbers. Show that if

$$E\{|X_{tn}|\} = g_t h_n, \quad t = 1, 2, \dots, n, \quad n = 1, 2, \dots,$$

then

$$\sum_{t=1}^n X_{tn} = O_p\left(h_n \sum_{t=1}^n g_t\right).$$

Show that  $X_{tn} = O_p(g_t h_n)$  does not imply that

$$\sum_{t=1}^n X_{tn} = O_p\left(h_n \sum_{t=1}^n g_t\right).$$

22. Let the model (5.7.1)–(5.7.2) hold with  $X_t = t$  and  $\mathbf{V}_{uu} = \text{diag}\{e^\theta, e^{2\theta}, \dots, e^{n\theta}\}$  for  $\theta \in \Theta = (1, \infty)$ . Let  $M_n = G_n^{1/2}$ . Suppose an estimator of  $\theta^0$ , the true value, is available such that  $\hat{\theta} - \theta^0 = O_p(n^{-1/2})$ .

- (a) Show that

$$\|\mathbf{V}_{uu}^{1/2} \hat{\mathbf{V}}_{uu}^{-1} \mathbf{V}_{uu}^{1/2} - \mathbf{I}_n\|_2$$

does not go to zero in probability, where  $\|\mathbf{A}\|_2$  is the square root of the largest eigenvalue of  $\mathbf{A}'\mathbf{A}$ .

- (b) Show that the assumption (5.7.8) holds for this model.

23. Let

$$f_n(z) = a_{0n} + a_{1n}z + \cdots + a_{kn}z^k,$$

and let

$$f(z) = a_0 + a_1 z + \cdots + a_k z^k = 0$$

have distinct roots  $\bar{z}_1, \bar{z}_2, \dots, \bar{z}_q$  with multiplicities  $r_1, r_2, \dots, r_q$ . Let

$$a_{in} - a_i = O_p(n^{-\xi})$$

for  $i = 1, 2, \dots, k$  and some  $\xi > 0$ . Prove that, given  $\epsilon > 0$ , there exists an  $N$  such that the probability is greater than  $1 - \epsilon$  that exactly  $r_i$  of the roots of  $f_n(z)$  are within  $M_\epsilon n^{-\xi}$  of  $\bar{z}_i$  for  $i = 1, 2, \dots, q$ .

**24.** Consider the regression model

$$Y_t = X_{t1}\beta_1 + X_{t2}\beta_2 + e_t,$$

where

$$(X_{t1}, X_{t2}) = (t + u_{t1}, t + u_{t2}), \\ \mathbf{u}_t = (u_{t1}, u_{t2})' \sim \text{II}(\mathbf{0}, \sigma_u^2 \mathbf{I}),$$

$e_t \sim \text{II}(0, \sigma_e^2)$ , and  $e_t$  is independent of  $\mathbf{u}_t$  for all  $t$  and  $j$ . The least squares estimator of  $\hat{\beta} = (\beta_1, \beta_2)'$  is

$$\hat{\beta} = \left( \sum_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \right)^{-1} \sum_{t=1}^n \mathbf{X}_t' \mathbf{Y}_t.$$

The variance of  $\hat{\beta}$  conditional on  $\mathbf{X}_t$ ,  $t = 1, 2, \dots, n$ , is

$$V\{\hat{\beta} | (\mathbf{X}_1, \dots, \mathbf{X}_n)\} = \left( \sum_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \right)^{-1} \sigma_e^2.$$

Show that

$$\underset{n \rightarrow \infty}{p\lim} n^{-3} \sum_{t=1}^n \mathbf{X}_t' \mathbf{X}_t = 3^{-1} \mathbf{J}' \mathbf{J},$$

where  $\mathbf{J} = (1, 1)$ . Construct a sequence of matrices  $\{\mathbf{M}_n\}$  such that

$$\underset{n \rightarrow \infty}{p\lim} \mathbf{M}_n^{-1} \sum_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \mathbf{M}_n^{-1} = \mathbf{I}$$

and such that

$$\mathbf{M}_n'(\hat{\beta} - \beta^0) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}).$$

**25.** Use Theorem 5.3.4 to prove the assertion made in Example 5.5.1 that

$$\mathbf{M}_n^{-1} \mathbf{U}_n(\beta^0) \xrightarrow{D} N(\mathbf{0}, \mathbf{V}_{\beta\beta}).$$

**26.** Let  $(a_{1n}, a_{2n})' \sim N[(-1.40, 0.49)', n^{-1} \mathbf{I}]$ . Let  $(m_{1n}, m_{2n})$  be the roots of

$$m^2 + a_{1n}m + a_{2n} = 0,$$

where  $|m_{1n}| \geq |m_{2n}|$ . What can you say about the distribution of  $n^{1/2}(m_{1n} - 0.7, m_{2n} - 0.7)$  as  $n$  increases?

**27.** Let  $(a_{1n}, a_{2n})' \sim N[(-1.20, 0.61)', n^{-1} \mathbf{I}]$ . Let  $(m_{1n}, m_{2n})$  be the roots of

$$m^2 + a_{1n}m + a_{2n} = 0.$$

What can you say about the distribution of  $n^{1/2}(m_{1n} - m_1, m_{2n} - m_2)'$ , where  $(m_1, m_2)$  are the roots of

$$m^2 - 1.20m + 0.61 = 0?$$

*Hint:* See Exercise 10 of Chapter 4.

28. Consider the sequence of polynomials

$$g_n(m) = m^2 - 3n^{-1}m + 2n^{-2}.$$

What is the order of the difference of the two roots? How do you explain the difference between this result and the example of the text? Use the proof of Theorem 5.8.1 to prove a general result for the order of the difference of the two roots of a second order polynomial that is converging to a polynomial with repeated roots.

29. Let  $\{\mathbf{Y}_i\}_{i=1}^{\infty}$  be a sequence of  $k$ -dimensional random variables, where  $\mathbf{Y}_i \sim NI(\mu \mathbf{J}, \Sigma)$ , and  $\mathbf{J}$  is a  $k$ -dimensional column vector of ones. Find the limiting distribution, as  $n \rightarrow \infty$ , of

$$\hat{\mu}_g = (\mathbf{J}' \hat{\Sigma}^{-1} \mathbf{J})^{-1} \mathbf{J}' \hat{\Sigma}^{-1} \bar{\mathbf{y}},$$

where  $\bar{\mathbf{y}} = n^{-1} \sum_{i=1}^n \mathbf{Y}_i$  and

$$\hat{\Sigma} = (n-1)^{-1} \sum_{i=1}^n (\mathbf{Y}_i - \bar{\mathbf{y}})(\mathbf{Y}_i - \bar{\mathbf{y}})'.$$

Compare the variance of the limiting distribution of  $\hat{\mu}_g$  with that of  $\tilde{\mu} = k^{-1} \mathbf{J}' \bar{\mathbf{y}}$ .

30. Prove the following.

**Corollary 5.1.6.2.** Let  $\{X_n\}$  be a sequence of scalar random variables such that

$$X_n = a + O_p(r_n),$$

where  $r_n \rightarrow 0$  as  $n \rightarrow \infty$ . If  $g(x)$  and  $g'(x)$  are continuous at  $a$ , then

$$g(X_n) = g(a) + g'(a)(X_n - a) + o_p(r_n).$$

31. Let

$$Y_t = e^{\beta t} + a_t, \quad t = 0, 1, 2, \dots,$$

where  $\beta \in (-\infty, 0)$  and  $a_i \sim NI(0, \sigma^2)$ . Let  $\hat{\beta}$  be the value of  $\beta$  that minimizes

$$Q_n(\beta) = \sum_{i=0}^n (Y_i - e^{\beta t_i})^2$$

for  $\beta \in (-\infty, 0)$ . Are the conditions of Theorem 5.5.1 satisfied for  $M_n = 1$  and  $a_n = n^\zeta$  for some  $\zeta > 0$ ? What do you conclude about the consistency of  $\hat{\beta}$ ?

## CHAPTER 6

# Estimation of the Mean and Autocorrelations

In this chapter we shall derive some large sample results for the sampling behavior of the estimated mean, covariances, and autocorrelations.

### 6.1. ESTIMATION OF THE MEAN

Consider a stationary time series  $X_t$ , with mean  $\mu$ , which we desire to estimate. If it were possible to obtain a number of independent realizations, then the average of the realization averages would converge in mean square to  $\mu$  as the number of realizations increased. That is, given  $m$  samples of  $n$  observations each,

$$\text{Var}\left\{\frac{1}{m} \sum_{j=1}^m \bar{x}_{(j)}\right\} \leq \frac{1}{m} \gamma_X(0),$$

where  $\bar{x}_{(j)} = n^{-1} \sum_{t=1}^n X_{(j)t}$  is the mean of the  $n$  observations from the  $j$ th realization.

However, in many areas of application, it is difficult or impossible to obtain multiple realizations. For example, most economic time series constitute a single realization. Therefore, the question becomes whether or not we can use the average of a single realization to estimate the mean. Clearly, the sample mean  $\bar{x}_n$  is an unbiased estimator for the mean of a covariance stationary time series. If the mean square error of the sample mean as an estimator of the population mean approaches zero as the number of observations included in the mean increases, we say that the time series is *ergodic* for the mean. We now investigate conditions under which the time series is ergodic for the mean. Theorem 6.1.1 demonstrates that the sample mean may be a consistent estimator for nonstationary time series if the nonstationarity is of a transient nature. The theorem follows Parzen (1962).

**Theorem 6.1.1.** Let  $\{X_t: t \in (1, 2, \dots)\}$  be a time series satisfying

$$\lim_{t \rightarrow \infty} E\{X_t\} = \mu,$$

$$\lim_{n \rightarrow \infty} \text{Cov}\{\bar{x}_n, X_n\} = 0,$$

where  $\bar{x}_n = n^{-1} \sum_{t=1}^n X_t$ . Then

$$\lim_{n \rightarrow \infty} E\{(\bar{x}_n - \mu)^2\} = 0.$$

**Proof.** Now

$$E\{(\bar{x}_n - \mu)^2\} = \text{Var}\{\bar{x}_n\} + \left( \frac{1}{n} \sum_{t=1}^n E\{X_t\} - \mu \right)^2$$

where the second term on the right converges to zero by Lemma 3.1.5. Furthermore,

$$\begin{aligned} \text{Var}\{\bar{x}_n\} &= \frac{1}{n^2} \sum_{i=1}^n \sum_{j=1}^n \text{Cov}\{X_i, X_j\} \\ &= \frac{2}{n^2} \sum_{i=1}^n \sum_{j=1}^i \text{Cov}\{X_i, X_j\} - \frac{1}{n^2} \sum_{i=1}^n \text{Var}\{X_i\} \\ &= \frac{2}{n^2} \sum_{i=1}^n i \text{Cov}\{\bar{x}_i, X_i\} - \frac{1}{n^2} \sum_{i=1}^n \text{Var}\{X_i\} \\ &\leq \frac{2}{n} \sum_{i=1}^n |\text{Cov}\{\bar{x}_i, X_i\}|, \end{aligned}$$

which also converges to zero by Lemma 3.1.5. ▲

**Corollary 6.1.1.1.** Let  $\{X_t\}$  be a stationary time series whose covariance function  $\gamma(h)$  converges to zero as  $h$  gets large. Then

$$\lim_{n \rightarrow \infty} \text{Var}\{\bar{x}_n\} = 0.$$

**Proof.** By Lemma 3.1.5, the convergence of  $\gamma(h)$  to zero implies that  $\text{Cov}\{\bar{x}_n, X_n\} = (1/n) \sum_{h=0}^{n-1} \gamma(h)$  converges to zero, and the result follows by Theorem 6.1.1. ▲

**Corollary 6.1.1.2.** A stationary time series with absolutely summable covariance function is ergodic for the mean. Furthermore,

$$\lim_{n \rightarrow \infty} n \text{Var}\{\bar{x}_n\} = \sum_{h=-\infty}^{\infty} \gamma(h).$$

**Proof.** The assumption of absolute summability implies

$$\lim_{h \rightarrow \infty} \gamma(h) = 0,$$

and ergodicity follows from Corollary 6.1.1.1. We have

$$n \operatorname{Var}\{\bar{x}_n\} = \frac{1}{n} \sum_{j=1}^n \sum_{i=1}^n \gamma(i-j) = \frac{1}{n} \sum_{h=-n+1}^{n-1} (n - |h|) \gamma(h),$$

and  $n \operatorname{Var}\{\bar{x}_n\}$  converges to the stated result by Lemma 3.1.4.  $\blacktriangle$

**Theorem 6.1.2.** If the spectral density of a stationary time series  $X_t$  is continuous, then

$$\lim_{n \rightarrow \infty} n \operatorname{Var}\{\bar{x}_n\} = 2\pi f(0), \quad (6.1.1)$$

where  $f(0)$  is the spectral density of  $X_t$  evaluated at zero.

**Proof.** By Theorem 3.1.10 the Fourier series of a continuous periodic function is uniformly summable by the method of Cesàro. The autocovariances  $\gamma(k)$  are equal to  $\pi$  times the  $a_k$  of that theorem. Therefore,

$$\begin{aligned} 2\pi f(0) &= \lim_{n \rightarrow \infty} 2\pi \left[ \frac{a_0}{2} + \frac{1}{n} \sum_{r=2}^n \sum_{k=1}^{r-1} a_k \right] \\ &= \lim_{n \rightarrow \infty} \left[ \gamma(0) + \frac{2}{n} \sum_{r=2}^n \sum_{k=1}^{r-1} \gamma(k) \right] \\ &= \lim_{n \rightarrow \infty} \left[ \gamma(0) + \frac{2}{n} \sum_{r=1}^n (n-r)\gamma(r) \right] \\ &= \lim_{n \rightarrow \infty} n \operatorname{Var}\{\bar{x}_n\}. \end{aligned} \quad \blacktriangle$$

Since the absolute summability of the covariance function implies that  $f(\omega)$  is continuous, it follows that (6.1.1) holds for a time series with absolutely summable covariance function. Thus, for a wide class of time series, the sample mean has a variance that is declining at the rate  $n^{-1}$ . In large samples the variance is approximately the spectral density evaluated at zero multiplied by  $2\pi n^{-1}$ .

To investigate the efficiency of the sample mean as an estimator of  $\mu$ , we write

$$Y_t = \mu + Z_t,$$

where  $Z_t$  is a time series with zero mean, and we define  $V$  to be the covariance matrix for  $n$  observations on  $Y_t$ . Thus,

$$V = E\{zz'\},$$

where  $z$  is the column vector of  $n$  observations on  $Z_t$ . If the covariance matrix is

known and nonsingular the best linear unbiased estimator of the mean is given by

$$\hat{\mu} = (\mathbf{1}' \mathbf{V}^{-1} \mathbf{1})^{-1} \mathbf{1}' \mathbf{V}^{-1} \mathbf{y}, \quad (6.1.2)$$

where  $\mathbf{1}$  is a column vector composed of  $n$  ones and  $\mathbf{y}$  is the vector of  $n$  observations on  $Y_t$ . The variance of  $\hat{\mu}$  is

$$\text{Var}\{\hat{\mu}\} = (\mathbf{1}' \mathbf{V}^{-1} \mathbf{1})^{-1}, \quad (6.1.3)$$

whereas the variance of  $\bar{y}_n = n^{-1} \sum_{t=1}^n Y_t$  is

$$\text{Var}\{\bar{y}_n\} = n^{-2} \mathbf{1}' \mathbf{V} \mathbf{1}. \quad (6.1.4)$$

Let  $Y_t$  be a  $p$ th order stationary autoregressive process defined by

$$(Y_t - \mu) + \sum_{j=1}^p \alpha_j (Y_{t-j} - \mu) = e_t, \quad (6.1.5)$$

where the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables, and the roots of the characteristic equation are less than one in absolute value. For known  $\alpha_j$ , the estimator (6.1.2) can be constructed by transforming the observations into a sequence of uncorrelated constant variance observations. Using the Gram-Schmidt orthogonalization procedure, we obtain

$$\begin{aligned} W_1 &= \delta_{11} Y_1, \\ W_2 &= \delta_{22} Y_2 - \delta_{21} Y_1, \\ &\vdots \\ W_p &= \delta_{pp} Y_p - \sum_{j=1}^{p-1} \delta_{p,p-j} Y_{p-j}, \\ W_t &= Y_t + \sum_{j=1}^p \alpha_j Y_{t-j}, \quad t = p+1, p+2, \dots, n, \end{aligned} \quad (6.1.6)$$

where  $\delta_{11} = \gamma^{-1/2}(0)\sigma$ ,  $\delta_{22} = \{[1 - \rho^2(1)]\gamma(0)\}^{-1/2}\sigma$ ,  $\delta_{21} = \rho(1)\{[1 - \rho^2(1)]\gamma(0)\}^{-1/2}\sigma$ , and so forth. The expected values are

$$E\{W_1\} = \delta_{11} \mu,$$

$$E\{W_2\} = (\delta_{22} - \delta_{21})\mu,$$

$$E\{W_p\} = \left( \delta_{pp} - \sum_{j=1}^{p-1} \delta_{p,p-j} \right) \mu, \\ E\{W_t\} = \left( 1 + \sum_{j=1}^p \alpha_j \right) \mu, \quad t = p+1, p+2, \dots, n.$$

In matrix notation we let  $\mathbf{T}$  denote the transformation defined in (6.1.6). Then  $\mathbf{T}'\mathbf{T}\sigma^{-2} = \mathbf{V}^{-1}$ ,  $E\{\mathbf{Ty}\} = \mathbf{T}\mathbf{1}\mu$ , and

$$\hat{\mu} = (\mathbf{1}'\mathbf{T}'\mathbf{T}\mathbf{1})^{-1} \mathbf{1}'\mathbf{T}'\mathbf{Ty}. \quad (6.1.7)$$

With the aid of this transformation, we can demonstrate that the sample mean has the same asymptotic efficiency as the best linear unbiased estimator.

**Theorem 6.1.3.** Let  $Y_t$  be the stationary  $p$ th order autoregressive process defined in (6.1.5). Then

$$\lim_{n \rightarrow \infty} n \operatorname{Var}\{\bar{y}_n\} = \lim_{n \rightarrow \infty} n \operatorname{Var}\{\hat{\mu}\},$$

where  $\hat{\mu}$  is defined in (6.1.2).

**Proof.** Without loss of generality, we let  $\sigma^2 = 1$ . The spectral density of  $Y_t$  is then

$$f_Y(\omega) = \frac{1}{2\pi} \left[ \sum_{j=0}^p \alpha_j e^{-i\omega j} \sum_{j=0}^p \alpha_j e^{i\omega j} \right]^{-1},$$

where  $\alpha_0 = 1$ . With the exception of  $2p^2$  terms in the upper left and lower right corners of  $\mathbf{T}'\mathbf{T}$ , the elements of  $\mathbf{V}^{-1}$  are given by

$$v^{ir} = \begin{cases} \sum_{r=0}^{p-h} \alpha_r \alpha_{r+h}, & |i-r| = h \leq p, \\ 0 & \text{otherwise.} \end{cases} \quad (6.1.8)$$

The values of the elements  $v^{ir}$  in (6.1.8) depend only on  $|i-r|$ , and we recognize  $\sum_{r=0}^{|h|} \alpha_r \alpha_{r+|h|} = \gamma_m(h)$ , say, as the covariance function of a  $p$ th order moving average. Therefore, for  $n > 2p$ ,

$$\begin{aligned} \frac{1}{n} \mathbf{1}' \mathbf{V}^{-1} \mathbf{1} &= \frac{1}{n} \sum_{i=1}^n \sum_{r=1}^n v^{ir} \\ &= \frac{1}{n} \sum_{i=1}^n \sum_{r=1}^n \gamma_m(i-r) + \frac{1}{n} \sum_{i=1}^p \sum_{r=1}^p [v^{ir} - \gamma_m(i-r)] \\ &\quad + \frac{1}{n} \sum_{i=n-p+1}^n \sum_{r=n-p+1}^n [v^{ir} - \gamma_m(i-r)] \end{aligned}$$

$$= \frac{1}{n} \sum_{i=1}^n \sum_{r=1}^n \gamma_m(i-r) + O(n^{-1}).$$

It follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} n \operatorname{Var}\{\hat{\mu}\} &= \lim_{n \rightarrow \infty} \left[ \frac{1}{n} \sum_{i=1}^n \sum_{r=1}^n \gamma_m(i-r) \right]^{-1} = \left| \sum_{j=0}^p \alpha_j \right|^{-2} \\ &= 2\pi f_Y(0) = \lim_{n \rightarrow \infty} n \operatorname{Var}\{\bar{y}_n\}. \end{aligned}$$
▲

Using the fact that a general class of spectral densities can be approximated by the spectral density of an autoregressive process (see Theorem 4.3.4), Grenander and Rosenblatt (1957) have shown that the mean and certain other linear estimators have the same asymptotic efficiency as the generalized least squares estimator for time series with spectral densities in that class.

## 6.2. ESTIMATORS OF THE AUTOCOVARIANCE AND AUTOCORRELATION FUNCTIONS

While the sample mean is a natural estimator to consider for the mean of a stationary time series, a number of estimators have been proposed for the covariance function. If the mean is known and, without loss of generality, taken to be zero, the estimator

$$\tilde{\gamma}(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} X_t X_{t+h} \quad (6.2.1)$$

is seen to be the mean of  $n-h$  observations from the time series, say,

$$Z_{th} = X_t X_{t+h}.$$

For stationary time series,  $E[Z_{th}] = \gamma_X(h)$  for all  $t$ , and it follows that  $\tilde{\gamma}(h)$  is an unbiased estimator of  $\gamma(h)$ . In most practical situations the mean is unknown and must be estimated. We list below two possible estimators of the covariance function when the mean is estimated. In both expressions  $h$  is taken to be greater than or equal to zero:

$$\gamma^*(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} (X_t - \bar{x}_n)(X_{t+h} - \bar{x}_n), \quad (6.2.2)$$

$$\hat{\gamma}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{x}_n)(X_{t+h} - \bar{x}_n). \quad (6.2.3)$$

It is clear that these estimators differ by factors that become small at the rate  $n^{-1}$ . Unlike  $\tilde{\gamma}(h)$ , neither of the estimators is unbiased. The bias is given in Theorem 6.2.2 of this section.

The estimator  $\hat{\gamma}(h)$  can be shown to have smaller mean square error than  $\gamma^t(h)$  for certain types of time series. This estimator also has the advantage of guaranteeing positive definiteness of the estimated covariance function. In most of our applications we shall use the estimator  $\hat{\gamma}(h)$ .

As one might expect, the variances of the estimated autocovariances are much more complicated than those of the mean. The theorem we present is due to Bartlett (1946). A result needed in the proof will be useful in later sections, and we state it as a lemma.

**Lemma 6.2.1.** Let  $\{\delta_j\}_{j=-\infty}^{\infty}$  and  $\{c_j\}_{j=-\infty}^{\infty}$  be two absolutely summable sequences. Then, for fixed integers  $r, h$ , and  $d, d \geq 0$ ,

$$\begin{aligned}\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{s=1}^{n+d} \sum_{t=1}^n \delta_{s-t+r} c_{s-t+h} &= \sum_{p=-\infty}^{\infty} \delta_{p+r} c_{p+h} \\ &= \sum_{p=-\infty}^{\infty} \delta_p c_{p+h-r}.\end{aligned}$$

**Proof.** Let  $p = s - t$ . Then

$$\begin{aligned}\frac{1}{n} \sum_{s=1}^{n+d} \sum_{t=1}^n \delta_{s-t+r} c_{s-t+h} &= \frac{1}{n} \sum_{p=0}^{n-1} \sum_{s=p+1}^n \delta_{p+r} c_{p+h} \\ &\quad + \frac{1}{n} \sum_{p=-(n-1)}^{-1} \sum_{s=1}^{n+p} \delta_{p+r} c_{p+h} \\ &\quad + \frac{1}{n} \sum_{s=n+1}^{n+d} \sum_{t=1}^n \delta_{s-t+r} c_{s-t+h} \\ &= \sum_{p=-(n-1)}^{n-1} \frac{(n-|p|)}{n} \delta_{p+r} c_{p+h} \\ &\quad + \frac{1}{n} \sum_{s=n+1}^{n+d} \sum_{t=1}^n \delta_{s-t+r} c_{s-t+h}.\end{aligned}$$

Now,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{s=n+1}^{n+d} \sum_{t=1}^n \delta_{s-t+r} c_{s-t+h} \leq \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{p=1}^{n+d} d |\delta_{p+r}| |c_{p+h}| = 0$$

with the inequality resulting from the inclusion of additional terms and the introduction of absolute values.  $\blacktriangleleft$

**Theorem 6.2.1.** Let the time series  $\{X_t\}$  be defined by

$$X_t = \sum_{j=-\infty}^{\infty} \alpha_j e_{t-j},$$

where the sequence  $\{\alpha_j\}$  is absolutely summable and the  $e_i$  are independent  $(0, \sigma^2)$  random variables with  $E\{e_i^4\} = \eta\sigma^4$ . Then, for fixed  $h$  and  $q$  ( $h \geq q \geq 0$ ),

$$\begin{aligned} & \lim_{n \rightarrow \infty} (n - q) \text{Cov}\{\tilde{\gamma}(h), \tilde{\gamma}(q)\} \\ &= (\eta - 3)\gamma(h)\gamma(q) + \sum_{p=-\infty}^{\infty} [\gamma(p)\gamma(p-h+q) + \gamma(p+q)\gamma(p-h)], \end{aligned} \quad (6.2.4)$$

where  $\tilde{\gamma}(h)$  is defined in (6.2.1).

**Proof.** Using

$$E\{e_t e_u e_v e_w\} = \begin{cases} \eta\sigma^4, & t = u = v = w, \\ \sigma^4 & \text{if subscripts are equal in} \\ & \text{pairs but not all equal} \\ 0 & \text{otherwise,} \end{cases}$$

it follows that

$$\begin{aligned} E\{X_t X_{t+h} X_{t+h+p} X_{t+h+p+q}\} &= (\eta - 3)\sigma^4 \sum_{j=-\infty}^{\infty} \alpha_j \alpha_{j+h} \alpha_{j+h+p} \alpha_{j+h+p+q} \\ &\quad + \gamma(h)\gamma(q) + \gamma(h+p)\gamma(p+q) \\ &\quad + \gamma(h+p+q)\gamma(p). \end{aligned} \quad (6.2.5)$$

Thus,

$$\begin{aligned} & E\{\tilde{\gamma}(h)\tilde{\gamma}(q)\} - \gamma(h)\gamma(q) \\ &= \frac{1}{(n-h)(n-q)} E\left\{ \sum_{s=1}^{n-q} \sum_{t=1}^{n-h} X_s X_{s+h} X_t X_{t+h} \right\} - \gamma(h)\gamma(q) \\ &= \frac{(\eta - 3)\sigma^4}{(n-h)(n-q)} \sum_{s=1}^{n-q} \sum_{t=1}^{n-h} \sum_{j=-\infty}^{\infty} \alpha_j \alpha_{j+h} \alpha_{j+s-t} \alpha_{j+s-t+h} \\ &\quad + \frac{1}{(n-h)(n-q)} \sum_{s=1}^{n-q} \sum_{t=1}^{n-h} [\gamma(s-t)\gamma(s-t-h+q) \\ &\quad \quad + \gamma(s-t+q)\gamma(s-t-h)]. \end{aligned} \quad (6.2.6)$$

Applying Lemma 6.2.1 to equation (6.2.6), we have

$$\begin{aligned} & \lim_{n \rightarrow \infty} (n - q) \text{Cov}\{\tilde{\gamma}(h), \tilde{\gamma}(q)\} \\ &= (\eta - 3)\sigma^4 \sum_{j=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} \alpha_j \alpha_{j+h} \alpha_{j+p} \alpha_{j+p+q} \\ &\quad + \sum_{p=-\infty}^{\infty} [\gamma(p)\gamma(p-h+q) + \gamma(p+q)\gamma(p-h)]. \end{aligned} \quad \blacktriangle$$

For normal time series  $\eta = 3$ , and we have

$$\text{Cov}\{\tilde{\gamma}(h), \tilde{\gamma}(q)\} = \frac{1}{n-h} \sum_{p=-\infty}^{\infty} [\gamma(p)\gamma(p-h+q) + \gamma(p+q)\gamma(p-h)].$$

**Corollary 6.2.1.1.** Given a time series  $\{e_t : t \in (0, \pm 1, \pm 2, \dots)\}$ , where the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables, then for  $h, q \geq 0$ ,

$$\text{Cov}\{\tilde{\gamma}_e(h), \tilde{\gamma}_e(q)\} = \begin{cases} \frac{2\sigma^4}{n}, & h = q = 0, \\ \frac{\sigma^4}{n-h}, & h = q \neq 0, \\ 0 & \text{otherwise.} \end{cases}$$

**Proof.** Reserved for the reader. ▲

In Theorem 6.2.1 the estimator was constructed assuming the mean to be known. As we have mentioned, the estimation of the unknown mean introduces a bias into the estimated covariance. However, the variance formulas presented in Theorem 6.2.1 remain valid approximations for the estimator defined in equation (6.2.2).

**Theorem 6.2.2.** Given fixed  $h \geq q \geq 0$  and a time series  $X_t$  satisfying the assumptions of Theorem 6.2.1,

$$E\{\hat{\gamma}(h) - \gamma(h)\} = -\frac{|h|}{n} \gamma(h) - \frac{n-|h|}{n} \text{Var}\{\bar{x}_n\} + O(n^{-2})$$

and

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{n^2}{(n-h)} \text{Cov}\{\hat{\gamma}(h), \hat{\gamma}(q)\} \\ = (\eta - 3)\gamma(h)\gamma(q) + \sum_{p=-\infty}^{\infty} [\gamma(p)\gamma(p-h+q) + \gamma(p+q)\gamma(p-h)]. \end{aligned}$$

**Proof.** From the definition of  $\hat{\gamma}(h)$ , we obtain

$$\begin{aligned} \hat{\gamma}(h) &= \frac{1}{n} \left[ \sum_{t=1}^{n-h} X_t X_{t+h} - \bar{x}_n \sum_{t=1}^{n-h} (X_t + X_{t+h}) + (n-h)\bar{x}_n^2 \right] \\ &= \frac{n-h}{n} \tilde{\gamma}(h) + \frac{\bar{x}_n}{n} \left( \sum_{t=1}^h X_t + \sum_{t=n-h+1}^n X_t - 2h\bar{x}_n \right) - \frac{n-h}{n} \bar{x}_n^2. \end{aligned}$$

Now

$$\left| E\left\{ \bar{x}_n \left( \sum_{t=1}^h X_t + \sum_{t=n-h+1}^n X_t - 2h\bar{x}_n \right) \right\} \right| \leq \frac{2h}{n} \sum_{j=-\infty}^{\infty} |\gamma(j)|,$$

and we have the first result. Since by an application of (6.2.5) we have  $\text{Var}\{\tilde{x}_n\} = O(n^{-2})$ , the second conclusion also follows.  $\blacktriangle$

Using Theorems 6.2.1 and 6.2.2 and the results of Chapter 5, we can approximate the mean and variance of the estimated correlation function. If the mean is known, we consider the estimated autocorrelation

$$\tilde{r}(h) = [\tilde{\gamma}(0)]^{-1} \tilde{\gamma}(h),$$

and if the mean is unknown, the estimator

$$\hat{r}(h) = [\hat{\gamma}(0)]^{-1} \hat{\gamma}(h). \quad (6.2.7)$$

If the denominator of the estimator is zero, we define the estimator to be zero.

**Theorem 6.2.3.** Let the time series  $\{X_t\}$  be defined by

$$X_t = \sum_{j=-\infty}^{\infty} \alpha_j e_{t-j},$$

where the sequence  $\{\alpha_j\}$  is absolutely summable, and the  $e_i$  are independent  $(0, \sigma^2)$  random variables with  $E\{e_i^6\} = \eta\sigma^6$ . Then for fixed  $h$  and  $q$ ,

$$\begin{aligned} \text{Cov}\{\tilde{r}(h), \tilde{r}(q)\} &= [\gamma(0)]^{-2} [\text{Cov}\{\tilde{\gamma}(h), \tilde{\gamma}(q)\} - \rho(h)\text{Cov}\{\tilde{\gamma}(0), \tilde{\gamma}(q)\} \\ &\quad - \rho(q)\text{Cov}\{\tilde{\gamma}(0), \tilde{\gamma}(h)\} + \rho(h)\rho(q)\text{Var}\{\tilde{\gamma}(0)\}] \\ &\quad + O(n^{-2}) \\ &= \frac{1}{n} \sum_{p=-\infty}^{\infty} [\rho(p)\rho(p-h+q) + \rho(p+q)\rho(p-h) \\ &\quad - 2\rho(q)\rho(p)\rho(p-h) - 2\rho(h)\rho(p)\rho(p-q) \\ &\quad + 2\rho(h)\rho(q)\rho^2(p)] + O(n^{-2}), \end{aligned} \quad (6.2.8)$$

$$\text{Cov}\{\hat{r}(h), \hat{r}(q)\} = \text{Cov}\{\tilde{r}(h), \tilde{r}(q)\} + O(n^{-2}),$$

$$\begin{aligned} E\{\tilde{r}(h) - \rho(h)\} &= [\gamma(0)]^{-2} [\rho(h)\text{Var}\{\tilde{\gamma}(0)\} - \text{Cov}\{\tilde{\gamma}(0), \tilde{\gamma}(h)\}] \\ &\quad + O(n^{-2}), \end{aligned}$$

$$\begin{aligned} E\{\hat{r}(h)\} &= \frac{n-h}{n} \rho(h) - [\gamma(0)]^{-1} [1 - \rho(h)] \text{Var}\{\tilde{x}_n\} \\ &\quad + [\gamma(0)]^{-2} [\rho(h)\text{Var}\{\tilde{\gamma}(0)\} - \text{Cov}\{\tilde{\gamma}(h), \tilde{\gamma}(0)\}] \\ &\quad + O(n^{-2}). \end{aligned}$$

**Proof.** The estimated autocorrelations are bounded and are differentiable functions of the estimated covariances on a closed set containing the true parameter vector as an interior point. Furthermore, the derivatives are bounded on that set. Hence, the conditions of Theorem 5.4.3 are met with  $\{\tilde{\gamma}(h), \tilde{\gamma}(q), \tilde{\gamma}(0)\}$

playing the role of  $\{X_n\}$  of that theorem. Since the function  $\tilde{r}(h)$  is bounded, we take  $\alpha = 1$ .

Expanding  $[\tilde{r}(h) - \rho(h)][\tilde{r}(q) - \rho(q)]$  through third order terms and using Theorem 5.4.1 to establish that the expected value of the third order moments is  $O(n^{-2})$ , we have the result (6.2.8). The remaining results are established in a similar manner.  $\Delta$

For the first order autoregressive time series  $X_t = \rho X_{t-1} + e_t$ , it is relatively easy to evaluate the variances of the estimated autocorrelations. We have, for  $h > 0$ ,

$$\begin{aligned}\text{Var}\{\hat{r}(h)\} &\doteq \frac{n-h}{n^2} \sum_{p=-\infty}^{\infty} (\rho^{|2p|} + \rho^{|p+h|+|p-h|} - 4\rho^{|h|}\rho^{|p|+|p-h|} + 2\rho^{|2h|}\rho^{|2p|}) \\ &= \frac{n-h}{n^2} \left[ \frac{(1+\rho^2)(1-\rho^{2h})}{1-\rho^2} - 2h\rho^{2h} \right].\end{aligned}\quad (6.2.9)$$

We note that for large  $h$ ,

$$\text{Var}\{\hat{r}(h)\} \doteq \frac{n-h}{n^2} \frac{1+\rho^2}{1-\rho^2}. \quad (6.2.10)$$

For a time series where the correlations approach zero rapidly, the variance of  $\hat{r}(h)$  for large  $h$  can be approximated by the first term of (6.2.8). That is, for such a time series and for  $h$  such that  $\rho(h) \doteq 0$ , we have

$$\text{Var}\{\hat{r}(h)\} \doteq \frac{1}{n} \sum_{p=-\infty}^{\infty} \rho^2(p). \quad (6.2.11)$$

We are particularly interested in the behavior of the estimated autocorrelations for a time series of independent random variables, since this is often a working hypothesis in time series analysis.

**Corollary 6.2.3.** Let  $\{e_t\}$  be a sequence of independent  $(0, \sigma^2)$  random variables with sixth moment  $\eta\sigma^6$ . Then, for  $h \geq q > 0$ ,

$$E\{\hat{r}(h)\} = -\frac{n-h}{n(n-1)} + O(n^{-2}),$$

$$\text{Cov}\{\hat{r}(h), \hat{r}(q)\} = \begin{cases} \frac{n-h}{n^2} + O(n^{-2}), & h = q \neq 0, \\ O(n^{-2}) & \text{otherwise.} \end{cases}$$

**Proof.** Omitted.  $\Delta$

For large  $n$  the bias in  $\hat{r}(h)$  is negligible. However, it is easy to reduce the bias

in the null case of independent random variables. It is suggested that the estimator

$$\hat{\rho}(h) = \hat{r}(h) + \frac{n-h}{(n-1)^2} \{1 - [\hat{r}(h)]^2\} \quad (6.2.12)$$

be used for hypothesis testing when only a small number of observations are available. For time series of the type specified in Corollary 6.2.3 and  $h, q > 0$ ,

$$E\{\hat{\rho}(h)\} = O(n^{-2}),$$

$$\text{Cov}\{\hat{\rho}(h), \hat{\rho}(q)\} = \begin{cases} \frac{n-h}{n^2} + O(n^{-2}), & h = q > 0, \\ O(n^{-2}) & \text{otherwise.} \end{cases}$$

In the next section we prove that the  $\hat{r}(h)$  and  $\hat{\rho}(h)$  are approximately normally distributed. The approximate distribution of the autocorrelations will be adequate for most purposes, but we mention one exact distributional result. A statistic closely related to the first order autocorrelation is the von Neumann ratio:<sup>1</sup>

$$d_v = \frac{\sum_{t=2}^n (X_t - X_{t-1})^2}{\sum_{t=1}^n (X_t - \bar{x}_n)^2}. \quad (6.2.13)$$

We see that

$$\begin{aligned} d_v &= \frac{\sum_{t=2}^n (X_t - \bar{x}_n)^2 - 2 \sum_{t=2}^n (X_t - \bar{x}_n)(X_{t-1} - \bar{x}_n) + \sum_{t=1}^{n-1} (X_t - \bar{x}_n)^2}{\sum_{t=1}^n (X_t - \bar{x}_n)^2} \\ &= 2 \left[ 1 - \hat{r}(1) - \frac{(X_1 - \bar{x}_n)^2 + (X_n - \bar{x}_n)^2}{2 \sum_{t=1}^n (X_t - \bar{x}_n)^2} \right]. \end{aligned}$$

If the  $X_t$  are normal independent  $(\mu, \sigma^2)$  random variables, it is possible to show that  $E\{d_v\} = 2$ . Therefore,  $r_v = \frac{1}{2}(d_v - 2)$  is an unbiased estimator of zero in that case. von Neumann (1941) and Hart (1942) have given the exact distribution of  $r_v$  under the assumption that  $X_t$  is a sequence of normal independent  $(\mu, \sigma^2)$  random variables. Tables of percentage points are given in Hart (1942) and in Anderson (1971, p. 345). Inspection of these tables demonstrates that the

<sup>1</sup> The ratio is sometimes defined with the multiplier  $n/(n-1)$ .

percentage points of  $t_v = r_v(n+1)^{1/2}(1-r_v^2)^{-1/2}$  are approximately those of Student's  $t$  with  $n+3$  degrees of freedom for  $n$  greater than 10.

Clearly the distribution of  $\hat{\rho}(1)[1 - \hat{\rho}^2(1)]^{-1/2}(n+1)^{1/2}$ , where  $\hat{\rho}(1)$  is defined in (6.2.12), is close to the distribution of  $t_v$  and therefore may also be approximated by Student's  $t$  with  $n+3$  degrees of freedom when the observations are independent normal random variables.

Kendall and Stuart (1966) and Anderson (1971) present discussions of the distributional theory of statistics such as  $d_v$ .

### 6.3. CENTRAL LIMIT THEOREMS FOR STATIONARY TIME SERIES

The results of this chapter have, so far, been concerned with the mean and variance of certain sample statistics computed from a single realization. In order to perform tests of hypotheses or set confidence limits for the underlying parameters, some distribution theory is required.

That the mean of a finite moving average is, in the limit, normally distributed is a simple extension of the central limit theorems of Section 5.3.

**Proposition 6.3.1.** Let  $\{X_t: t \in (0, \pm 1, \pm 2, \dots)\}$  be defined by

$$X_t = \mu + \sum_{j=0}^m b_j e_{t-j},$$

where  $b_0 = 1$ ,  $\sum_{j=0}^m b_j \neq 0$ , and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables. Assume  $n^{1/2}\bar{e}_n$  converges in distribution to a  $N(0, \sigma^2)$  random variable. Then,

$$n^{1/2}(\bar{x}_n - \mu) \xrightarrow{\mathcal{D}} N\left(0, \sigma^2 \left[ \sum_{j=0}^m b_j \right]^2\right).$$

**Proof.** We have

$$\begin{aligned} n^{1/2}(\bar{x}_n - \mu) &= n^{-1/2} \sum_{t=1}^n (X_t - \mu) \\ &= n^{-1/2} \sum_{t=1}^n (e_t + b_1 e_{t-1} + b_2 e_{t-2} + \cdots + b_m e_{t-m}) \\ &= n^{-1/2} \sum_{j=0}^m b_j \sum_{t=1}^n e_t + n^{-1/2} \sum_{s=1}^m \sum_{j=s}^m b_j e_{1-s} \\ &\quad - n^{-1/2} \sum_{s=0}^{m-1} \sum_{j=s+1}^m b_j e_{n-s}. \end{aligned}$$

Both  $n^{-1/2} \sum_{s=0}^{m-1} \sum_{j=s+1}^m b_j e_{n-s}$  and  $n^{-1/2} \sum_{s=1}^m \sum_{j=s}^m b_j e_{1-s}$  converge in probability to zero as  $n$  increases, by Chebyshev's inequality.

Therefore, the limiting distribution of  $n^{1/2}(\bar{x}_n - \mu)$  is the limiting distribution of  $n^{1/2} \sum_{j=0}^m b_j \bar{e}_n$ , and the result follows.  $\blacktriangle$

If  $\sum_{j=0}^m b_j = 0$ , then  $\bar{x}_n - \mu$  has a variance that approaches zero at a rate faster than  $n^{-1}$ . The reader can verify this by considering, for example, the time series  $X_t = \mu + e_t - e_{t-1}$ . In such a case the theorem holds in the sense that  $n^{1/2}(\bar{x}_n - \mu)$  is converging to the singular (zero variance) normal distribution.

Moving average time series of independent random variables are special cases of a more general class of time series called *m-dependent*.

**Definition 6.3.1.** The sequence of random variables  $\{Z_t: t \in (0, \pm 1, \pm 2, \dots)\}$  is said to be *m-dependent* if  $s - r > m$ , where  $m$  is a positive integer, implies that the two sets

$$(\dots, Z_{r-2}, Z_{r-1}, Z_r), \quad (Z_s, Z_{s+1}, Z_{s+2}, \dots)$$

are independent.

We give a theorem for such time series due to Hoeffding and Robbins (1948).

**Theorem 6.3.1.** Let  $\{Z_t: t \in (1, 2, \dots)\}$  be a sequence of *m-dependent* random variables with  $E\{Z_t\} = 0$ ,  $E\{Z_t^2\} = \sigma_t^2 < \beta < \infty$ , and  $E\{|Z_t|^{2+2\delta}\} \leq \beta^{2+2\delta}$  for some  $\delta > 0$ . Let the limit

$$\lim_{p \rightarrow \infty} p^{-1} \sum_{j=1}^p A_{t+j} = A,$$

$A \neq 0$ , be uniform for  $t = 1, 2, \dots$ , where

$$A_t = E[Z_{t+m}^2] + 2 \sum_{j=1}^m E\{Z_{t+m-j} Z_{t+m}\}.$$

Then

$$n^{-1/2} \sum_{i=1}^n Z_i \xrightarrow{\mathcal{D}} N(0, A).$$

**Proof.** Fix an  $\alpha$ ,  $0 < \alpha < 0.25 - \epsilon_\delta$ , where

$$0.25 > \epsilon_\delta > \max\{0, 0.25(1+2\delta)^{-1}(1-2\delta)\}.$$

Let  $k$  be the largest integer less than  $n^\alpha$ , and let  $p$  be the largest integer less than  $k^{-1}n$ . Define

$$Y_i = \sum_{j=1}^{k-m} Z_{(i-1)k+j}, \quad i = 1, 2, \dots, p,$$

$$S_p = n^{-1/2} \sum_{i=1}^p Y_i.$$

Then the difference

$$D_n = n^{-1/2} \sum_{t=1}^n Z_t - S_p = n^{-1/2} \left[ \sum_{i=1}^{p-1} \left( \sum_{j=1}^m Z_{ki-m+j} \right) + \sum_{t=pk-m+1}^n Z_t \right].$$

For  $n$  large enough so that  $k > 2m$ , the sums  $\sum_{j=1}^m Z_{ki-m+j}$ ,  $i = 1, 2, \dots, p$ , are independent. By the assumption on the moments, we have

$$\begin{aligned} \text{Var} \left\{ \sum_{j=1}^m Z_{ki-m+j} \right\} &\leq m^2 \beta^2, \\ \text{Var} \left\{ \sum_{t=pk-m+1}^n Z_t \right\} &\leq (k+m)^2 \beta^2. \end{aligned}$$

It follows that

$$\text{Var}\{D_n\} \leq n^{-1} \beta^2 \{m^2(p-1) + (k+m)^2\} = o(1),$$

and  $n^{-1/2} \sum_{t=1}^n Z_t$  converges in mean square to  $S_p$ . Since  $Z_t$  is correlated only with  $Z_{t-m}, Z_{t-m+1}, \dots, Z_{t+m-1}, Z_{t+m}$ , the addition of a  $Z_t$  to a sum containing the  $m$  preceding terms,  $Z_{t-1}, Z_{t-2}, \dots, Z_{t-m}$ , increases the variance by the amount  $A_{t-m}$ . Therefore,

$$\text{Var}\{Y_i\} = \text{Var} \left\{ \sum_{j=1}^{k-m} Z_{(i-1)k+j} \right\} = \text{Var} \left\{ \sum_{j=1}^m Z_{(i-1)k+j} \right\} + \sum_{j=1}^{k-2m} A_{(i-1)k+j}$$

and

$$\text{Var}\{S_p\} = n^{-1} \sum_{i=1}^p \text{Var} \left\{ \sum_{j=1}^m Z_{(i-1)k+j} \right\} + n^{-1} \sum_{i=1}^p \sum_{j=1}^{k-2m} A_{(i-1)k+j}.$$

Since  $\text{Var}\{\sum_{j=1}^m Z_{(i-1)k+j}\} \leq m^2 \beta^2$ , and  $p^{-1} \sum_{j=1}^p A_{(i-1)k+j}$  converges uniformly, we have

$$\lim_{p \rightarrow \infty} n^{-1} \sum_{i=1}^p \text{Var}\{Y_i\} = \lim_{p \rightarrow \infty} p^{-1} \sum_{i=1}^p \text{Var}(k^{-1/2} Y_i) = A.$$

Now

$$E\{|k^{-1/2} Y_i|^{2+2\delta}\} = k^{1+\delta} E \left\{ \left| k^{-1} \sum_{j=1}^{k-m} Z_{(i-1)k+j} \right|^{2+2\delta} \right\} \leq k^{1+\delta} \beta^{2+2\delta}$$

and

$$\lim_{p \rightarrow \infty} \frac{\sum_{i=1}^p E\{|k^{-1/2} Y_i|^{2+2\delta}\}}{\left( \sum_{i=1}^p E\{|k^{-1/2} Y_i|^2\} \right)^{1+\delta}} \leq \lim_{p \rightarrow \infty} \frac{pk^{1+\delta} \beta^{2+2\delta}}{(pA)^{1+\delta}} = 0.$$

Hence,  $p^{-1/2} \sum_{i=1}^p k^{-1/2} Y_i$  satisfies the conditions of Liapounov's central limit theorem, and the result follows.  $\blacktriangle$

For the  $m$ th order moving average, the  $A_t$  of Theorem 6.3.1 is the same for all  $t$  and is

$$\begin{aligned}\gamma(0) + 2 \sum_{j=1}^m \gamma(j) &= \sum_{j=0}^m b_j^2 \sigma^2 + 2 \sum_{j=1}^m \sum_{s=0}^{m-j} b_s b_{s+j} \sigma^2 \\ &= \left( \sum_{j=0}^m b_j \right)^2 \sigma^2,\end{aligned}$$

which agrees with Proposition 6.3.1.

The results of Proposition 6.3.1 and Theorem 6.3.1 may be generalized to infinite moving averages of independent random variables. Our proofs follow those of Diananda (1953) and Anderson (1959, 1971). We first state a lemma required in the proofs of the primary theorems.

**Lemma 6.3.1.** Let the random variables  $\xi_n$  with distribution functions  $F_{\xi_n}(z)$  be defined by

$$\xi_n = S_{kn} + D_{kn}$$

for  $k = 1, 2, \dots$  and  $n = 1, 2, \dots$ . Let

$$p\lim_{k \rightarrow \infty} D_{kn} = 0$$

uniformly in  $n$ . Assume

$$F_{S_{kn}}(z) \xrightarrow{C} F_{\psi_k}(z) \quad \text{as } n \rightarrow \infty$$

and

$$F_{\psi_k}(z) \xrightarrow{C} F_{\xi}(z) \quad \text{as } k \rightarrow \infty.$$

Then

$$F_{\xi_n}(z) \xrightarrow{C} F_{\xi}(z) \quad \text{as } n \rightarrow \infty.$$

**Proof.** Let  $z_0$  be a point of continuity of  $F_{\xi}(z)$ , and fix  $\delta > 0$ . There is an  $\epsilon > 0$  such that

$$|F_{\xi}(z) - F_{\xi}(z_0)| < 0.25\delta$$

for  $z$  in  $[z_0 - \epsilon, z_0 + \epsilon]$ , and  $F_{\xi}(z)$  is continuous at  $z_0 - \epsilon$  and at  $z_0 + \epsilon$ . By

hypothesis, there exists a  $K_0$  such that

$$\begin{aligned}|F_{\psi_k}(z_0 - \epsilon) - F_\xi(z_0 - \epsilon)| &< 0.25\delta, \\|F_{\psi_k}(z_0 + \epsilon) - F_\xi(z_0 + \epsilon)| &< 0.25\delta\end{aligned}$$

for  $k > K_0$ . This means that

$$|F_{\psi_k}(z) - F_\xi(z_0)| < 0.5\delta$$

for  $k > K_0$  and  $z_0 - \epsilon \leq z \leq z_0 + \epsilon$ .

Now there is a  $K_1$  such that for  $k \geq K_1$ ,

$$P\{|\xi_n - S_{kn}| > \epsilon\} < 0.25\delta$$

for all  $n$ . By the arguments used in the proof of Theorem 5.2.1, this implies that

$$F_{S_{kn}}(z - \epsilon) - 0.25\delta \leq F_{\xi_n}(z) \leq F_{S_{kn}}(z + \epsilon) + 0.25\delta$$

for all  $z$ .

Fix  $k$  at the maximum of  $K_0$  and  $K_1$ . Let  $z_1$  and  $z_2$  be continuity points of  $F_{\psi_k}(z)$  such that  $z_0 - \epsilon \leq z_1 < z_0$  and  $z_0 < z_2 \leq z_0 + \epsilon$ . Then there exists an  $N$  such that for  $n > N$ ,

$$\begin{aligned}|F_{S_{kn}}(z_1) - F_{\psi_k}(z_1)| &< 0.25\delta, \\|F_{S_{kn}}(z_2) - F_{\psi_k}(z_2)| &< 0.25\delta.\end{aligned}$$

Therefore, for  $n > N$ ,

$$\begin{aligned}F_\xi(z_0) - \delta &\leq F_{\psi_k}(z_1) - 0.5\delta \leq F_{S_{kn}}(z_1) - 0.25\delta \\&\leq F_{\xi_n}(z_0) \leq F_{S_{kn}}(z_2) + 0.25\delta \leq F_{\psi_k}(z_2) + 0.5\delta \\&\leq F_\xi(z_0) + \delta.\end{aligned}\quad \blacktriangle$$

The following two lemmas are the convergence in probability and almost sure results analogous to the convergence in distribution result of Lemma 6.3.1.

**Lemma 6.3.2.** Let  $(\xi_n, S_{kn}, D_{kn})$ ,  $k = 1, 2, \dots$ ,  $n = 1, 2, \dots$ , be random variables defined on the probability space  $(\Omega, \mathcal{A}, P)$  satisfying

$$\xi_n = S_{kn} + D_{kn}. \quad (6.3.1)$$

Assume:

(i) one has

$$\operatorname{plim}_{k \rightarrow \infty} D_{kn} = 0 \quad \text{uniformly in } n, \quad (6.3.2)$$

(ii) for every fixed  $k$  there is an  $S_k$  satisfying

$$\operatorname{plim}_{n \rightarrow \infty} (S_{kn} - S_k) = 0, \quad (6.3.3)$$

(iii) one has

$$\operatorname{plim}_{k \rightarrow \infty} (S_k - \xi) = 0. \quad (6.3.4)$$

Then

$$\operatorname{plim}_{n \rightarrow \infty} (\xi_n - \xi) = 0.$$

**Proof.** Omitted. ▲

**Lemma 6.3.3.** Let the assumption (6.3.1) of Lemma 6.3.2 hold. Replace the assumptions (6.3.2), (6.3.3), and (6.3.4) with the following:

(i) one has

$$\lim_{k \rightarrow \infty} D_{kn} = 0 \quad \text{a.s., uniformly in } n; \quad (6.3.5)$$

(ii) for every fixed  $k$ , there is an  $S_k$  satisfying

$$\lim_{n \rightarrow \infty} (S_{kn} - S_k) = 0 \quad \text{a.s.}; \quad (6.3.6)$$

(iii) one has

$$\lim_{k \rightarrow \infty} (S_k - \xi) = 0 \quad \text{a.s.} \quad (6.3.7)$$

Then

$$\lim_{n \rightarrow \infty} (\xi_n - \xi) = 0 \quad \text{a.s.}$$

**Proof.** Omitted. ▲

We use Lemma 6.3.2. to prove that the sample mean of an infinite moving average converges in probability to the population mean under weak conditions on the  $e_i$ . For example, the conclusion holds for independently identically distributed  $e_i$  with zero mean. Corollary 6.1.1.2 used finite variance to obtain a stronger conclusion.

**Theorem 6.3.2.** Let the time series  $X_t$  be defined by

$$X_t = \mu + \sum_{j=-\infty}^{\infty} w_j e_{t-j},$$

where  $\sum_{j=-\infty}^{\infty} |w_j| < \infty$ . Assume that  $E[|e_t|] < M < \infty$  for all  $t$  and that

$$n^{-1} \sum_{t=1}^n e_t \xrightarrow{P} 0.$$

Then  $\bar{x}_n \xrightarrow{P} \mu$ .

**Proof.** We apply Lemma 6.3.2. Let  $\xi_n = \bar{x}_n - \mu$ ,

$$S_{kn} = n^{-1} \sum_{t=1}^n \sum_{j=-k}^k w_j e_{t-j} = n^{-1} \sum_{j=-k}^k w_j \sum_{t=1}^n e_{t-j}, \quad (6.3.8)$$

and

$$D_{kn} = n^{-1} \sum_{t=1}^n \sum_{|j|>k} w_j e_{t-j}.$$

By the assumption that the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables,  $n^{-1} \sum_{t=1}^n e_{t-j}$  converges to zero in probability as  $n \rightarrow \infty$ . Hence, the sum (6.3.8) converges to zero ( $= S_k$ ) as  $n \rightarrow \infty$  for every fixed  $k$ , and condition (ii) of Lemma 6.3.2 is satisfied. Condition (iii) is trivially satisfied for  $\xi = 0$ . Now

$$\begin{aligned} P\{|D_{kn}| > \epsilon\} &\leq \epsilon^{-1} E\left\{ n^{-1} \sum_{t=1}^n \sum_{|j|>k} |w_j| |e_{t-j}| \right\} \\ &\leq M \epsilon^{-1} \sum_{|j|>k} |w_j| \end{aligned}$$

for some finite  $M$ , by Chebyshev's inequality. Hence, condition (i) of Lemma 6.3.2 is satisfied and the conclusion follows.  $\blacktriangle$

Using Lemma 6.3.1, we show that if the  $e_t$  of an infinite moving average satisfy a central limit theorem, then the moving average also satisfies a central limit theorem.

**Theorem 6.3.3.** Let  $X_t$  be a covariance stationary time series satisfying

$$X_t = \sum_{j=0}^{\infty} \alpha_j e_{t-j},$$

where  $\sum_{j=0}^{\infty} |\alpha_j| < \infty$ ,  $\sum_{j=0}^{\infty} \alpha_j \neq 0$ , and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables. Assume

$$n^{1/2} \bar{e}_n \xrightarrow{\mathcal{L}} N(0, \sigma^2).$$

Then

$$n^{-1/2} \sum_{t=1}^n X_t \xrightarrow{\mathcal{L}} N\left(0, \sum_{h=-\infty}^{\infty} \gamma_X(h)\right),$$

where

$$\sum_{h=-\infty}^{\infty} \gamma_X(h) = \left( \sum_{j=0}^{\infty} \alpha_j \right)^2 \sigma^2.$$

**Proof.** Let

$$(Y_{tk}, W_{tk}) = \left( \sum_{j=0}^k \alpha_j e_{t-j}, \sum_{j=k+1}^{\infty} \alpha_j e_{t-j} \right),$$

and define the normalized sums

$$(S_{kn}, D_{kn}) = n^{-1/2} \sum_{t=1}^n (Y_{tk}, W_{tk}).$$

For a fixed  $k$ ,  $W_{tk}$  is a stationary time series such that

$$\gamma_W(h) = \sum_{j=k+1}^{\infty} \alpha_j \alpha_{j+|h|} \sigma^2, \quad h = 0, \pm 1, \pm 2, \dots$$

It follows that

$$\begin{aligned} \text{Var}\{D_{kn}\} &= \frac{1}{n} \sum_{h=-\infty}^{n-1} (n - |h|) \gamma_W(h) \\ &\leq \sigma^2 \sum_{j=k+1}^{\infty} \alpha_j^2 + 2\sigma^2 \sum_{h=1}^{n-1} \sum_{j=k+1}^{\infty} |\alpha_j \alpha_{j+h}| \\ &\leq \sigma^2 \left( \sum_{j=k+1}^{\infty} |\alpha_j| \right)^2. \end{aligned}$$

Therefore, by Chebyshev's inequality,  $D_{kn}$  converges to zero as  $k \rightarrow \infty$  uniformly in  $n$ . For fixed  $k$ ,  $Y_{tk}$  is a finite moving average of  $e$ , that satisfies a central limit theorem. Therefore, by Proposition 6.3.1, as  $n$  tends to infinity,

$$S_{kn} \xrightarrow{\mathcal{L}} N\left[0, \left(\sum_{j=0}^k \alpha_j\right)^2 \sigma^2\right].$$

As  $k$  increases, the variance of the normal distribution converges to  $(\sum_{j=0}^{\infty} \alpha_j)^2 \sigma^2$ . The conclusion follows by Lemma 6.3.1.  $\blacktriangle$

The stationary autoregressive moving average is a special case of an infinite moving average. Because the weights decline exponentially for the autoregressive

moving average, the sample mean of  $Y_t$  can be expressed as a function of the sample mean of the  $e_t$ , with remainder of order in probability  $n^{-1}$ .

**Corollary 6.3.3.** Let  $Y_t$  be a stationary finite order autoregressive moving average satisfying

$$\sum_{j=0}^p a_j(Y_{t-j} - \mu) = \sum_{i=0}^q b_i e_{t-i},$$

where  $a_0 = 1$ ,  $b_0 = 1$ , the roots of the autoregressive characteristic equation are less than one in absolute value, and  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables. Then

$$\bar{y}_n = \mu + n^{-1} \sum_{t=1}^n \left( \sum_{j=0}^{\infty} \gamma_j \right) e_t + O_p(n^{-1}),$$

where the  $\gamma_j$  are defined in Theorem 2.7.1 and  $\sum_{j=0}^{\infty} \gamma_j = (\sum_{j=0}^p a_j)^{-1} \sum_{i=0}^q b_i$ . If  $n^{1/2} \bar{e}_n$  converges in distribution to a  $N(0, \sigma^2)$  random variable, then

$$n^{1/2}(\bar{y}_n - \mu) \xrightarrow{\mathcal{D}} N\left[0, \left(\sum_{j=0}^{\infty} \gamma_j\right)^2 \sigma^2\right].$$

**Proof.** By the representation of Theorem 2.7.1,

$$\begin{aligned} \sum_{t=1}^n (Y_t - \mu) &= \sum_{t=1}^n \left( \sum_{j=0}^{\infty} \gamma_j \right) e_t + \sum_{i=1}^{\infty} \sum_{j=i}^{n+i-1} \gamma_j e_{t-i} \\ &\quad - \sum_{i=0}^{n-1} \sum_{j=i}^{\infty} \gamma_{j+1} e_{n-i}. \end{aligned}$$

Now, by Exercise 2.24, there is some  $M$  and  $0 < \lambda < 1$  such that  $|\gamma_j| < M\lambda^j$  for all  $j \geq 0$ . Hence, there is a  $K$  such that

$$V\left\{ \sum_{i=1}^{\infty} \sum_{j=i}^{n+i-1} \gamma_j e_{t-i} \right\} \leq \sum_{i=1}^{\infty} \lambda^{2i} K^2 \sigma^2 = (1 - \lambda^2)^{-1} \lambda^2 K^2 \sigma^2$$

and

$$V\left\{ \sum_{i=0}^{n-1} \sum_{j=i}^{\infty} \gamma_{j+1} e_{n-i} \right\} \leq (1 - \lambda^2)^{-1} \lambda^2 K^2 \sigma^2$$

for all  $n$ . Dividing the variances by  $n^2$ , we obtain the first result. The limiting distribution result is an immediate consequence of the order in probability result and the assumption that  $n^{1/2} \bar{e}_n$  converges in distribution.  $\blacktriangle$

We now obtain a central limit theorem for a linear function of a realization of a

stationary time series. We shall state the assumptions of this theorem in a manner that permits us to use the Lindeberg central limit theorem.

**Theorem 6.3.4.** Let  $\{X_t; t \in T = (0, \pm 1, \pm 2, \dots)\}$  be a time series defined by

$$X_t = \sum_{j=0}^{\infty} \alpha_j e_{t-j},$$

where  $\sum_{j=0}^{\infty} |\alpha_j| < \infty$ , and the  $e_t$  are independent  $(0, \sigma^2)$  random variables with distribution functions  $F_t(e)$  such that

$$\lim_{\delta \rightarrow \infty} \sup_{t \in T} \int_{|e| > \delta} e^2 dF_t(e) = 0.$$

Furthermore, let  $\{C_t\}_{t=1}^{\infty}$  be a sequence of fixed real numbers satisfying

$$(i) \quad \lim_{n \rightarrow \infty} \sum_{t=1}^n C_t^2 = \infty,$$

$$(ii) \quad \lim_{n \rightarrow \infty} \left( \sum_{t=1}^n C_t^2 \right)^{-1} C_n^2 = 0,$$

$$(iii) \quad \lim_{n \rightarrow \infty} \left( \sum_{t=1}^n C_t^2 \right)^{-1} \sum_{i=1}^{n-h} C_i C_{i+h} = g(h), \quad h = 0, \pm 1, \pm 2, \dots.$$

Let  $V = \sum_{h=-\infty}^{\infty} g(h) \gamma_X(h) \neq 0$ . Then

$$\left( \sum_{t=1}^n C_t^2 \right)^{-1/2} \sum_{t=1}^n C_t X_t \xrightarrow{\mathcal{D}} N(0, V).$$

**Proof.** Following the proof of Theorem 6.3.3, we set

$$(Y_{ik}, W_{ik}) = \left( \sum_{j=0}^k \alpha_j e_{i-j}, \sum_{j=k+1}^{\infty} \alpha_j e_{i-j} \right)$$

and note that

$$\begin{aligned} \text{Var} \left\{ \left( \sum_{t=1}^n C_t^2 \right)^{-1/2} \sum_{t=1}^n C_t W_{ik} \right\} &= \left( \sum_{t=1}^n C_t^2 \right)^{-1} \sum_{t=1}^n \sum_{j=1}^n C_t C_j \gamma_w(t-j) \\ &\leq \sum_{h=-(n-1)}^{n-1} |\gamma_w(h)|, \end{aligned}$$

which, by the proof of Theorem 6.3.3, can be made arbitrarily small by choosing  $k$  sufficiently large.

For fixed  $k$ ,  $Y_{ik}$  is a finite moving average, and, following the proof of Theorem

6.3.1, we have

$$\left( \sum_{t=1}^n C_t^2 \right)^{-1/2} \left( \sum_{t=1}^n C_t Y_{it} \right) = \left( \sum_{t=1}^n C_t^2 \right)^{-1/2} \sum_{t=1}^n \sum_{j=0}^k b_j e_t + o_p(1),$$

where  $b_j = \sum_{i=0}^k \alpha_i C_{i+j}$ , and the random variables  $b_j e_t$  are independent with mean zero and variance  $b_j^2 \sigma^2$ . Let  $V_n = \sum_{t=1}^n b_t^2 \sigma^2$ ,

$$S_n = V_n^{-1/2} \sum_{t=1}^n b_t e_t,$$

$$M_n = \sup_{1 \leq t \leq n} b_t^2,$$

and consider, for  $\delta > 0$ ,

$$\lim_{n \rightarrow \infty} V_n^{-1} \sum_{t=1}^n b_t^2 \int_{R_t} e^2 dF_t(e) \leq \lim_{n \rightarrow \infty} \sum_{t=1}^n V_n^{-1} b_t^2 \sup_{1 \leq t \leq n} \int_{R_0} e^2 dF_t(e),$$

where

$$R_t = \{e: |e| > V_n^{1/2} |b_t|^{-1} \delta\}$$

and

$$R_0 = \{e: |e| > M_n^{-1/2} V_n^{1/2} \delta\}.$$

Clearly,  $R_t \subset R_0$  for all  $t \leq n$ . Assumptions i and ii imply that

$$\lim_{n \rightarrow \infty} \sup_{1 \leq t \leq n} \left( \sum_{j=1}^n C_j^2 \right)^{-1} C_t^2 = 0.$$

Because  $V = \lim_{n \rightarrow \infty} (\sum_{t=1}^n C_t^2)^{-1} \sum_{t=1}^n b_t^2 \sigma^2 \neq 0$ , we have

$$\lim_{n \rightarrow \infty} \sup_{1 \leq t \leq n} \left( \sum_{r=1}^n b_r^2 \right)^{-1} b_t = 0,$$

which, in turn, implies that  $\lim_{n \rightarrow \infty} M_n^{-1/2} V_n^{-1/2} = \infty$ .

By assumption, the supremum of the integral over  $R_0$  goes to zero. Hence, the assumptions of the Lindeberg central limit theorem are met, and  $S_n$  converges in distribution to a normal random variable with zero mean and unit variance. Since

$$\lim_{n \rightarrow \infty} \left( \sum_{t=1}^n C_t^2 \right)^{-1} \sum_{t=1}^n b_t^2 = g(0) \sum_{j=0}^k \alpha_j^2 + 2 \sum_{h=1}^k g(h) \sum_{j=0}^{k-h} \alpha_j \alpha_{j+h}$$

and

$$\lim_{k \rightarrow \infty} \sigma^2 \left\{ g(0) \sum_{j=0}^k \alpha_j^2 + 2 \sum_{h=1}^k g(h) \sum_{j=0}^{k-h} \alpha_j \alpha_{j+h} \right\} = \sum_{h=-\infty}^{\infty} g(h) \gamma_X(h),$$

the result follows by Lemma 6.3.1.  $\blacktriangle$

Because a stationary finite order autoregressive moving average time series can be expressed as an infinite moving average with absolutely summable covariance function, the conclusions of Theorem 6.3.4 hold for such time series. Also, the condition that the  $e_t$  are independent  $(0, \sigma^2)$  random variables can be replaced by the condition that the  $e_t$  are martingale differences with  $E[e_t^2 | \mathcal{A}_{t-1}] = \sigma^2$  a.s. and bounded  $2 + \delta$  ( $\delta > 0$ ) moments, where  $\mathcal{A}_{t-1}$  is the sigma-field generated by  $e_1, \dots, e_{t-1}$ .

We now investigate the large sample properties of the estimated autocovariances and autocorrelations. We use Lemma 6.3.2 to demonstrate that the estimated autocovariances converge in probability to the true values. The assumptions and conclusions of Theorem 6.3.5 differ from those of Theorem 6.2.1. In Theorem 6.2.1, the existence of fourth moments enabled us to obtain variances of the estimated autocovariances. In Theorem 6.3.5, weaker assumptions are used to obtain convergence in probability of the sample autocovariances.

**Theorem 6.3.5.** Let the stationary time series  $Y_t$  be defined by

$$Y_t = \sum_{j=-\infty}^{\infty} w_j e_{t-j},$$

where  $\sum_{j=-\infty}^{\infty} |w_j| < \infty$  and  $e_t \sim II(0, \sigma^2)$ . Then

$$\plim_{n \rightarrow \infty} \tilde{\gamma}(h) = \plim_{n \rightarrow \infty} \hat{\gamma}(h) = \gamma(h) = \sigma^2 \sum_{j=-\infty}^{\infty} w_j w_{j+h}.$$

**Proof.** Now

$$\begin{aligned} \tilde{\gamma}(h) &= (n-h)^{-1} \sum_{i=1}^{n-h} \sum_{j=-\infty}^{\infty} w_j e_{t-j} \sum_{i=-\infty}^{\infty} w_i e_{t-i+h} \\ &= (n-h)^{-1} \sum_{i=1}^{n-h} \left[ \sum_{j=-\infty}^{\infty} w_j w_{j+h} e_{t-j}^2 + \sum_{i \neq j} w_j w_{i+h} e_{t-j} e_{t-i} \right]. \end{aligned}$$

Also,

$$\plim_{n \rightarrow \infty} (n-h)^{-1} \sum_{i=1}^{n-h} \sum_{j=-\infty}^{\infty} w_j w_{j+h} e_{t-j}^2 = \gamma(h)$$

by Theorem 6.3.2.

To show that the term

$$(n-h)^{-1} \sum_{t=1}^{n-h} \sum_{i \neq j} w_j w_{i+h} e_{i-j} e_{t-i}$$

converges to zero, we apply Lemma 6.3.2, letting

$$S_{kn} = (n-h)^{-1} \sum_{t=1}^{n-h} \sum_{\substack{|i| \leq k \\ |j| \leq k \\ i \neq j}} w_j w_{i+h} e_{i-j} e_{t-i},$$

$$D_{kn}^{(1)} = (n-h)^{-1} \sum_{t=1}^{n-h} \sum_{\substack{|i| > k \\ |j| > k \\ i \neq j}} w_j w_{i+h} e_{i-j} e_{t-i},$$

$$D_{kn}^{(2)} = (n-h)^{-1} \sum_{t=1}^{n-h} \sum_{\substack{|i| > k \\ |j| \leq k \\ i \neq j}} w_j w_{i+h} e_{i-j} e_{t-i},$$

$$D_{kn}^{(3)} = (n-h)^{-1} \sum_{t=1}^{n-h} \sum_{\substack{|i| \leq k \\ |j| > k \\ i \neq j}} w_j w_{i+h} e_{i-j} e_{t-i},$$

and

$$D_{kn} = D_{kn}^{(1)} + D_{kn}^{(2)} + D_{kn}^{(3)}.$$

For fixed  $i, j$  with  $i \neq j$ , using Chebyshev's inequality, it can be proved that

$$\operatorname{plim}_{n \rightarrow \infty} (n-h)^{-1} \sum_{t=1}^{n-h} w_j w_{i+h} e_{i-j} e_{t-i} = 0.$$

Hence, for any fixed  $k$ ,  $\operatorname{plim}_{n \rightarrow \infty} S_{kn} = 0$ .

Now, by Chebyshev's inequality,

$$P\{|D_{kn}^{(1)}| > \epsilon\} \leq \epsilon^{-1} \left( \sum_{|i| > k} \sum_{|j| > k \atop i \neq j} |w_j| |w_{i+h}| \sigma^2 \right).$$

Therefore,

$$\operatorname{plim}_{k \rightarrow \infty} D_{kn}^{(1)} = 0 \quad \text{uniformly in } n.$$

In a similar manner, we can prove

$$\operatorname{plim}_{k \rightarrow \infty} [D_{kn}^{(2)} + D_{kn}^{(3)}] = 0 \quad \text{uniformly in } n,$$

and it follows that  $\hat{\gamma}(h)$  converges in probability to  $\gamma(h)$ . The result for  $\hat{\gamma}(h)$  follows because, by the proof of Theorem 6.2.2,  $\hat{\gamma}(h) - \gamma(h) = O_p(n^{-1})$ .  $\blacktriangle$

**Corollary 6.3.5.** Let the stationary time series  $Y_t$  satisfy

$$Y_t = \sum_{j=-\infty}^{\infty} w_j e_{t-j},$$

where  $\sum_{j=-\infty}^{\infty} |w_j| < \infty$ ,

$$E\{(e_t, e_t^2) | \mathcal{A}_{t-1}\} = (0, \sigma^2) \quad \text{a.s.}$$

$$E\{|e_t|^{2+\delta} | \mathcal{A}_{t-1}\} < L < \infty \quad \text{a.s.}$$

for some  $\delta > 0$ , and  $\mathcal{A}_{t-1}$  is the sigma-field generated by  $e_{t-1}, e_{t-2}, \dots$ . Then

$$\lim_{n \rightarrow \infty} \hat{\gamma}(h) = \lim_{n \rightarrow \infty} \hat{\gamma}(h) = \gamma(h) = \sigma^2 \sum_{j=-\infty}^{\infty} w_j w_{j+h}.$$

**Proof.** The proof parallels that of Theorem 6.3.5. ▲

The vector of a finite number of sample autocovariances converges in distribution to a normal vector under mild assumptions.

**Theorem 6.3.6.** Let  $Y_t$  be the stationary time series defined by

$$Y_t = \sum_{j=-\infty}^{\infty} b_j e_{t-j},$$

where the  $e_t$  are independent  $(0, \sigma^2)$  random variables with fourth moment  $\eta\sigma^4$  and finite  $4 + \delta$  absolute moment for some  $\delta > 0$ , and the  $b_j$  are absolutely summable. Let  $K$  be fixed. Then the limiting distribution of  $n^{1/2}[\hat{\gamma}(0) - \gamma(0), \hat{\gamma}(1) - \gamma(1), \dots, \hat{\gamma}(K) - \gamma(K)]'$  is multivariate normal with mean zero and covariance matrix  $\mathbf{V}$  whose elements are defined by (6.2.4) of Theorem 6.2.1.

**Proof.** The estimated covariance, for  $h = 0, 1, 2, \dots, K$ , is

$$\begin{aligned} \hat{\gamma}(h) &= n^{-1} \sum_{t=1}^{n-h} (Y_t - \bar{y}_n)(Y_{t+h} - \bar{y}_n) \\ &= n^{-1} \sum_{t=1}^{n-h} Y_t Y_{t+h} - n^{-1} \bar{y}_n \sum_{t=1}^{n-h} (Y_t + Y_{t+h}) + n^{-1}(n-h)\bar{y}_n^2 \end{aligned} \quad (6.3.9)$$

and the last two terms, when multiplied by  $n^{1/2}$ , converge in probability to zero. Therefore, in investigating the limiting distribution of  $n^{1/2}[\hat{\gamma}(h) - \gamma(h)]$  we need only consider the first term on the right of (6.3.9). Let

$$Y_t = X_{mt} + W_{mt},$$

where

$$\begin{aligned}(X_{mt}, W_{mt}) &= \left( \sum_{j=-m}^m b_j e_{t-j}, \sum_{|j|>m} b_j e_{t-j} \right) \\ &= \left( \sum_{j=-\infty}^{\infty} \alpha_j e_{t-j}, \sum_{j=-\infty}^{\infty} \beta_j e_{t-j} \right),\end{aligned}$$

$\alpha_j = b_j$  if  $|j| \leq m$  and  $\alpha_j = 0$  otherwise, and  $\beta_j = b_j$  if  $|j| \geq m+1$  and  $\beta_j = 0$  otherwise. Then

$$\begin{aligned}n^{-1/2} \sum_{t=1}^{n-h} (X_{mt} + W_{mt})(X_{m,t+h} + W_{m,t+h}) \\ = n^{-1/2} \left[ \sum_{t=1}^{n-h} X_{mt} X_{m,t+h} + \sum_{t=1}^{n-h} X_{mt} W_{m,t+h} \right. \\ \left. + \sum_{t=1}^{n-h} W_{mt} X_{m,t+h} + \sum_{t=1}^{n-h} W_{mt} W_{m,t+h} \right] \\ \stackrel{(say)}{=} S_{mhn} + D_{mn},\end{aligned}$$

where  $S_{mhn} = n^{-1/2} \sum_{t=1}^{n-h} X_{mt} X_{m,t+h}$ . Following the proof of Theorem 6.2.1, we have

$$\begin{aligned}E \left\{ (n-h)^{-1} (n-q)^{-1} \sum_{s=1}^{n-q} \sum_{t=1}^{n-h} X_{mt} X_{m,t+h} W_{ms} W_{m,s+q} \right\} - \gamma_{X_m}(h) \gamma_{W_m}(q) \\ = (\eta-3)\sigma^4 (n-h)^{-1} (n-q)^{-1} \sum_{s=1}^{n-q} \sum_{t=1}^{n-h} \sum_{j=-\infty}^{\infty} \alpha_j \alpha_{j+h} \beta_{j+s-t} \beta_{j+s-t+q} \\ + (n-h)^{-1} (n-q)^{-1} \sum_{s=1}^{n-q} \sum_{t=1}^{n-h} \left[ \sum_{j=-\infty}^{\infty} \alpha_j \beta_{j+s-t} \right. \\ \times \left. \sum_{i=-\infty}^{\infty} \alpha_{i+h} \beta_{i+q+s-t} \sum_{j=-\infty}^{\infty} \alpha_j \beta_{j+q+s-t} \sum_{i=-\infty}^{\infty} \alpha_{i+h} \beta_{i+s-t} \right] \sigma^4.\end{aligned}$$

Now,

$$\sum_{p=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} |\alpha_j| |\beta_{j+p}| \leq \sum_{p=-\infty}^{\infty} |\beta_p| \sum_{j=-\infty}^{\infty} |\alpha_j| < L_m,$$

where  $L_m \rightarrow 0$  uniformly in  $n$  as  $m \rightarrow \infty$  because  $\sum_{|j|>m} b_j \rightarrow 0$  as  $m \rightarrow \infty$ . Therefore,  $D_{mn} \rightarrow 0$  uniformly in  $n$  as  $m \rightarrow \infty$ .

Consider a linear combination of the errors in the estimated covariances of  $X_{mt}$ ,

$$\begin{aligned}S_{m,n} &= n^{1/2} \sum_{h=0}^K \lambda_h \left[ n^{-1} \sum_{t=1}^{n-h} X_{mt} X_{m,t+h} - \gamma_{X_m}(h) \right] \\ &= n^{-1/2} \sum_{h=0}^K \sum_{t=1}^{n-h} \lambda_h [Z_{th} - E\{Z_{th}\}] - n^{-1/2} \sum_{h=0}^K h \lambda_h \gamma_{X_m}(h),\end{aligned}$$

where the  $\lambda_h$  are arbitrary real numbers (not all zero) and

$$Z_{th} = X_{mr} X_{m,t+h}, \quad h = 0, 1, 2, \dots, K.$$

Now,  $Z_{th}$  is an  $(m+h)$ -dependent covariance stationary time series with mean  $\gamma_{X_m}(h)$  and covariance function

$$\begin{aligned} \gamma_{Z_h}(s) &= (\eta - 3)\sigma^4 \sum_{j=-\infty}^{\infty} \alpha_j \alpha_{j+h} \alpha_{j+s} \alpha_{j+s+h} \\ &\quad + \gamma_{X_m}^2(h) + \gamma_{X_m}^2(s) + \gamma_{X_m}(s+h)\gamma_{X_m}(s-h), \end{aligned} \quad (6.3.10)$$

where we have used (6.2.5). Thus, the weighted average of the  $Z_{th}$ 's,

$$U_t = \sum_{h=0}^K \lambda_h Z_{th} = \sum_{h=0}^K \lambda_h X_{mr} X_{m,t+h},$$

is a stationary time series. Furthermore, the time series  $U_t$  is  $(m+K)$ -dependent, it has finite  $2 + \delta/2$  moment, and

$$\lim_{n \rightarrow \infty} n^{-1/2} \sum_{h=0}^K h \lambda_h \gamma_{X_m}(h) = 0.$$

Therefore, by Theorem 6.3.1,  $S_{mn}$  converges in distribution to a normal random variable. Since the  $\lambda_h$  are arbitrary, the vector random variable

$$n^{1/2}[\hat{\gamma}_{X_m}(0) - \gamma_{X_m}(0), \hat{\gamma}_{X_m}(1) - \gamma_{X_m}(1), \dots, \hat{\gamma}_{X_m}(K) - \gamma_{X_m}(K)]$$

converges in distribution to a multivariate normal by Theorem 5.3.3, where the covariance matrix is defined by (6.3.10). Because  $E\{X_{mr} X_{m,t+h}\}$  converges to  $E\{Y_t Y_{t+h}\}$  as  $m \rightarrow \infty$ , the conditions of Lemma 6.3.1 are satisfied and we obtain the conclusion. ▲

Theorem 6.3.6 can be proven for  $e$ , that are martingale differences. See Exercises 6.23 and 6.24, and Hannan and Heyde (1972).

Generally it is the estimated autocorrelations that are subjected to analysis, and hence their limiting distribution is of interest.

**Corollary 6.3.6.1.** Let the assumptions of Theorem 6.3.6 hold. Then the vector  $n^{1/2}[\hat{\rho}(1) - \rho(1), \hat{\rho}(2) - \rho(2), \dots, \hat{\rho}(K) - \rho(K)]'$  converges in distribution to a multivariate normal with mean zero and covariance matrix  $G$ , where the  $hq$ th element of  $G$  is  $\sum_{p=-\infty}^{\infty} [\rho(p)\rho(p-h+q) + \rho(p+q)\rho(p-h) - 2\rho(q)\rho(p)\rho(p-h) - 2\rho(h)\rho(p)\rho(p-q) + 2\rho(h)\rho(q)\rho^2(p)].$

**Proof.** Since the  $\hat{\rho}(h)$  are continuous differentiable functions of the  $\hat{\gamma}(h)$ , the result follows from Theorems 5.1.4, 6.2.3, and 6.3.5. ▲

Observe that if the original time series  $X_t$  is a sequence of independent identically distributed random variables with finite moments, the sample correlations will be nearly independent in large samples. Because of the importance of this result in the testing of time series for independence, we state it as a corollary.

**Corollary 6.3.6.2.** Let the time series  $e_t$  be a sequence of independent  $(0, \sigma^2)$  random variables with uniformly bounded  $4 + \delta$  moments for some  $\delta > 0$ . Let  $\hat{\rho}(h)$  be defined by (6.2.12), and let  $K$  be a fixed integer. Then  $n(n-h)^{-1/2} \hat{\rho}(h)$ ,  $h = 1, 2, \dots, K$ , converge in distribution to independent normal  $(0, 1)$  random variables.

**Proof.** Omitted. ▲

**Example 6.3.1.** The quarterly seasonally adjusted United States unemployment rate from 1948 to 1972 is given in Table 6.3.1 and displayed in Figure 6.3.1. The mean unemployment rate is 4.77. The autocorrelation function estimated using (6.2.7) is given in Figure 6.3.2 and Table 6.3.2. This plot is sometimes called a correlogram.

To carry out statistical analyses we assume the time series can be treated as a stationary time series with finite sixth moment. If the original time series was a sequence of uncorrelated random variables, we would expect about 95% of the estimated correlations to fall between the lines plus and minus  $1.96n^{-1}(n-h)^{1/2}$ . Obviously unemployment is not an uncorrelated time series. Casual inspection of the correlogram might lead one to conclude that the time series contains a periodic component with a period of about 54 quarters, since the estimated correlations for  $h$  equal to 21 through 33 are negative and below the 1.96 sigma bounds for an uncorrelated time series. However, because the time series is highly correlated at small lags, the variance of the estimated correlations at large lags is much larger than the variance of correlations computed from a white noise sequence.

The first few autocorrelations of the unemployment time series are in good agreement with those generated by the second order autoregressive process

$$X_t = 1.5356X_{t-1} - 0.6692X_{t-2} + e_t,$$

where the  $e_t$  are uncorrelated  $(0, 0.1155)$  random variables. We shall discuss the estimation of the parameters of autoregressive time series in Chapter 8. However, the fact that the sample autocorrelations are consistent estimators of the population correlations permits us to obtain consistent estimators of the autoregressive parameters of a second order process from (2.5.7). The general agreement between the correlations for the second order autoregressive process and the sample correlations is clear from Table 6.3.2.

The roots of the second order autoregressive process are  $0.768 \pm 0.282i$ . We recall that the correlation function of such a second order process can be written as

$$\rho(h) = b_1 m_1^h + b_2 m_2^h, \quad h = 0, 1, 2, \dots,$$

Table 6.3.1. U.S. Unemployment Rate (Quarterly Seasonally Adjusted) 1948-1972

Year	Quarter	Rate	Year	Quarter	Rate	Year	Quarter	Rate
1948	1	3.73	1957	1	3.93	1966	1	3.87
	2	3.67		2	4.10		2	3.80
	3	3.77		3	4.23		3	3.77
	4	3.83		4	4.93		4	3.70
1949	1	4.67	1958	1	6.30	1967	1	3.77
	2	5.87		2	7.37		2	3.83
	3	6.70		3	7.33		3	3.83
	4	6.97		4	6.37		4	3.93
1950	1	6.40	1959	1	5.83	1968	1	3.73
	2	5.57		2	5.10		2	3.57
	3	4.63		3	5.27		3	3.53
	4	4.23		4	5.60		4	3.43
1951	1	3.50	1960	1	5.13	1969	1	3.37
	2	3.10		2	5.23		2	3.43
	3	3.17		3	5.53		3	3.60
	4	3.37		4	6.27		4	3.60
1952	1	3.07	1961	1	6.80	1970	1	4.17
	2	2.97		2	7.00		2	4.80
	3	3.23		3	6.77		3	5.17
	4	2.83		4	6.20		4	5.87
1953	1	2.70	1962	1	5.63	1971	1	5.93
	2	2.57		2	5.53		2	5.97
	3	2.73		3	5.57		3	5.97
	4	3.70		4	5.53		4	5.97
1954	1	5.27	1963	1	5.77	1972	1	5.83
	2	5.80		2	5.73		2	5.77
	3	5.97		3	5.50		3	5.53
	4	5.33		4	5.57		4	5.30
1955	1	4.73	1964	1	5.47			
	2	4.40		2	5.20			
	3	4.10		3	5.00			
	4	4.23		4	5.00			
1956	1	4.03	1965	1	4.90			
	2	4.20		2	4.67			
	3	4.13		3	4.37			
	4	4.13		4	4.10			

Sources. *Business Statistics*, 1971 Biennial Edition, pp. 68 and 233 and *Survey of Current Business*, January 1972 and January 1973. Quarterly data are the averages of monthly data.

where

$$(b_1, b_2) = (m_2 - m_1)^{-1} [m_2 - \rho(1), \rho(1) - m_1].$$

For the unemployment time series the estimated parameters are  $\hat{b}_1 = 0.500 -$

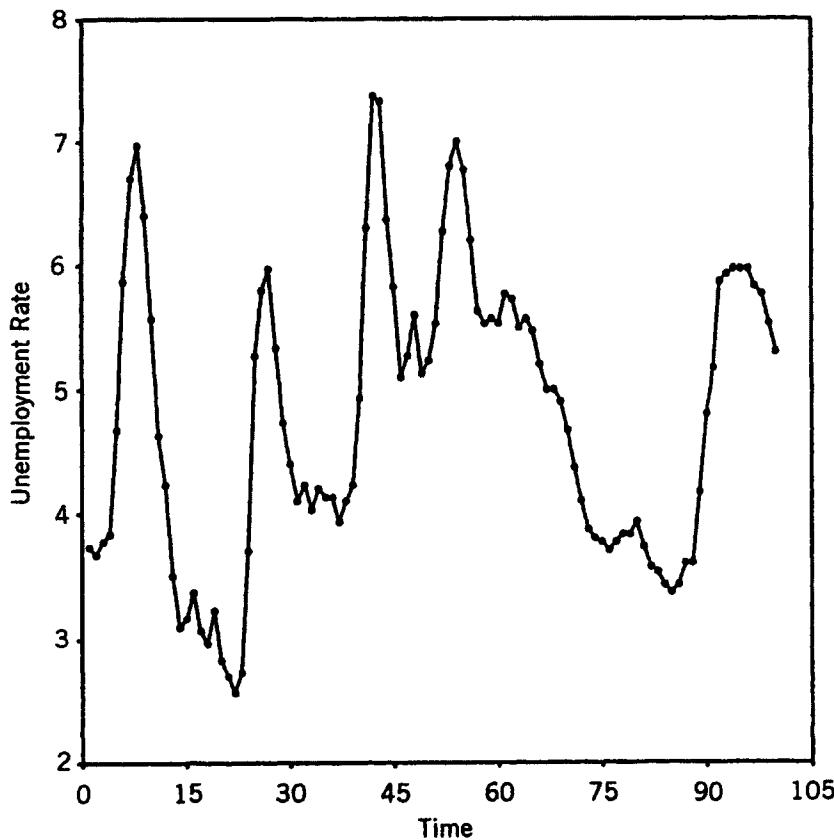


Figure 6.3.1. United States quarterly seasonally adjusted unemployment rate.

$0.270\epsilon$  and  $\hat{b}_2 = \hat{b}_1^* = 0.500 + 0.270\epsilon$ . Using these values, we can estimate the variance of the estimated autocorrelations for large lags using (6.2.11). We have

$$\begin{aligned} n \hat{\text{Var}}\{\hat{r}(h)\} &\doteq \sum_{s=-\infty}^{\infty} \hat{\rho}^2(s) \\ &= \hat{b}_1^2 \frac{1 + \hat{m}_1^2}{1 - \hat{m}_1^2} + \hat{b}_2^2 \frac{1 + \hat{m}_2^2}{1 - \hat{m}_2^2} + 2\hat{b}_1\hat{b}_2 \frac{1 + \hat{m}_1\hat{m}_2}{1 - \hat{m}_1\hat{m}_2} \\ &= 4.812. \end{aligned}$$

Thus the estimated standard error of the estimated autocorrelations at large lags is about 0.22, and the observed correlations at lags near 27 could arise from such a process.

Given that the time series was generated by the second order autoregressive mechanism, the variance of the sample mean can be estimated by Corollary

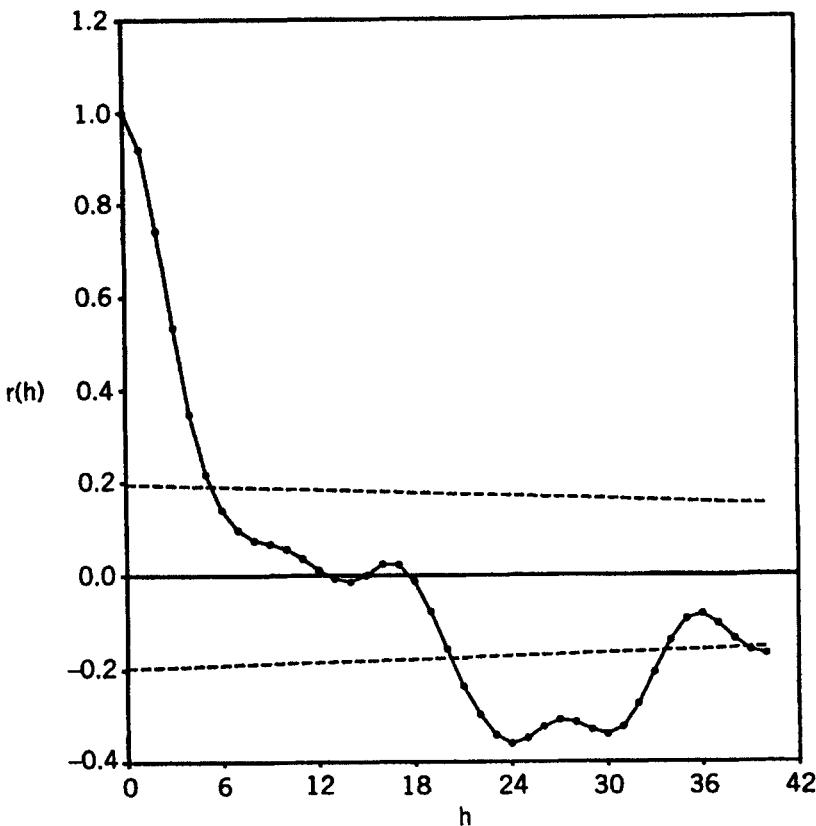


Figure 6.3.2. Correlogram of quarterly seasonally adjusted unemployment rate.

#### 6.1.1.2. By that corollary

$$\begin{aligned}
 n \hat{\text{Var}}\{\bar{x}_n\} &= \sum_{h=-\infty}^{\infty} \hat{\gamma}(h) = \hat{\gamma}(0) \sum_{h=-\infty}^{\infty} \hat{\rho}(h) \\
 &= \frac{0.1155}{(1 - 1.5356 + 0.6692)^2} \\
 &= 6.47 .
 \end{aligned}$$

For this highly correlated process the variance of the mean is about five times that of an uncorrelated time series with the same variance. ▲▲

## 6.4. ESTIMATION OF THE CROSS COVARIANCES

In our Section 1.7 discussion of vector valued time series we introduced the  $k \times k$

**Table 6.3.2. Estimated Autocorrelations, Quarterly U.S. Seasonally Adjusted Unemployment Rate, 1948-72**

Lag <i>h</i>	Estimated Correlations	$ 1.96n^{-1}(n-h)^{1/2} $	Correlations for Second Order Process
0	1.0000	—	1.0000
1	0.9200	0.1950	0.9200
2	0.7436	0.1940	0.7436
3	0.5348	0.1930	0.5262
4	0.3476	0.1920	0.3105
5	0.2165	0.1910	0.1247
6	0.1394	0.1900	-0.0164
7	0.0963	0.1890	-0.1085
8	0.0740	0.1880	-0.1557
9	0.0664	0.1870	-0.1665
10	0.0556	0.1859	-0.1515
11	0.0352	0.1849	-0.1212
12	0.0109	0.1839	-0.0848
13	-0.0064	0.1828	-0.0490
14	-0.0135	0.1818	-0.0186
15	0.0004	0.1807	0.0043
16	0.0229	0.1796	0.0190
17	0.0223	0.1786	0.0263
18	-0.0126	0.1775	0.0277
19	-0.0762	0.1764	0.0249
20	-0.1557	0.1753	0.0197
21	-0.2351	0.1742	0.0136
22	-0.2975	0.1731	0.0077
23	-0.3412	0.1720	0.0027
24	-0.3599	0.1709	-0.0010
25	-0.3483	0.1697	-0.0033
26	-0.3236	0.1686	-0.0044
27	-0.3090	0.1675	-0.0046
28	-0.3142	0.1663	-0.0041
29	-0.3299	0.1652	-0.0032
30	-0.3396	0.1640	-0.0022
31	-0.3235	0.1628	-0.0012
32	-0.2744	0.1616	-0.0004
33	-0.2058	0.1604	0.0002
34	-0.1378	0.1592	0.0006
35	-0.0922	0.1580	0.0007
36	-0.0816	0.1568	0.0008
37	-0.1027	0.1556	0.0007
38	-0.1340	0.1543	0.0005
39	-0.1590	0.1531	0.0004
40	-0.1669	0.1518	0.0002

covariance matrix

$$\Gamma(h) = E\{(\mathbf{X}_t - \boldsymbol{\mu})(\mathbf{X}_{t+h} - \boldsymbol{\mu})'\},$$

where  $\mathbf{X}_t$  is a stationary  $k$ -dimensional time series and  $\boldsymbol{\mu} = E\{\mathbf{X}_t\}$ . The  $j$ th diagonal element of the matrix is the autocovariance of  $X_{jt}$ ,  $\gamma_{jj}(h) = E\{(X_{jt} - \mu_j)(X_{j,t+h} - \mu_j)\}$ , and the  $ij$ th element is the cross covariance between  $X_{it}$  and  $X_{jt}$ ,

$$\gamma_{ij}(h) = E\{(X_{it} - \mu_i)(X_{j,t+h} - \mu_j)\}.$$

Expressions for the estimated cross covariance analogous to those of (6.2.1) and (6.2.3) are

$$\hat{\gamma}_{ij}(h) = \begin{cases} \frac{1}{n-h} \sum_{t=1}^{n-h} X_{it} X_{j,t+h}, & h = 0, 1, \dots, n-1, \\ \frac{1}{n+h} \sum_{t=-h}^n X_{it} X_{j,t+h}, & h = -1, -2, \dots, -(n-1), \end{cases} \quad (6.4.1)$$

for the means known and taken to be zero, and

$$\hat{\gamma}_{ij}(h) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-h} (X_{it} - \bar{x}_{in})(X_{j,t+h} - \bar{x}_{jn}), & h = 0, 1, \dots, n-1, \\ \frac{1}{n} \sum_{t=-h}^n (X_{it} - \bar{x}_{in})(X_{j,t+h} - \bar{x}_{jn}), & h = -1, -2, \dots, -(n-1), \end{cases} \quad (6.4.2)$$

where the unknown means are estimated by  $\bar{x}_{in} = n^{-1} \sum_{t=1}^n X_{it}$ . By (1.7.4), we can also write

$$\hat{\gamma}_{ij}(-h) = \hat{\gamma}_{ji}(h) = \frac{1}{n} \sum_{t=1}^{n-h} (X_{jt} - \bar{x}_{jn})(X_{i,t+h} - \bar{x}_{in}), \quad h = 0, 1, \dots, n-1.$$

The corresponding estimators of the cross correlations are

$$\tilde{r}_{ij}(h) = [\hat{\gamma}_{ii}(0) \hat{\gamma}_{jj}(0)]^{-1/2} \hat{\gamma}_{ij}(h) \quad (6.4.3)$$

and

$$\hat{r}_{ij}(h) = [\hat{\gamma}_{ii}(0) \hat{\gamma}_{jj}(0)]^{-1/2} \hat{\gamma}_{ij}(h) \quad (6.4.4)$$

By our earlier results (see Theorem 6.2.2) the estimation of the mean in estimator (6.4.2) introduces a bias that is  $O(n^{-1})$  for time series with absolutely summable covariance function. The properties of the estimated cross covariances are analogous to the properties of the estimated autocovariance given in Theorems 6.2.1 and 6.2.3. To simplify the derivation, we only present the results for normal time series.

**Theorem 6.4.1.** Let the bivariate stationary normal time series  $\mathbf{X}_t$  be such that

$$\sum_{h=-\infty}^{\infty} |\gamma_{ii}(h)| < \infty, \quad i = 1, 2.$$

Then

$$\begin{aligned} \lim_{n \rightarrow \infty} n \operatorname{Cov}\{\hat{\gamma}_{12}(h), \hat{\gamma}_{12}(q)\} &= \sum_{p=-\infty}^{\infty} \gamma_{11}(p) \gamma_{22}(p+q-h) \\ &\quad + \sum_{p=-\infty}^{\infty} \gamma_{12}(p+q) \gamma_{21}(p-h). \end{aligned}$$

**Proof.** Letting the mean vector be zero and  $h, q \geq 0$ , we have

$$E\{\hat{\gamma}_{12}(h) \hat{\gamma}_{12}(q)\} = E\left\{ \frac{1}{n^2} \sum_{i=1}^{n-h} X_{1i} X_{2,i+h} \sum_{s=1}^{n-q} X_{1s} X_{2,s+q} \right\} + O(n^{-1}),$$

where the remainder term enters because the mean is estimated. Evaluating the expectation, we have

$$\begin{aligned} E\{[\hat{\gamma}_{12}(h) - \gamma_{12}(h)][\hat{\gamma}_{12}(q) - \gamma_{12}(q)]\} &= \frac{1}{n^2} \sum_{i=1}^{n-h} \sum_{s=1}^{n-q} \gamma_{11}(s-i) \gamma_{22}(s-i+q-h) \\ &\quad + \frac{1}{n^2} \sum_{i=1}^{n-h} \sum_{s=1}^{n-q} \gamma_{12}(s+q-i) \gamma_{21}(s-i-h) + O(n^{-1}). \end{aligned}$$

Using Lemma 6.2.1 to take the limit, we have the stated result.  $\blacktriangle$

Since the cross correlations are simple functions of the cross covariances, we can obtain a similar expression for the covariance of the estimated cross correlations.

**Corollary 6.4.1.1.** Given the bivariate stationary normal time series of Theorem 6.4.1,

$$\begin{aligned} \lim_{n \rightarrow \infty} n \operatorname{Cov}\{\hat{r}_{12}(h), \hat{r}_{12}(q)\} &= \sum_{p=-\infty}^{\infty} \{ \rho_{11}(p) \rho_{22}(p+q-h) + \rho_{12}(p+q) \rho_{21}(p-h) \\ &\quad - \rho_{12}(h)[\rho_{11}(p) \rho_{21}(p+q) + \rho_{22}(p) \rho_{21}(p-q)] \\ &\quad - \rho_{12}(q)[\rho_{11}(p) \rho_{21}(p+h) + \rho_{22}(p) \rho_{21}(p-q)] \\ &\quad + \rho_{12}(h) \rho_{12}(q) [\frac{1}{2} \rho_{11}^2(p) + \rho_{12}^2(p) + \frac{1}{2} \rho_{22}^2(p)] \}. \end{aligned}$$

**Proof.** By Theorem 5.5.1 we may use the first term in Taylor's series to obtain the leading term in the covariance expression. Evaluating  $\text{Cov}\{\hat{\gamma}_{12}(h), \hat{\gamma}_{12}(q)\}$ ,  $\text{Cov}\{\hat{\gamma}_{12}(h), \frac{1}{2}[\hat{\gamma}_{11}(0) + \hat{\gamma}_{22}(0)]\}$ ,  $\text{Cov}\{\hat{\gamma}_{12}(q), \frac{1}{2}[\hat{\gamma}_{11}(0) + \hat{\gamma}_{22}(0)]\}$ , and  $\text{Var}\{\frac{1}{2}[\hat{\gamma}_{11}(0) + \hat{\gamma}_{22}(0)]\}$  by the methods of Theorem 6.2.1, we obtain the conclusion.  $\blacktriangle$

Perhaps the most important aspect of these rather cumbersome results is that the covariances are decreasing at the rate  $n^{-1}$ . Also, certain special cases are of interest. One working hypothesis is that the two time series are uncorrelated. If  $X_{1t}$  is a sequence of independent normal random variables, we obtain a particularly simple result.

**Corollary 6.4.1.2.** Let  $X_{1t}$  be a bivariate stationary normal time series satisfying

$$\sum_{h=-\infty}^{\infty} |\gamma_{22}(h)| < \infty,$$

$$\gamma_{11}(h) = \begin{cases} \sigma_1^2, & h = 0, \\ 0, & \text{otherwise,} \end{cases}$$

and

$$\gamma_{12}(h) = 0, \quad \text{all } h.$$

Then

$$\lim_{n \rightarrow \infty} n \text{Cov}\{\hat{r}_{12}(h), \hat{r}_{12}(q)\} = \rho_{22}(q - h).$$

In the null case, the variance of the estimated cross correlation is approximately  $n^{-1}$ , and the correlation between estimated cross correlations is the autocorrelation of  $X_{2t}$  multiplied by  $n^{-1}$ . If the two time series are independent and neither time series is autocorrelated, then the estimated cross correlations are uncorrelated with an approximate variance of  $n^{-1}$ .

It is possible to demonstrate that the sample covariances are consistent estimators under much weaker conditions.

**Lemma 6.4.1.** Let

$$(X_{1t}, X_{2t}) = \left( \sum_{j=0}^{\infty} \alpha_j e_{1,t-j}, \sum_{j=0}^{\infty} \beta_j e_{2,t-j} \right),$$

where  $\{\alpha_j\}$  and  $\{\beta_j\}$  are absolutely summable and  $\{\mathbf{e}_t\} = \{(e_{1t}, e_{2t})'\}$  is a sequence of independent  $(0, \Sigma)$  random variables. Assume  $E\{|e_{it}|^{2+\delta}\} < L$  for  $i = 1, 2$  and some

$\delta > 0$ , or that the  $e_i$  are identically distributed. Then

$$n^{-1} \sum_{i=1}^{n-h} X_{1t}, X_{2,t+h} \xrightarrow{P} \gamma_{X_1 X_2}(h), \quad h = 0, 1, \dots.$$

**Proof.** Define

$$(Y_{1t}, Y_{2t}) = \left( \sum_{j=0}^k \alpha_j e_{1,t-j}, \sum_{j=0}^k \beta_j e_{2,t-j} \right),$$

$$(D_{1t}, D_{2t}) = \left( \sum_{j=k+1}^{\infty} \alpha_j e_{1,t-j}, \sum_{j=k+1}^{\infty} \beta_j e_{2,t-j} \right),$$

fix  $h$ , and consider

$$\frac{1}{n} \sum_{i=1}^{n-h} Y_{1t} Y_{2,t+h} = \frac{1}{n} \sum_{i=1}^{n-h} \sum_{j=0}^k \sum_{l=0}^k \alpha_j \beta_l e_{1,t-j} e_{2,t-l+h}.$$

If  $j \neq i - h$ ,

$$\text{Var} \left\{ \frac{1}{n} \sum_{i=1}^{n-h} \alpha_j \beta_l e_{1,t-j} e_{2,t-l+h} \right\} = \frac{n-h}{n^2} \alpha_j^2 \beta_l^2 \sigma_1^2 \sigma_2^2,$$

where  $\sigma_i^2$  is the variance of  $e_{it}$ . If  $j = i - h$  and  $\sigma_{12} = E\{e_{1t} e_{2t}\}$ , then

$$\frac{1}{n} \sum_{i=1}^{n-h} \alpha_j \beta_{j+h} e_{1,t-j} e_{2,t-j} \xrightarrow{P} \alpha_j \beta_{j+h} \sigma_{12}$$

by the weak law of large numbers. [See, for example, Chung (1968, p. 104).] Now

$$\begin{aligned} \left| \frac{1}{n} \sum_{i=1}^{n-h} X_{1t} X_{2,t+h} - \frac{1}{n} \sum_{i=1}^{n-h} Y_{1t} Y_{2,t+h} \right| \\ \leq \left| \frac{1}{n} \sum_{i=1}^{n-h} Y_{1t} D_{2t} \right| + \left| \frac{1}{n} \sum_{i=1}^{n-h} Y_{2t} D_{1t} \right| + \left| \frac{1}{n} \sum_{i=1}^{n-h} D_{1t} D_{2t} \right|. \end{aligned}$$

By Chebyshev's inequality

$$P \left\{ \left| \frac{1}{n} \sum_{i=1}^{n-h} D_{1t}^2 \right| > \epsilon \right\} \leq \sum_{j=k+1}^{\infty} \alpha_j^2 \sigma_1^2 \epsilon^{-1},$$

and it follows that

$$p\lim_{k \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n-h} Y_{1t} D_{2t} = 0,$$

$$p\lim_{k \rightarrow \infty} \frac{1}{n} \sum_{i=1}^{n-h} Y_{2t} D_{1t} = 0,$$

uniformly in  $n$ . Convergence in probability implies convergence in distribution, and by an application of Lemma 6.3.1 we have that  $n^{-1} \sum_{t=1}^{n-h} X_{1,t} X_{2,t+h}$  converges in distribution to the constant  $\gamma_{X_1 X_2}(h)$ . The result follows by Lemma 5.2.1.  $\blacktriangle$

**Theorem 6.4.2.** Let  $\{e_{1,t}\}$  and  $\{X_t\}$  be independent time series, where  $\{e_{1,t}\}$  is a sequence of independent  $(0, \sigma_1^2)$  random variables with uniformly bounded third moment and  $\{X_t\}$  is defined by

$$X_t = \sum_{j=0}^{\infty} \alpha_j e_{2,t-j},$$

where  $\sum_{j=0}^{\infty} |\alpha_j| < \infty$  and  $\{e_{2,t}\}$  is a sequence of independent  $(0, \sigma_2^2)$  random variables with uniformly bounded third moment. Then, for fixed  $h > 0$ ,

$$n^{1/2} \hat{r}_{12}(h) \xrightarrow{\mathcal{D}} N(0, 1).$$

**Proof.** We write

$$\begin{aligned} n^{1/2} \hat{r}_{12}(h) &= \left[ \left( n^{-1} \sum_{t=1}^n e_{1,t}^2 \right)^{1/2} \left( n^{-1} \sum_{t=1}^n X_t^2 \right)^{1/2} \right]^{-1} n^{-1/2} \sum_{t=1}^{n-h} e_{1,t} X_{t+h} \\ &\quad + O_p(n^{-1/2}) \end{aligned}$$

and note that, by Lemma 6.4.1,

$$\begin{aligned} p\lim n^{-1} \sum_{t=1}^n e_{1,t}^2 &= \sigma_1^2, \\ p\lim n^{-1} \sum_{t=1}^n X_t^2 &= \gamma_{XX}(0). \end{aligned}$$

For fixed  $h$ , the time series

$$Z_{th} = e_{1,t} X_{t+h}$$

is weakly stationary with

$$\begin{aligned} E\{Z_{th}\} &= 0, \\ E\{Z_{th}^2\} &= \sigma_1^2 \gamma_{XX}(0), \end{aligned}$$

and bounded third moment. Asymptotic normality follows by a modest extension of Theorem 6.3.3.  $\blacktriangle$

**Example 6.4.1.** In Table 6.4.1 we present the sample autocorrelations and

**Table 6.4.1. Sample Correlation Functions for Suspended Sediment in Des Moines River at Boone, Iowa and Saylorville, Iowa**

$h$	Autocorrelation		Cross correlation Boone-Saylorville $\hat{r}_{12}(h)$
	Boone $\hat{r}_{11}(h)$	Saylorville $\hat{r}_{22}(h)$	
-12	0.20	0.10	0.10
-11	0.19	0.13	0.07
-10	0.22	0.16	0.08
-9	0.25	0.16	0.08
-8	0.29	0.16	0.13
-7	0.29	0.17	0.18
-6	0.29	0.15	0.21
-5	0.32	0.18	0.21
-4	0.39	0.27	0.23
-3	0.48	0.42	0.30
-2	0.62	0.60	0.40
-1	0.76	0.81	0.53
0	1.00	1.00	0.64
1	0.76	0.81	0.74
2	0.62	0.60	0.67
3	0.48	0.42	0.53
4	0.39	0.27	0.42
5	0.32	0.18	0.32
6	0.29	0.15	0.26
7	0.29	0.17	0.26
8	0.29	0.16	0.25
9	0.25	0.16	0.29
10	0.22	0.16	0.31
11	0.19	0.13	0.28
12	0.20	0.10	0.33

cross correlations for the bivariate time series  $\mathbf{Y}_t = (Y_{1t}, Y_{2t})'$ , where  $Y_{1t}$  is the logarithm of suspended sediment in the water of the Des Moines River at Boone, Iowa, and  $Y_{2t}$  is the logarithm of suspended sediment in the water at Saylorville, Iowa. Saylorville is approximately 48 miles downstream from Boone. The sample data were 205 daily observations collected from April to October 1973.

There are no large tributaries entering the Des Moines River between Boone and Saylorville, and a correlation between the readings at the two points is expected. Since Saylorville is some distance downstream, the correlation pattern should reflect the time required for water to move between the two points. In fact, the largest sample cross correlation is between the Saylorville reading at time  $t + 1$  and the Boone reading at time  $t$ . Also, estimates of  $\gamma_{12}(h)$ ,  $h > 0$ , are consistently larger than the estimates of  $\gamma_{12}(-h)$ .

The Boone time series was discussed in Section 4.5. There it was assumed that

the time series  $Y_{1,t}$  could be represented as the sum of the "true process" and a measurement error. The underlying true value  $X_{1,t}$  was assumed to be a first order autoregressive process with parameter 0.81. If we define the time series

$$W_{3t} = Y_{1,t} - 0.81Y_{1,t-1},$$

$$W_{4t} = Y_{2,t} - 0.81Y_{2,t-1},$$

the transformed true process for Boone is a sequence of uncorrelated random variables, although the observed time series  $W_{3t}$  will show a small negative first order autocorrelation.

Table 6.4.2 contains the first few estimated correlations for  $W_t$ . Note that the estimated cross correlation at zero is quite small. Under the null hypothesis that the cross correlations are zero and that the autocorrelations of  $W_{3t}$  and  $W_{4t}$  are zero

**Table 6.4.2. Sample Correlation Functions for Transformed Suspended Sediment in Des Moines River at Boone, Iowa and Saylorville, Iowa**

$h$	Autocorrelation		Cross Correlation $W_{3t}$ with $W_{4t}$
	$W_{3t}$	$W_{4t}$	
-12	0.05	0.05	0.04
-11	-0.06	0.02	-0.07
-10	0.03	0.08	0.05
-9	0.01	0.01	0.11
-8	0.11	0.04	0.03
-7	0.05	0.07	0.08
-6	-0.03	-0.06	0.10
-5	0.04	-0.12	0.03
-4	-0.01	-0.07	-0.05
-3	-0.07	0.02	-0.01
-2	0.05	-0.04	-0.04
-1	-0.17	0.15	0.06
0	1.00	1.00	0.06
1	-0.17	0.15	0.41
2	0.05	-0.04	0.24
3	-0.07	0.02	-0.02
4	-0.01	-0.07	0.01
5	0.04	-0.12	0.04
6	-0.03	-0.06	-0.10
7	0.05	0.07	0.08
8	0.11	0.04	-0.08
9	0.01	0.01	0.05
10	0.03	0.08	0.16
11	-0.06	0.02	0.08
12	0.05	0.05	0.01

after a lag of two, the estimated variance of the sample cross correlations is

$$\begin{aligned}\hat{\text{Var}}\{\hat{r}_{34}(h)\} &= \frac{1}{n} \sum_{p=-2}^2 \hat{r}_{33}(p) \hat{r}_{44}(p) \\ &= \frac{1}{205} [1 + 2(-0.17)(0.15) + 2(0.05)(-0.04)] \\ &= 0.0046.\end{aligned}$$

Under these hypotheses, the estimated standard error of the estimated cross correlation is 0.068.

The hypothesis of zero cross correlation is rejected by the estimates  $\hat{r}_{34}(1)$ ,  $\hat{r}_{34}(2)$ , since they are several times as large as the estimated standard error. The two nonzero sample cross correlations suggest that the input-output model is more complicated than that of a simple integer delay. It might be a simple delay of over one day, or it is possible that the mixing action of moving water produces a more complicated lag structure.  $\blacktriangle\blacktriangle$

## REFERENCES

- Section 6.1.** Grenander and Rosenblatt (1957), Hannan (1970), Parzen (1958, 1962).
- Section 6.2.** Anderson (1971), Bartlett (1946, 1966), Hart (1942), Kendall (1954), Kendall and Stuart (1966), Mariott and Pope (1954), von Neumann (1941, 1942).
- Section 6.3.** Anderson (1959, 1971), Anderson and Walker (1964), Diananda (1953), Eicker (1963), Hannan and Heyde (1972), Hoeffding and Robbins (1948), Moran (1947).
- Section 6.4.** Bartlett (1966), Box, Jenkins, and Reinsel (1994), Hannan (1970).

## EXERCISES

1. Let  $Y_t = \mu + X_t$ , where  $X_t = e_t + 0.4e_{t-1}$  and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables. Compute the variance of  $\bar{y}_n = n^{-1} \sum_{t=1}^n Y_t$ . What is  $\lim_{n \rightarrow \infty} n \text{Var}\{\bar{y}_n\}$ ?
2. Let the time series  $\{Y_t: t \in (1, 2, \dots)\}$  be defined by

$$Y_t = \alpha + \rho Y_{t-1} + e_t,$$

where  $Y_0$  is fixed,  $\{e_t: t \in (1, 2, \dots)\}$  is a sequence of independent  $(0, \sigma^2)$  random variables, and  $|\rho| < 1$ . Find  $E\{Y_t\}$  and  $\text{Var}\{Y_t\}$ . Show that  $Y_t$  satisfies the conditions of Theorem 6.1.1. What value does the sample mean of  $Y_t$  converge to?

3. Let  $X_t = e_t + 0.5e_{t-1}$ , where the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables. Letting

$$\tilde{\gamma}_X(h) = \frac{1}{n-h} \sum_{t=1}^{n-h} X_t X_{t+h},$$

$$\tilde{r}_X(h) = \tilde{\gamma}_X(h)/\tilde{\gamma}_X(0),$$

find  $\text{Var}\{\tilde{\gamma}_X(h)\}$  for  $h = 0, 1, 2, 3$ ;  $\text{Cov}\{\tilde{\gamma}_X(0), \tilde{\gamma}_X(h)\}$  for  $h = 1, 2, 3$ ;  $\text{Var}\{\tilde{r}_X(h)\}$  for  $h = 0, 1, 2, 3$ ; and  $\text{Cov}\{\tilde{r}_X(1), \tilde{r}_X(h)\}$  for  $h = 2, 3, 4$ .

4. Evaluate (6.2.4) for the first order autoregressive process

$$X_t = \rho X_{t-1} + e_t,$$

where  $|\rho| < 1$  and the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables.

5. Given the finite moving average

$$X_t = \sum_{j=1}^M a_j e_{t-j},$$

where the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables, is there a distance  $d = h - q$  such that  $\tilde{\gamma}(h)$  and  $\tilde{\gamma}(q)$  are uncorrelated?

6. Use the realization (10, 1, 10) and equation (6.2.2) to construct the estimated  $(3 \times 3)$  covariance matrix for a realization of size 3. Show that the resulting matrix is not positive definite. Use the fact that

$$\frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{x}_n)(X_{t+h} - \bar{x}_n) = \frac{1}{n} \sum_{j=1}^{2n-1} Z_{mj} Z_{m+h,j}$$

for  $m = 0, 1, \dots, n-1$ ,  $h = 0, 1, \dots, n-1$ , where, for  $j = 1, 2, \dots, 2n-1$ ,

$$Z_{mj} = \begin{cases} X_{j-m} - \bar{x}_n, & j = m+1, m+2, \dots, m+n, \\ 0 & \text{otherwise,} \end{cases}$$

to prove that (6.2.3.) yields an estimated covariance matrix that is always positive semidefinite.

7. Give the variance of  $\bar{x}_n$ ,  $n = 1, 2, \dots$ , for  $\{X_t: t \in (1, 2, \dots)\}$  defined by

$$X_t = \mu + e_t - e_{t-1},$$

where  $\{e_t: t \in (0, 1, 2, \dots)\}$  is a sequence of independent identically distributed  $(0, \sigma^2)$  random variables. Do you think there is a function  $w_n$  such that  $w_n(\bar{x}_n - \mu) \rightarrow N(0, 1)$ ?

8. Prove the following result, which is used in Theorem 6.3.4. If the sequence  $\{c_i\}$  is such that

$$\lim_{n \rightarrow \infty} \left( \sum_{i=1}^n c_i^2 \right)^{-1} c_n^2 = 0,$$

then

$$\lim_{n \rightarrow \infty} \sup_{1 \leq i \leq n} \left( \sum_{j=1}^n c_j^2 \right)^{-1} c_i^2 = 0.$$

9. Prove Corollary 6.3.5.

10. The data in the accompanying table are the average weekly gross hours per production worker on the payrolls of manufacturing establishments (seasonally adjusted).

Year	Quarter			
	I	II	III	IV
1948	40.30	40.23	39.97	39.70
1949	39.23	38.77	39.23	39.30
1950	39.70	40.27	40.90	40.97
1951	40.90	40.93	40.43	40.37
1952	40.63	40.33	40.60	41.07
1953	41.00	40.87	40.27	39.80
1954	39.53	39.47	39.60	39.90
1955	40.47	40.73	40.60	40.93
1956	40.60	40.30	40.27	40.47
1957	40.37	39.97	39.80	39.13
1958	38.73	38.80	39.40	39.70
1959	40.23	40.53	40.20	40.03
1960	40.17	39.87	39.63	39.07
1961	39.27	39.70	39.87	40.40
1962	40.27	40.53	40.47	40.27
1963	40.37	40.40	40.50	40.57
1964	40.40	40.77	40.67	40.90
1965	41.27	41.10	41.03	41.30
1966	41.53	41.47	41.33	41.10
1967	40.60	40.43	40.67	40.70
1968	40.60	40.63	40.83	40.77
1969	40.57	40.73	40.63	40.53
1970	40.17	39.87	39.73	39.50
1971	39.80	39.39	39.77	40.07
1972	40.30	40.67	40.67	40.87

Sources: *Business Statistics*, 1971, pp. 74 and 237, and *Survey of Current Business*, Jan. 1972, Jan. 1973. The quarterly data are the averages of monthly data.

- (a) Estimate the covariance function  $\gamma(h)$ , assuming the mean unknown.  
 (b) Estimate the correlation function  $\rho(h)$ , assuming the mean unknown.  
 (c) Using large sample theory, test the hypothesis  $H_0: \rho(1) = 0$ , assuming  $\rho(h) = 0, h > 1$ .
11. Using Hart's (1942) tables for the percentage points of  $d_v$  or Anderson's (1971) tables for  $r_v$ , obtain the percentage points for  $t_v = r_v(n+1)^{1/2}(1 - r_v^2)^{-1/2}$  for  $n = 10$  and  $15$ . Compare these values with the percentage points of Student's  $t$  with  $13$  and  $18$  degrees of freedom.
12. Using the truncation argument of Theorem 6.3.3, complete the proof of Theorem 6.4.2 by showing that

$$(n-h)^{-1/2} \sum_{i=1}^{n-h} Z_{ih}$$

converges in distribution to a normal random variable.

13. Denoting the data of Exercise 10 by  $X_{1t}$ , and that of Table 6.3.1 by  $X_{2t}$ , compute the cross covariance and cross correlation functions. Define

$$\begin{aligned} Y_{1t} &= X_{1t} - 1.53X_{1,t-1} - 0.66X_{1,t-2}, \\ Y_{2t} &= X_{2t} - 1.53X_{2,t-1} - 0.66X_{2,t-2}. \end{aligned}$$

Compute the cross covariance and cross correlation functions for  $(Y_{1t}, Y_{2t})$ . Plot the cross correlation function of  $(Y_{1t}, Y_{2t})$ . Compute the variance of  $\hat{\rho}_{Y_1 Y_2}(h)$  under the assumption that  $Y_{1t}$  is a sequence of uncorrelated random variables. Plot the standard error on your figure.

14. Let  $X_t$  be a time series with positive continuous spectral density  $f_X(\omega)$ .
- (a) Show that, given  $\epsilon > 0$ , one may define two moving average time series  $W_{1t}$  and  $W_{2t}$  with spectral densities

$$\begin{aligned} f_{W_1}(\omega) &= \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{q_1} \beta_{1j} e^{-i\omega j} \right|^2, \\ f_{W_2}(\omega) &= \frac{\sigma^2}{2\pi} \left| \sum_{j=0}^{q_2} \beta_{2j} e^{-i\omega j} \right|^2 \end{aligned}$$

such that

$$f_X(\omega) - \epsilon \leq f_{W_1}(\omega) \leq f_X(\omega) \leq f_{W_2}(\omega) \leq f_X(\omega) + \epsilon.$$

(b) Let  $\{a_t: t = 1, 2, \dots\}$  be such that

$$\begin{aligned} & \lim_{n \rightarrow \infty} \sum_{t=1}^n a_t^2 = \infty, \\ & \lim_{n \rightarrow \infty} \frac{\sum_{t=1}^n a_t a_{t+h}}{\sum_{t=1}^n a_t^2} = g(h), \quad h = 0, \pm 1, \pm 2, \dots, \\ & \sum_{h=-\infty}^{\infty} |g(h)| < \infty. \end{aligned}$$

Define  $f_g(\omega) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} g(h) e^{-i\omega h}$ . Show that

$$\lim_{n \rightarrow \infty} \text{Var} \left\{ \left( \sum_{t=1}^n a_t^2 \right)^{-1/2} \sum_{t=1}^n a_t X_t \right\} = \int_{-\pi}^{\pi} f_X(\omega) f_g(\omega) d\omega.$$

See Exercise 4.15 and Grenander and Rosenblatt (1957, Chapter 7).

**15.** Prove Lemma 6.3.2.

**16.** Let  $X_t = \sum_{j=0}^q b_j e_{t-j}$ . Show that

$$\bar{x}_n = n^{-1} \left( \sum_{j=0}^q b_j \sum_{i=1}^n e_i + \sum_{s=1}^q \sum_{j=s}^q b_j e_{1-s} - \sum_{s=0}^{q-1} \sum_{j=s+1}^q b_j e_{n-s} \right).$$

**17.** Let  $Y_t$  be the stationary autoregressive moving average

$$\sum_{j=0}^p a_j (Y_{t-j} - \mu) = \sum_{i=0}^q b_i e_{t-i}.$$

Prove that if  $\sum_{i=0}^q b_i = 0$ , then  $V\{\bar{y}_n\} = O(n^{-2})$ .

**18.** Let  $X_t$  be a covariance stationary time series satisfying

$$X_t = \sum_{j=0}^{\infty} \alpha_j e_{t-j},$$

where  $\sum_{j=0}^{\infty} |\alpha_j| < \infty$ ,  $\sum_{j=0}^{\infty} \alpha_j \neq 0$ , and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables. Show that

$$\sum_{h=-\infty}^{\infty} \gamma_X(h) = \left( \sum_{j=0}^{\infty} \alpha_j \right)^2 \sigma^2.$$

**19.** Let

$$Y_t = \sum_{j=1}^{\infty} c_j a_{t-j} \quad \text{and} \quad X_t = \sum_{i=1}^{\infty} b_i e_{t-i},$$

where  $\sum_{j=1}^{\infty} |c_j| < \infty$ ,  $\sum_{i=1}^{\infty} |b_i| < \infty$ , and

$$\begin{pmatrix} a_t \\ e_t \end{pmatrix} \sim \Pi \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{aa} & \sigma_{ae} \\ \sigma_{ea} & \sigma_{ee} \end{pmatrix} \right).$$

Prove

$$n^{-1} \sum_{t=1}^n Y_t X_t \xrightarrow{P} \sum_{j=1}^{\infty} c_j b_j \sigma_{ae}.$$

**20.** Let  $Y_t$  be a stationary time series with absolutely summable covariance function, and let

$$X_t = \sum_{j=0}^{\infty} \alpha_j Y_{t-j},$$

where  $|\alpha_j| < M\lambda^j$  for some  $M < \infty$  and  $0 < \lambda < 1$ . Show that

$$n^{-1} \sum_{t=1}^n \left( \sum_{j=0}^{t-1} \alpha_j Y_{t-j} \right)^2 = n^{-1} \sum_{t=1}^n X_t^2 + O_p(n^{-1}).$$

**21.** Construct an alternative proof for Corollary 6.3.3 by averaging both sides of the defining equation to obtain

$$\begin{aligned} \left( \sum_{j=0}^p a_j \right) (\bar{y}_n - \mu) &= \left( \sum_{i=0}^q b_i \right) \bar{e}_n + \sum_{i=0}^q b_i n^{-1} \sum_{j=1}^i (e_{j-i} - e_{n-i+j}) \\ &\quad - \sum_{j=0}^p a_j n^{-1} \sum_{k=1}^j (Y_{k-j} - Y_{n-j+k}). \end{aligned}$$

**22.** For a moving average of order  $q$ , show that  $n^{1/2} \hat{r}(h) \xrightarrow{d} N[0, 1 + 2 \sum_{k=1}^q \rho^2(k)]$  for  $|h| > q$ .

**23.** Prove the following.

**Result 6.3.1.** Let  $e_t$  be a sequence of random variables, let  $\mathcal{A}_t$  be the sigma-field generated by  $\{e_s : s \leq t\}$ , and assume the  $e_t$  satisfy

- (a)  $E\{(e_t, e_t^2 - \sigma^2, e_t^4 - \eta\sigma^4) | \mathcal{A}_{t-1}\} = (0, 0, 0)$  a.s.,
- (b)  $\operatorname{plim}_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n E\{e_t^3 e_{t-h} | \mathcal{A}_{t-1}\} = \lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n E\{e_t^3 e_{t-h}\} = 0$  for every fixed  $h > 0$ ,
- (c)  $E\{|e_t|^{4+2\delta}\} < M < \infty$  for some  $\delta > 0$ .

Then, for any fixed  $K$ ,

$$n^{1/2}[\tilde{\gamma}_\epsilon(0) - \sigma^2, \tilde{\gamma}_\epsilon(1), \dots, \tilde{\gamma}_\epsilon(K)]' \xrightarrow{\mathcal{L}} N[0, \sigma^4 \text{diag}(\eta - 1, 1, \dots, 1)].$$

**24.** Using Result 6.3.1 of Exercise 23, prove the following.

**Result 6.3.2.** Let  $Y_t = \sum_{i=0}^q \beta_i e_{t-i}$ , where the  $e_t$  satisfy the conditions of Result 6.3.1. Then, for any fixed  $K$ ,

$$n^{1/2}[\tilde{\gamma}_Y(0) - \gamma_Y(0), \tilde{\gamma}_Y(1) - \gamma_Y(1), \dots, \tilde{\gamma}_Y(K) - \gamma_Y(K)]' \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{V}),$$

where the elements of  $\mathbf{V}$  are defined by (6.2.4) of Theorem 6.2.1.

## CHAPTER 7

# The Periodogram, Estimated Spectrum

In this chapter we shall investigate estimators of the spectral density of time series with absolutely summable covariance function. The spectral density of a time series was defined in Section 4.1 as the Fourier transform of the covariance function. We also noted in Section 4.2 that the variances of the random variables defined by the Fourier coefficients of the original time series are, approximately, multiples of the spectral density. These two results suggest methods of estimating the spectral density.

The study of the Fourier coefficients was popular among economists in the 1920s and 1930s, and the student of economics may be interested in the discussion of Davis (1941) and the studies cited by Tintner (1952). Also see Granger and Hatanaka (1964) and Nold (1972). Recent applications are more common in the engineering literature. See, for example, Marple (1987).

### 7.1. THE PERIODOGRAM

Given a finite realization from a time series, we can represent the  $n$  observations by the trigonometric polynomial

$$X_t = \frac{a_0}{2} + \sum_{k=1}^m (a_k \cos \omega_k t + b_k \sin \omega_k t), \quad (7.1.1)$$

where

$$\omega_k = \frac{2\pi k}{n}, \quad k = 0, 1, 2, \dots, m,$$

$$a_k = \frac{2 \sum_{t=1}^n X_t \cos \omega_k t}{n}, \quad k = 0, 1, 2, \dots, m,$$

$$b_k = \frac{2 \sum_{t=1}^n X_t \sin \omega_k t}{n}, \quad k = 1, 2, \dots, m,$$

and we have assumed  $n$  odd and equal to  $2m + 1$ . As before, we recognize the Fourier coefficients as regression coefficients. We can use the standard regression analysis to partition the total sum of squares for the  $n$  observations. The sum of squares removed by the regression of  $X_t$  on  $\cos \omega_k t$  is the regression coefficient multiplied by the sum of cross products; that is, the sum of squares due to  $a_k$  is

$$\frac{n}{2} a_k^2 = \frac{2}{n} \left( \sum_{t=1}^n X_t \cos \omega_k t \right)^2, \quad k = 1, 2, \dots, m, \quad (7.1.2)$$

and the sum of squares removed by  $\cos \omega_k t$  and  $\sin \omega_k t$  is

$$\frac{n}{2} (a_k^2 + b_k^2) = \frac{2}{n} \left[ \left( \sum_{t=1}^n X_t \cos \omega_k t \right)^2 + \left( \sum_{t=1}^n X_t \sin \omega_k t \right)^2 \right]. \quad (7.1.3)$$

We might call the quantity in (7.1.3) the sum of squares associated with frequency  $\omega_k$ . Thus the total sum of squares for the  $n = 2m + 1$  observations may be partitioned into  $m + 1$  components. One component is associated with the mean. Each of the remaining  $m$  components is the sum of the two squares associated with the  $m$  nonzero frequencies. The partition is displayed in Table 7.1.1.

Should the number of observations be even and denoted by  $2m$ , there is only one regression variable associated with the  $m$ th frequency:  $\cos \pi t$ . Then the sum of squares for the  $m$ th frequency has one degree of freedom and is given by  $n^{-1}(\sum_{t=1}^n X_t \cos \pi t)^2 = \frac{1}{4} n a_m^2$ .

One might divide all of the sums of squares of Table 7.1.1 by the degrees of

Table 7.1.1. Analysis of Variance Table for a Sample of Size  $n = 2m + 1$

Source	Degrees of Freedom	Sum of Squares
Mean	1	$n \bar{x}_n^2 = \frac{1}{4} n a_0^2$
Frequency $\omega_1 = 2\pi/n$	2	$(n/2)(a_1^2 + b_1^2)$
Frequency $\omega_2 = 4\pi/n$	2	$(n/2)(a_2^2 + b_2^2)$
:	:	:
Frequency $\omega_m = 2\pi m/n$	2	$(n/2)(a_m^2 + b_m^2)$
Total	$n$	$\sum_{t=1}^n X_t^2$

freedom and consider the mean squares. However, it is more common to investigate the sums of squares, multiplying the sums of squares with one degree of freedom by two. The function of frequency given by these normalized sums of squares is called the *periodogram*. Thus, the periodogram is defined by

$$I_n(\omega_k) = \frac{n}{2} (a_k^2 + b_k^2), \quad k = 1, 2, \dots, m, \quad (7.1.4)$$

where  $m$  is the smallest integer greater than or equal to  $(n - 1)/2$ .

Most computer programs designed to compute the periodogram for large data sets use an algorithm based on the fast Fourier transform. See Cooley, Lewis, and Welch (1967) for references on the fast Fourier transform. Singleton (1969) gives a Fortran program for the transform, and Bloomfield (1976, p. 61) discusses the procedure.

If  $\{X_i\}$  is a sequence of normal independent  $(0, \sigma^2)$  random variables, then the  $a_k$  and  $b_k$ , being linear combinations of the  $X_i$ , will be normally distributed. Since the sine and cosine functions are orthogonal, the  $a_k$  and  $b_k$  are independent. In this case those entries in Table 7.1.1 with two degrees of freedom divided by  $\sigma^2$  are distributed as independent chi-squares with two degrees of freedom.

The periodogram may also be defined in terms of the original observations as

$$I_n(\omega_k) = \frac{2}{n} \left[ \left( \sum_{i=1}^n X_i \cos \omega_k t_i \right)^2 + \left( \sum_{i=1}^n X_i \sin \omega_k t_i \right)^2 \right], \quad k = 0, 1, \dots, m. \quad (7.1.5)$$

Note that if we define the complex coefficients  $c_k$  by

$$c_k = \frac{1}{n} \sum_{i=1}^n X_i e^{i\omega_k t_i},$$

then

$$I_n(\omega_k) = 2nc_k c_k^* = \frac{2}{n} \left| \sum_{i=1}^n X_i e^{i\omega_k t_i} \right|^2. \quad (7.1.6)$$

Thus, the periodogram ordinate at  $\omega_k$  is a multiple of the squared norm of the complex Fourier coefficient of the time series associated with the frequency  $\omega_k$ .

The periodogram is also expressible as a multiple of the Fourier transform of the estimated covariance function. If  $\omega_k \neq 0$ , we can write

$$\begin{aligned} \sum_{i=1}^n X_i \cos \omega_k t_i &= \sum_{i=1}^n (X_i - \mu) \cos \omega_k t_i, \\ \sum_{i=1}^n X_i \sin \omega_k t_i &= \sum_{i=1}^n (X_i - \mu) \sin \omega_k t_i, \end{aligned}$$

where  $\mu = E\{X_i\}$ . Therefore,

$$\begin{aligned} I_n(\omega_k) &= \frac{2}{n} \left[ \left\{ \sum_{t=1}^n (X_t - \mu) \cos \omega_k t \right\}^2 + \left\{ \sum_{t=1}^n (X_t - \mu) \sin \omega_k t \right\}^2 \right] \\ &= \frac{2}{n} \left[ \sum_{t=1}^n \sum_{j=1}^n (X_t - \mu)(X_j - \mu) \cos \omega_k t \cos \omega_k j \right. \\ &\quad \left. + \sum_{t=1}^n \sum_{j=1}^n (X_t - \mu)(X_j - \mu) \sin \omega_k t \sin \omega_k j \right] \\ &= \frac{2}{n} \left[ \sum_{t=1}^n \sum_{j=1}^n (X_t - \mu)(X_j - \mu) \cos \omega_k(t-j) \right] \end{aligned}$$

for  $k = 1, 2, \dots, m$ . In this double sum there are  $n$  combinations with  $t-j=0$ ,  $n-1$  combinations with  $t-j=1$ , and so forth. Therefore, by letting  $p=t-j$ , we obtain the following result.

**Result 7.1.1.** The  $k$ th periodogram ordinate is given by

$$I_n(\omega_k) = \begin{cases} 2n\bar{x}_n^2, & k=0, \\ 4\pi\hat{\gamma}(p), & k=1, 2, \dots, m, \end{cases} \quad (7.1.7)$$

where

$$\hat{\gamma}(\omega) = \frac{1}{2\pi} \sum_{p=-\infty}^{\infty} \frac{n-|p|}{n} \tilde{\gamma}(p) \cos \omega p \quad (-\pi \leq \omega \leq \pi), \quad (7.1.8)$$

$$\tilde{\gamma}(-p) = \tilde{\gamma}(p) = \begin{cases} \frac{1}{n-p} \sum_{j=1}^{n-p} (X_j - \mu)(X_{j+p} - \mu), & 0 \leq p \leq n-1, \\ 0, & p > n-1. \end{cases} \quad (7.1.9)$$

The coefficients  $a_k$  and  $b_k$  ( $k \neq 0$ ) that are computed using  $X_t - \bar{x}_n$  are identical to those computed using  $X_t$ . Therefore, by substituting  $X_t - \bar{x}_n$  for  $X_t$  in (7.1.5), we can also write

$$I_n(\omega_k) = 2 \sum_{p=-\infty}^{\infty} \hat{\gamma}(p) \cos \omega_k p, \quad (7.1.10)$$

where  $k \neq 0$  and

$$\hat{\gamma}(h) = \hat{\gamma}(-h) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{x}_n)(X_{t+h} - \bar{x}_n), & 0 \leq h \leq n-1, \\ 0, & h > n-1. \end{cases}$$

The function  $\hat{f}(\omega)$  of equation (7.1.8) is a continuous periodic function of  $\omega$  defined for all  $\omega$ . The periodogram has been defined only for the discrete set of points  $\omega_k = 2\pi k/n$ ,  $k = 0, 1, 2, \dots, m$ . In investigating the limiting properties of the periodogram, it is convenient to have a function defined for all  $\omega \in [0, \pi]$ . To this end we introduce the function

$$K(n, \omega) = k \quad \text{for } \frac{\pi(2k-1)}{n} < \omega \leq \frac{\pi(2k+1)}{n}, \quad k = 0, \pm 1, \pm 2, \dots,$$

and take

$$I_n(\omega) = I_n(\omega_{K(n, \omega)}) .$$

Thus,  $I_n(\omega)$  for  $\omega \in [0, \pi]$  is a step function that takes the value  $I_n(\omega_k)$  on the interval  $(\pi(2k-1)/n, \pi(2k+1)/n)$ .

The word periodogram is used with considerable flexibility in the literature. Despite the apparent conflict in terms, our definition of the periodogram as a function of frequency rather than of period is a common one. We have chosen to define the periodogram for the discrete frequencies  $\omega_k = 2\pi k/n$ ,  $k = 0, 1, 2, \dots, m$ , and to extend it to all  $\omega$  as a step function. An alternative definition of the periodogram is  $4\pi\hat{f}(\omega)$ , where  $\hat{f}(\omega)$  is given in (7.1.8), in which case one automatically has a function for all  $\omega$ . Our definition will prove convenient in obtaining the limiting properties of estimators of the spectrum.

The distributional properties of the periodogram ordinates are easily obtained when the time series is normal white noise. To establish the properties of the periodogram for other time series, we first obtain the limiting value of the expectation of  $I_n(\omega)$  for a time series with absolutely summable covariance function.

**Theorem 7.1.1.** Let  $X_t$  be a stationary time series with  $E\{X_t\} = \mu$  and absolutely summable covariance function. Then

$$\lim_{n \rightarrow \infty} E\{I_n(\omega)\} = 4\pi f(\omega), \quad \omega \neq 0,$$

$$\lim_{n \rightarrow \infty} E\{I_n(0) - 2n\mu^2\} = 4\pi f(0), \quad \omega = 0 .$$

**Proof.** Since  $E\{\tilde{\gamma}(p)\} = \gamma(p)$ , it follows from (7.1.7) that

$$E\{I_n(0)\} = 2 \sum_{h=-(n-1)}^{n-1} \frac{n-|h|}{n} \gamma(h) + 2n\mu^2,$$

$$E\{I_n(\omega_k)\} = 2 \sum_{h=-(n-1)}^{n-1} \frac{n-|h|}{n} \gamma(h) \cos \omega_k h, \quad k = 1, 2, \dots, m .$$

Now

$$\left| \sum_{h=-n}^n \frac{|h|}{n} \gamma(h) \cos \omega_k h \right| \leq \sum_{h=-n}^n \frac{|h|}{n} |\gamma(h)| .$$

and the latter sum goes to zero as  $n \rightarrow \infty$  by Lemma 3.1.4. The sequence  $g_n(\omega) = 2 \sum_{h=-n+1}^{n-1} \gamma(h) \cos \omega h$  converges uniformly to  $4\pi f(\omega)$  by Corollary 3.1.8, and  $\omega_{K(n,\omega)}$  converges to  $\omega$  by construction.  $\Delta$

The normalized coefficients  $2^{-1/2} n^{1/2} a_k$  and  $2^{-1/2} n^{1/2} b_k$  are the random variables obtained by applying to the observations the transformation discussed in Section 4.2. Therefore, for a time series with absolutely summable covariance function, these random variables are in the limit uncorrelated and have variance given by a multiple of the spectral density evaluated at the associated frequency.

The importance of this result is difficult to overemphasize. For a wide class of time series we are able to transform an observed set of  $n$  observations into a set of  $n$  statistics that are nearly uncorrelated. All except the first, the sample mean, have zero expected value. The variance of these random variables is, approximately, a simple function of the spectral density.

**Theorem 7.1.2.** Let  $X_t$  be a time series defined by

$$X_t = \sum_{j=0}^{\infty} \alpha_j e_{t-j},$$

where  $\{\alpha_j\}$  is absolutely summable and the  $e_t$  are independent identically distributed  $(0, \sigma^2)$  random variables. Let  $f_X(\omega)$  be positive for all  $\omega$ . Then, for  $\omega$  and  $\lambda$  in  $(0, \pi)$  and  $\omega \neq \lambda$ , the sequences  $[2\pi f(\omega)]^{-1} I_n(\omega)$  and  $[2\pi f(\lambda)]^{-1} I_n(\lambda)$  converge in distribution to independent chi-square random variables, each with two degrees of freedom.

**Proof.** Consider

$$\begin{aligned} 2^{-1/2} n^{1/2} a_{K(n,\omega)} &= 2^{1/2} n^{-1/2} \sum_{t=1}^n \cos \omega_{K(n,\omega)} t X_t \\ &= 2^{1/2} n^{-1/2} \sum_{t=1}^n \cos \omega_{K(n,\omega)} t \left( \sum_{j=0}^r \alpha_j e_{t-j} + \sum_{j=r+1}^{\infty} \alpha_j e_{t-j} \right), \end{aligned}$$

where

$$\lim_{r \rightarrow \infty} \text{Var} \left\{ 2^{1/2} n^{-1/2} \sum_{t=1}^n \cos \omega_{K(n,\omega)} t \sum_{j=r+1}^{\infty} \alpha_j e_{t-j} \right\} = 0$$

uniformly in  $n$ . Fixing  $r$ , we have

$$2^{1/2} n^{-1/2} \sum_{t=1}^n \cos \omega_{K(n,\omega)} t \sum_{j=0}^r \alpha_j e_{t-j} = n^{-1/2} \sum_{t=1}^n \delta_{n\omega t} e_t + R_n,$$

where

$$\delta_{n\omega t} = 2^{1/2} \sum_{j=0}^r \alpha_j \cos \omega_{K(n,\omega)}(t+j),$$

$$R_n = 2^{1/2} n^{-1/2} \left[ \sum_{j=0}^{r-1} \sum_{s=j+1}^r \alpha_s e_{-j} \cos \omega_{K(n,\omega)}(s-j) \right. \\ \left. - \sum_{j=0}^{r-1} \sum_{s=j+1}^r \alpha_s e_{n-j} \cos \omega_{K(n,\omega)}(n+s-j) \right],$$

and  $R_n$  converges in probability to zero. As  $\delta_{n\omega t}$  is uniformly bounded by (say)  $M$ , we have for  $\epsilon > 0$

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \int_{|\delta_{n\omega t_i}| > \epsilon n^{1/2}} \delta_{n\omega t_i}^2 e^2 dF(e) \\ \leq \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \int_{|e| > \epsilon M^{-1} n^{1/2}} M^2 e^2 dF(e) = 0,$$

and  $n^{-1/2} \sum_{i=1}^n \delta_{n\omega t_i} e_i$  converges in distribution to a normal random variable by the Lindeberg condition. The asymptotic normality of  $2^{-1/2} n^{1/2} a_{K(n,\omega)}$  follows by Lemma 6.3.1. The same arguments hold for a linear combination of  $n^{1/2} a_{K(n,\omega)}$ ,  $n^{1/2} b_{K(n,\omega)}$ ,  $n^{1/2} a_{K(n,\lambda)}$ ,  $n^{1/2} b_{K(n,\lambda)}$ , so that  $2^{-1/2} n^{1/2} [a_{K(n,\omega)}, b_{K(n,\omega)}, a_{K(n,\lambda)}, b_{K(n,\lambda)}]$ ,  $\omega \neq \lambda$ , converges in distribution to a multivariate normal random variable. The covariance matrix is given by Theorem 4.2.1 and is

$$\text{diag}\{2\pi f(\omega), 2\pi f(\omega), 2\pi f(\lambda), 2\pi f(\lambda)\}.$$

Hence, by Theorem 5.2.4, the limiting distribution of  $I_n(\omega)/2\pi f(\omega)$  and  $I_n(\lambda)/2\pi f(\lambda)$  is that of two independent chi-square random variables with two degrees of freedom. ▲

Thus, for many nonnormal processes, we may treat the periodogram ordinates as multiples of chi-square random variables. If the original time series is a sequence of independent  $(0, \sigma^2)$  random variables, then the periodogram ordinates all have the same expected value. However, for a time series with a nonzero autocorrelation structure, the ordinates will have different expected values. These facts have been used in constructing tests based on the periodogram.

Perhaps it is most natural to use the periodogram to search for "cycles" or "periodicities" in the data. For example, let us hypothesize that a time series is well represented by

$$X_t = \mu + A \cos \omega t + B \sin \omega t + e_t, \quad (7.1.11)$$

where the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables and  $A$  and  $B$  are fixed.

First assume that  $\omega$  is known and of the form  $2\pi k/n$ , where  $k$  is an integer. To test the hypothesis that  $A = B = 0$  against the alternative  $A \neq 0$  or  $B \neq 0$ , we can use

$$F_{2m-2}^2 = \frac{(2m-2)(a_k^2 + b_k^2)}{2 \sum_{\substack{j=1 \\ j \neq k}}^m (a_j^2 + b_j^2)},$$

where  $F_{2m-2}^2$  has the  $F$ -distribution with 2 and  $2m - 2$  degrees of freedom. Note that the sum of squares for the mean is not included in the denominator, since we postulated a general mean  $\mu$ . If  $\omega$  cannot be expressed as  $2\pi k/n$ , where  $k$  is an integer, then the regression associated with (7.1.11) can be computed and the usual regression test constructed.

We sometimes believe a time series contains a periodic component, but are unwilling to postulate the periodic function to be a perfect sine wave. For example, we may feel that a monthly time series contains a seasonal component, a large portion of which is because of a high value for December. We know that any periodic function defined on the integers, with integral period  $H$ , can be represented by

$$\frac{a_0}{2} + \sum_{k=1}^{L(H)} \left( a_k \cos \frac{2\pi k}{H} t + b_k \sin \frac{2\pi k}{H} t \right),$$

where  $L(H)$  is the largest integer less than or equal to  $H/2$ . For the monthly time series one might postulate

$$Y_t = \mu + \sum_{k=1}^6 \left( A_k \cos \frac{2\pi k}{12} t + B_k \sin \frac{2\pi k}{12} t \right) + e_t,$$

To test the hypothesis of no seasonal effect (i.e., all  $A_k$  and  $B_k$  equal zero), we form Snedecor's  $F$  as the ratio of the mean square for the six seasonal frequencies to the mean square for the remaining frequencies. Note that these tests assume  $e_t$  to be a sequence of normal independent  $(0, \sigma^2)$  random variables.

The periodogram has also been used to search for "hidden periodicities." In the above examples we postulated the frequency or frequencies of interest and hence, under the null hypothesis, the ratio of the mean squares has the  $F$ -distribution. However, we might postulate the null model

$$X_t = \mu + e_t$$

and the alternative model

$$X_t = \mu + A \cos \omega t + B \sin \omega t + e_t,$$

where  $\omega$  is unknown.

In such a case one might search out the largest periodogram ordinate and ask if

this ordinate can reasonably be considered the largest in a random sample of size  $m$  selected from a distribution function that is a multiple of a chi-square with two degrees of freedom. A statistic that can be used to test the hypothesis is

$$\xi = \left( \frac{1}{m} \sum_{k=1}^m I_n(\omega_k) \right)^{-1} I_n(L),$$

where  $I_n(L)$  is the largest periodogram ordinate in a sample of  $m$  periodogram ordinates each with two degrees of freedom. Fisher (1929) demonstrated that, for  $g > 0$ ,

$$P\{m^{-1}\xi > g\} = \sum_{j=1}^k (-1)^{j-1} \binom{m}{j} (1-jg)^{m-1},$$

where  $\xi$  is constructed from the periodogram of a sequence of normal independent  $(\mu, \sigma^2)$  random variables and  $k$  is the largest integer less than  $g^{-1}$ . A table of the distribution of  $\xi$  is given by Davis (1941). Wilks (1962, p. 529) contains a derivation of the distribution. In Table 7.1.2 we give the 1, 5, and 10 percentage points for the distribution.

Under the null hypothesis that a time series is normal white noise, the periodogram ordinates are multiples of independent chi-squares, each with two degrees of freedom. Hence, any number of other "goodness of fit" tests are available to test the hypothesis of independence. Bartlett (1966, p. 318) suggested a test based on the normalized cumulative periodogram

$$C_k = \text{cum}\{I_n(\omega_k)\} = \left[ \sum_{j=1}^m I_n(\omega_j) \right]^{-1} \sum_{j=1}^k I_n(\omega_j). \quad (7.1.12)$$

The normalized cumulative periodogram for  $k = 1, 2, \dots, m-1$  has the same distribution function as that of an ordered sample of size  $m-1$  selected from the uniform  $(0, 1)$  distribution. Therefore, if we plot the normalized periodogram as a sample distribution function and apply the Kolmogorov-Smirnov test of the hypothesis that it is a sample distribution function for a sample of  $m-1$  selected from a uniform  $(0, 1)$  distribution, we have a test of the hypothesis that the original time series is white noise. This testing procedure has been discussed by Durbin (1967, 1969).

**Example 7.1.1.** In Section 9.2 a grafted quadratic trend is fitted to United States wheat yields from 1908 to 1971. The periodogram computed for the deviations from the trend is given in Table 7.1.3. As we are working with deviations from regression, the distributions of the test statistics are only approximately those discussed above. Durbin (1969) has demonstrated that the critical points for the Kolmogorov-Smirnov-type test statistic using deviations from regression differ from those for an unaltered time series by a quantity that is  $O(n^{-1})$ .

**Table 7.1.2. Percentage Points for the Ratio of Largest Periodogram Ordinate to the Average**

Number of Ordinates	Probability of a Larger Value		
	0.10	0.05	0.01
2	1.900	1.950	1.990
3	2.452	2.613	2.827
4	2.830	3.072	3.457
5	3.120	3.419	3.943
6	3.354	3.697	4.331
7	3.552	3.928	4.651
8	3.722	4.125	4.921
9	3.872	4.297	5.154
10	4.005	4.450	5.358
15	4.511	5.019	6.103
20	4.862	5.408	6.594
25	5.130	5.701	6.955
30	5.346	5.935	7.237
40	5.681	6.295	7.663
50	5.937	6.567	7.977
60	6.144	6.785	8.225
70	6.317	6.967	8.428
80	6.465	7.122	8.601
90	6.595	7.258	8.750
100	6.711	7.378	8.882
150	7.151	7.832	9.372
200	7.458	8.147	9.707
250	7.694	8.389	9.960
300	7.886	8.584	10.164
350	8.047	8.748	10.334
400	8.186	8.889	10.480
500	8.418	9.123	10.721
600	8.606	9.313	10.916
700	8.764	9.473	11.079
800	8.901	9.612	11.220
900	9.022	9.733	11.344
1000	9.130	9.842	11.454

To test the hypothesis that the largest ordinate is the largest in a random sample of 31 estimates, we form the ratio of the fourth ordinate to the average of ordinates 1 to 31:

**Table 7.1.3. Periodogram of Deviations of United States Wheat Yields from Trend**

<i>k</i>	Period in Years	Periodogram Ordinate
1	64.0	0.54
2	32.0	5.15
3	21.3	4.74
4	16.0	53.50
5	12.8	9.86
6	10.7	3.02
7	9.1	3.20
8	8.0	2.62
9	7.1	3.56
10	6.4	2.75
11	5.8	7.99
12	5.3	5.89
13	4.9	2.08
14	4.6	2.61
15	4.3	3.44
16	4.0	1.47
17	3.8	1.44
18	3.6	0.12
19	3.4	8.77
20	3.2	1.03
21	3.0	4.00
22	2.9	0.41
23	2.8	0.10
24	2.7	17.03
25	2.6	5.73
26	2.5	0.15
27	2.4	1.90
28	2.3	3.53
29	2.2	1.91
30	2.1	6.64
31	2.1	1.96
32	2.0	12.50

$$\xi = \frac{53.50}{5.391} = 9.92 .$$

From Table 7.1.2 we see that the 1% point for this ratio is about 7.28. Thus, the null hypothesis is rejected at this level. While we may be somewhat reluctant to accept the existence of a perfect sine cycle of length 16 years on the basis of 64 observations, it is unlikely that the deviation from trend is a white noise time series.

The cumulative periodogram for the wheat yield data is displayed in Figure

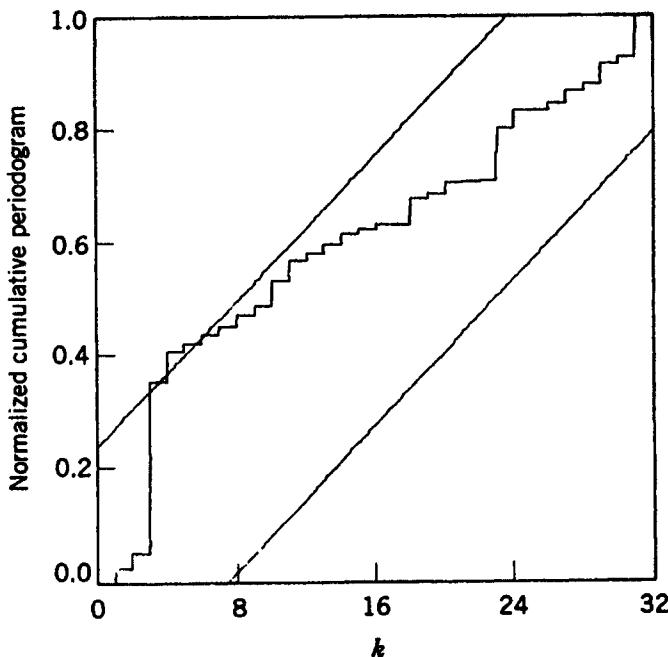


Figure 7.1.1. Normalized cumulative periodogram for deviations of wheat yield from trend.

7.1.1. We have followed the common practice of plotting the cumulative periodogram against  $k$ . The value of the cumulative periodogram for  $k$  is plotted on the interval  $(k - 1, k]$ . The upper and lower 5% bounds for the Kolmogorov-Smirnov test have been drawn in the figure. The lines are  $k/31 + 0.245$  and  $k/31 - 0.245$ , where 0.245 is the 5% point of the Kolmogorov-Smirnov statistic for sample size 31. The tables of Birnbaum (1952) indicate that for  $m - 1 > 30$ , the 95% point for the Kolmogorov-Smirnov statistic is approximately  $1.36(m - 1)^{-1/2}$  and the 99% point is approximately  $1.63(m - 1)^{-1/2}$ . In constructing the cumulative periodogram we included all 32 ordinates, even though the last ordinate has only one degree of freedom. Since the normalized cumulative periodogram passes above the upper 5% line, the data reject the hypothesis of independence, primarily because of the large ordinate at  $k = 4$ .  $\blacktriangle\blacktriangle$

## 7.2. SMOOTHING, ESTIMATING THE SPECTRUM

It is clear from the development of the distributional properties of the periodogram that increasing the sample size has little effect on the behavior of the estimated ordinate for a particular frequency. In fact, if the time series is normal  $(0, 1)$  white noise, the distribution of the periodogram ordinate is a two-degree-of-freedom chi-square independent of sample size. The number of periodogram ordinates

increases as the sample size increases, but the efficiency of the estimator for a particular frequency remains unchanged.

If the spectral density is a continuous function of  $\omega$ , it is natural to consider an average of local values of the periodogram to obtain a better estimate of the spectral density. Before treating such estimators, we present an additional result on the covariance properties of the periodogram.

In investigating local averages of the periodogram we find it convenient to make a somewhat stronger assumption about the rate at which the covariance function approaches zero. We shall assume that the covariances are such that

$$\sum_{h=-n}^n |h| |\gamma(h)| = O(n^{1/2}).$$

This is a fairly modest assumption and would be satisfied, for example, by any stationary finite autoregressive process. A sufficient condition for

$$\sum_{h=-n}^n |h| |\gamma(h)| = O(n^{1/2})$$

for a time series  $X_t$  is that

$$X_t = \sum_{j=0}^{\infty} \alpha_j e_{t-j},$$

where

$$\sum_{j=0}^{\infty} j^{1/2} |\alpha_j| < \infty$$

and  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables. This is because

$$\begin{aligned} n^{-1/2} \sum_{h=0}^n h |\gamma(h)| &\leq n^{-1/2} \sum_{h=0}^n \sum_{j=0}^{\infty} h |\alpha_j \alpha_{j+h}| \sigma^2 \\ &\leq \sum_{h=0}^n \sum_{j=0}^{\infty} h^{1/2} |\alpha_j| |\alpha_{j+h}| \sigma^2 \\ &\leq \sum_{j=0}^{\infty} \sum_{h=0}^{\infty} |\alpha_j| (j+h)^{1/2} |\alpha_{j+h}| \sigma^2 \\ &\leq \sigma^2 \sum_{j=0}^{\infty} |\alpha_j| \sum_{s=0}^{\infty} s^{1/2} |\alpha_s|. \end{aligned}$$

Before giving a theorem on the covariance properties of the periodogram, we present two lemmas useful in the proof.

**Lemma 7.2.1.** For integer  $p$ ,  $0 \leq p < n$ , and real  $L$ ,

$$\left| \sum_{j=1}^{n-p} \cos \omega_k(L+j) \right| \leq \min\{p, n-p\}$$

and

$$\left| \sum_{j=1}^{n-p} \sin \omega_k(L+j) \right| \leq \min\{p, n-p\},$$

where  $\omega_k = 2\pi k/n$ ,  $k = 1, 2, \dots, n-1$ .

**Proof.** Now

$$\begin{aligned} \sum_{j=1}^n \cos \omega_k(L+j) &= \sum_{j=1}^n (\cos \omega_k L \cos \omega_k j - \sin \omega_k L \sin \omega_k j) \\ &= (\cos \omega_k L) \sum_{j=1}^n \cos \omega_k j - (\sin \omega_k L) \sum_{j=1}^n \sin \omega_k j \\ &= 0 \quad \text{for } \omega_k \neq 0, \end{aligned}$$

where the zero sum follows from the arguments of Theorem 3.1.1. Now  $|\cos \theta| \leq 1$  for all  $\theta$ , and the result follows. The analogous argument holds for the sum of sines.  $\blacktriangle$

**Lemma 7.2.2.** For real  $L$

$$\begin{aligned} \left| \sum_{t=1}^n \sum_{j=1}^n \gamma(t-j) \cos \omega_k(L+j) \right| &\leq \sum_{p=-n+1}^{n-1} |p| |\gamma(p)|, \\ \left| \sum_{t=1}^n \sum_{j=1}^n \gamma(t-j) \sin \omega_k(L+j) \right| &\leq \sum_{p=-n+1}^{n-1} |p| |\gamma(p)|, \end{aligned}$$

where  $\omega_k = 2\pi k/n$ ,  $k = 1, 2, \dots, n-1$ .

**Proof.** We have

$$\begin{aligned} \left| \sum_{t=1}^n \sum_{j=1}^n \gamma(t-j) \cos \omega_k(L+j) \right| &= \left| \sum_{p=0}^{n-1} \sum_{j=1}^{n-p} \gamma(p) \cos \omega_k(L+j) \right. \\ &\quad \left. + \sum_{p=-n+1}^{-1} \sum_{j=-p+1}^n \gamma(p) \cos \omega_k(L+j) \right| \\ &\leq \sum_{p=-n+1}^{n-1} |p| |\gamma(p)|, \end{aligned}$$

where we have used Lemma 7.2.1. The sine result follows in a completely analogous manner.  $\blacktriangle$

**Theorem 7.2.1.** Let the time series  $X_t$  be defined by

$$X_t = \sum_{j=0}^{\infty} \alpha_j e_{t-j},$$

where the  $e_j$  are independent  $(0, \sigma^2)$  random variables with fourth moment  $\eta\sigma^4$  and

$$\sum_{j=1}^{\infty} j^{1/2} |\alpha_j| < \infty.$$

Then

$$\text{Cov}\{I_n(\omega_j), I_n(\omega_k)\} = \begin{cases} 2(4\pi)^2 f^2(0) + o(1), & \omega_j = \omega_k = 0, \\ (4\pi)^2 f^2(\omega_k) + o(1), & \omega_j = \omega_k, \omega_k \neq 0, \pi, \\ O(n^{-1}), & \omega_j \neq \omega_k. \end{cases}$$

Furthermore, for the sequence composed only of even-sized samples,

$$\text{Var}\{I_n(\pi)\} = 2(4\pi)^2 f^2(\pi) + o(1).$$

**Proof.** By the definition (7.1.6) of the periodogram, we have

$$I_n(\omega_k) = \frac{2}{n} \sum_{t=1}^n \sum_{s=1}^n X_t X_s e^{i\omega_k(t-s)}$$

and

$$\begin{aligned} & E\{I_n(\omega_k) I_n(\omega_j)\} - E\{I_n(\omega_k)\} E\{I_n(\omega_j)\} \\ &= E\left\{ \frac{4}{n^2} \sum_{t=1}^n \sum_{s=1}^n \sum_{u=1}^n \sum_{v=1}^n X_t X_s X_u X_v e^{i\omega_k(t-s)} e^{i\omega_j(u-v)} \right\} \\ &\quad - 4 \sum_{h=-(n-1)}^{n-1} \frac{n-|h|}{n} \gamma(h) e^{-i\omega_k h} \sum_{q=-(n-1)}^{n-1} \frac{n-|q|}{n} \gamma(q) e^{-i\omega_j q} \\ &= \frac{4(\eta-3)\sigma^4}{n^2} \sum_{t=1}^n \sum_{s=1}^n \sum_{u=1}^n \sum_{v=1}^n \sum_{r=-\infty}^{\infty} \alpha_r \alpha_{r+t-s} \alpha_{r+u-t-u} \alpha_{r+v-t-v} e^{i\omega_k(t-s)} e^{i\omega_j(u-v)} \\ &\quad + \frac{4}{n^2} \sum_{t=1}^n \sum_{s=1}^n \sum_{u=1}^n \sum_{v=1}^n \gamma(t-u) \gamma(s-v) e^{i\omega_k(t-u)} e^{i\omega_j(s-v)} e^{i(\omega_j + \omega_k)(u-s)} \\ &\quad + \frac{4}{n^2} \sum_{t=1}^n \sum_{s=1}^n \sum_{u=1}^n \sum_{v=1}^n \gamma(t-v) \gamma(u-s) e^{i\omega_k(t-v)} e^{i\omega_j(u-s)} e^{i(\omega_j - \omega_k)(s-v)}, \end{aligned} \tag{7.2.1}$$

where we have used (6.2.5) and  $\alpha_r = 0$ ,  $r < 0$ . Now,

$$\begin{aligned} & \frac{4}{n^2} \left| \sum_{t=1}^n \sum_{s=1}^n \sum_{u=1}^n \sum_{v=1}^n \sum_{r=-\infty}^{\infty} \alpha_r \alpha_{r+t-s} \alpha_{r+t-u} \alpha_{r+t-v} e^{i\omega_k(t-s)} e^{i\omega_j(u-v)} \right| \\ & \leq \frac{4}{n} \sum_{j=-\infty}^{\infty} \sum_{p=-\infty}^{\infty} \sum_{q=-\infty}^{\infty} \sum_{h=-\infty}^{\infty} |\alpha_j| |\alpha_{j+p}| |\alpha_{j+p+q}| |\alpha_{j+p+q+h}| \\ & = O(n^{-1}), \end{aligned}$$

by the absolute summability of  $\alpha_j$ .

If  $\omega_k = \omega_j = 0$ , or if  $n$  is even and  $\omega_k = \omega_j = \pi$ , the second term of (7.2.1) is

$$\frac{4}{n^2} \left[ \sum_{h=-(n-1)}^{n-1} (n - |h|) \gamma(h) \right]^2.$$

For  $\omega_k = \omega_j \neq 0, \pi$  the second term of (7.2.1) is

$$\begin{aligned} & \frac{4}{n^2} \sum_{t=1}^n \sum_{s=1}^n \sum_{u=1}^n \sum_{v=1}^n \gamma(t-u) \gamma(s-v) e^{i\omega_k[(t-u+2u)-(s-v+2v)]} \\ & \leq \left[ 2 \sum_{p=-(n-1)}^{n-1} \frac{|p|}{n} |\gamma(p)| \right]^2 = O(n^{-1}), \end{aligned}$$

where the inequality follows from Lemma 7.2.2. For  $\omega_k = \omega_j$  the third term reduces to

$$\left[ 2 \sum_{p=-(n-1)}^{n-1} \frac{n - |p|}{n} \gamma(p) e^{i\omega_k p} \right]^2.$$

By Lemma 7.2.2, for  $\omega_k \neq \omega_j$ ,

$$\left| \sum_{t=1}^n \sum_{u=1}^n \gamma(t-u) e^{i\omega_k(t-u)} e^{i(\omega_j + \omega_k)u} \right| \leq \sum_{p=-(n-1)}^{n-1} |p| |\gamma(p)|$$

and the absolute value of the second term shown above is less than  $[2n^{-1} \sum_{p=-(n-1)}^{n-1} |p| |\gamma(p)|]^2$ , which, by assumption, is  $O(n^{-1})$ . By a similar argument, the third term of (7.2.1) is  $O(n^{-1})$  when  $\omega_k \neq \omega_j$ . Thus, if  $\omega_k = \omega_j = \omega \neq 0, \pi$ ,

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{4}{n^2} \sum_{t=1}^n \sum_{s=1}^n \sum_{u=1}^n \sum_{v=1}^n \gamma(t-v) \gamma(u-s) e^{i\omega_{K(n,\omega)}(t-v)} e^{i\omega_{K(n,\omega)}(u-s)} \\ & = \lim_{n \rightarrow \infty} \left[ 2 \sum_{h=-(n-1)}^{n-1} \frac{n - |h|}{n} \gamma(h) e^{-i\omega h} \right]^2 = (4\pi)^2 f^2(\omega). \end{aligned}$$

If  $\omega_j = \omega_k = 0$ , then

$$\begin{aligned} & \lim_{n \rightarrow \infty} \frac{4}{n^2} \sum_{t=1}^n \sum_{s=1}^n \sum_{u=1}^n \sum_{v=1}^n \gamma(t-u)\gamma(s-v) \\ & + \lim_{n \rightarrow \infty} \frac{4}{n^2} \sum_{t=1}^n \sum_{s=1}^n \sum_{u=1}^n \sum_{v=1}^n \gamma(t-v)\gamma(u-s) \\ & = 2 \lim_{n \rightarrow \infty} \left[ 2 \sum_{h=-(n-1)}^{n-1} \frac{n-|h|}{n} \gamma(h) \right]^2 = 2(4\pi)^2 f^2(0), \end{aligned}$$

and the stated results follow.  $\blacktriangle$

**Corollary 7.2.1.** Let  $X_t$  be the time series defined in Theorem 7.2.1. In addition, assume the  $e_t$  are  $N(0, \sigma^2)$  random variables and that

$$\sum_{j=1}^{\infty} j|\alpha_j| < \infty.$$

Then

$$\text{Cov}\{I_n(\omega_j), I_n(\omega_k)\} = O(n^{-2}), \quad \omega_j \neq \omega_k.$$

**Proof.** Under normality, the first term of (7.2.1) is zero because  $\eta = 3$  for the normal distribution. Under the assumption that  $\sum_{j=1}^{\infty} j|\alpha_j| < \infty$ , we have

$$\sum_{-\infty}^{\infty} |h| |\gamma(h)| < \infty,$$

and the remaining two terms of (7.2.1) are  $O(n^{-2})$ . See Exercise 7.9.  $\blacktriangle$

Since the covariances between periodogram ordinates are small, the variance of the periodogram estimator of the spectral density at a particular frequency can be reduced by averaging adjacent periodogram ordinates. The simplest such estimator is defined by

$$\tilde{f}(\omega_k) = \frac{1}{2d+1} \sum_{j=-d}^d \hat{f}(\omega_{k+j}).$$

where

$$\hat{f}(\omega_k) = \frac{1}{4\pi} I_n(\omega_k).$$

In general, we consider the linear function of the periodogram ordinates

$$\tilde{f}(\omega_k) = \sum_{j=-d}^d W(j) \hat{f}(\omega_{k+j}), \tag{7.2.2}$$

where  $\sum_{j=-d}^d W(j) = 1$ . The weight function  $W(j)$  is typically symmetric about zero with a maximum at zero.

We may extend  $\tilde{f}(\omega_k)$  to all  $\omega$  by defining  $\tilde{f}(\omega) = \tilde{f}(\omega_{K(n,\omega)})$  where the function  $K(n, \omega)$  was introduced in Section 7.1. In practice the values of  $\tilde{f}(\omega_k)$  are often connected by lines to obtain a continuous function of  $\omega$ .

**Theorem 7.2.2.** Let the time series  $X_t$  be defined by

$$X_t = \sum_{j=0}^{\infty} \alpha_j e_{t-j},$$

where the  $e_t$  are independent  $(0, \sigma^2)$  random variables with fourth moment  $\eta\sigma^4$  and

$$\sum_{j=1}^{\infty} j^{1/2} |\alpha_j| < \infty.$$

Let  $d_n$  be an increasing sequence of positive integers satisfying

$$\lim_{n \rightarrow \infty} d_n = \infty,$$

$$\lim_{n \rightarrow \infty} \frac{d_n}{n} = 0.$$

Let the weight function  $W_n(j)$ ,  $j = 0, \pm 1, \pm 2, \dots, \pm d_n$ , satisfy

$$\begin{aligned} \sum_{j=-d_n}^{d_n} W_n(j) &= 1, \\ W_n(j) &= W_n(-j), \\ \lim_{n \rightarrow \infty} \sum_{j=-d_n}^{d_n} W_n^2(j) &= 0. \end{aligned}$$

Then  $\tilde{f}(\omega_{K(n,\omega)})$  defined by (7.2.2) satisfies

$$\lim_{n \rightarrow \infty} E\{\tilde{f}(\omega_{K(n,\omega)})\} = f(\omega),$$

$$\lim_{n \rightarrow \infty} \left( \sum_{j=-d_n}^{d_n} W_n^2(j) \right)^{-1} \text{Var}\{\tilde{f}(\omega_{K(n,\omega)})\} = \begin{cases} f^2(\omega), & \omega \neq 0, \pi, \\ 2f^2(\omega), & \omega = 0, \pi. \end{cases}$$

**Proof.** Now

$$E\{\tilde{f}(\omega_{K(n,\omega)})\} = \frac{1}{2\pi} \sum_{j=-d_n}^{d_n} W_n(j) \sum_{h=-(n-1)}^{n-1} \frac{n-|h|}{n} \gamma(h) e^{-i h \omega_{K(n,\omega) + 2\pi j/n}}$$

and

$$\lim_{n \rightarrow \infty} E\{\tilde{f}(\omega_{K(n, \omega)})\} = \lim_{n \rightarrow \infty} \sum_{j=-d_n}^{d_n} W_n(j) f(\omega_{K(n, \omega + 2\pi j/n)}).$$

Since  $f(\omega)$  is uniformly continuous, given  $\epsilon > 0$ , there exists an  $N$  such that for  $n > N$ ,  $|f(\delta) - f(\omega)| < \epsilon$  for  $\delta$  in the interval  $[\omega_{K(n, \omega - 2\pi d_n/n)}, \omega_{K(n, \omega + 2\pi d_n/n)}]$ . Therefore,

$$\lim_{n \rightarrow \infty} E\{\tilde{f}(\omega)\} = f(\omega).$$

By Theorem 7.2.1, for  $2\pi d_n/n \leq \omega \leq \pi - 2\pi d_n/n$ ,

$$\text{Var}\{\tilde{f}(\omega_{K(n, \omega)})\} = \sum_{j=-d_n}^{d_n} W_n^2(j) f^2(\omega_{K(n, \omega + 2\pi j/n)}) + o\left(\sum_{j=-d_n}^{d_n} W_n^2(j)\right),$$

and, by the argument used for the expectation,

$$\lim_{n \rightarrow \infty} \left( \sum_{j=-d_n}^{d_n} W_n^2(j) \right)^{-1} \text{Var}\{\tilde{f}(\omega)\} = f^2(\omega).$$

If  $\omega = 0$  or  $\pi$ , only  $d_n + 1$  (or  $d_n$ ) estimates  $\hat{f}(\omega_k)$  are averaged, since, for example,  $\hat{f}(2\pi/n) = \hat{f}(-2\pi/n)$ . Therefore,

$$\begin{aligned} \text{Var}\{\tilde{f}(0)\} &= W_n^2(0) f^2(0) \\ &+ 4 \sum_{j=1}^{d_n} W_n^2(j) f^2(\omega_{K(n, \omega + 2\pi j/n)}) + o\left(\sum_{j=-d_n}^{d_n} W_n^2(j)\right), \end{aligned}$$

and the result follows. ▲

**Corollary 7.2.2.** Let the assumptions of Theorem 7.2.2 be satisfied with  $W_n(j) = (2d_n + 1)^{-1}$ . Then

$$\text{Var}\{\tilde{f}(\omega_{K(n, \omega)})\} = \begin{cases} \frac{1}{2d_n + 1} f^2(\omega) + o(d_n^{-1}), & \omega \neq 0, \pi, \\ \frac{2}{2d_n + 1} f^2(\omega) + o(d_n^{-1}), & \omega = 0, \pi. \end{cases}$$

Perhaps it is worthwhile to pause and summarize our results for the periodogram estimators. First, the periodogram ordinates  $I_n(\omega_j)$  are the sums of squares associated with sine and cosine regression variables for the frequency  $\omega_j$ . For a

wide class of time series the  $I_n(\omega_j)$  are approximately independently distributed as  $[2\pi f(\omega_j)]\chi^2_2$ , that is, as a multiple of a chi-square with two degrees of freedom (Theorems 7.1.1 and 7.2.1).

If  $f(\omega)$  is a continuous function of  $\omega$ , then, for large  $n$ , adjacent periodogram ordinates have approximately the same mean and variance. Therefore, an average of  $2d + 1$  adjacent ordinates has approximately the same mean and a variance  $(2d + 1)^{-1}$  times that of the original ordinates. It is possible to construct a sequence of estimators based on realizations of increasing size wherein the number of adjacent ordinates being averaged increases (at a slower rate than  $n$ ) so that the average, when divided by  $4\pi$ , is a consistent estimator of  $f(\omega)$  (Theorem 7.2.2). The consistency result is less than fully appealing, since it does not tell us how many terms to include in the average for any particular time series. Some general conclusions are possible. For most time series the average of the periodogram ordinates will be a biased estimator of  $4\pi f(\omega)$ . For the largest portion of the range of most functions, this bias will increase as the number of terms being averaged increases. On the other hand, we can expect the variance of the average to decline as additional terms are added. (See Exercise 7.11.) Therefore, the mean square error of our average as an estimator of the spectral density will decline as long as the increase in the squared bias is less than the decrease in the variance. The white noise time series furnishes the limiting case. Since the spectral density is a constant function, the best procedure is to include all ordinates in the average [i.e., to use  $\tilde{\chi}(0)$  to estimate  $2\pi f(\omega)$  for all  $\omega$ ]. For a time series of known structure we could determine the optimum number of terms to include in the weight function. However, if we possess that degree of knowledge, the estimation problem is no longer of interest. The practitioner, as he works with data, will develop certain rules of thumb for particular kinds of data. For data with unknown structure it would seem advisable to construct several averages of varying length before reaching conclusions about the nature of the spectral density.

The approximate distributional properties of the smoothed periodogram can be used to construct a confidence interval for the estimated spectral density. Under the conditions of Theorem 7.2.2, the  $I_n(\omega_k)$  are approximately distributed as independent chi-squares, and therefore  $\tilde{f}(\omega)$  is approximately distributed as a linear combination of chi-square random variables. One common approximation to such a distribution is a chi-square distribution with degrees of freedom determined by the variance of the distribution.

**Result 7.2.1.** Let  $X_i$  satisfy the assumptions of Theorem 7.2.2 and let  $f(\omega) > 0$ . Then, for  $\pi d_n/n < \omega < \pi(1 - d_n/n)$ ,  $f^{-1}(\omega)\tilde{f}(\omega)$  is approximately distributed as a chi-square random variable divided by its degrees of freedom  $\nu$ , where

$$\nu = 2 \left[ \sum_{j=-d_n}^{d_n} W_n^2(j) \right]^{-1}.$$

An approximate  $1 - \alpha$  level confidence interval for  $f(\omega)$  can be constructed as

$$\frac{\nu \tilde{f}(\omega)}{\chi_{\nu, \alpha/2}^2} \leq f(\omega) \leq \frac{\nu \tilde{f}(\omega)}{\chi_{\nu, 1-(\alpha/2)}^2}, \quad (7.2.3)$$

where  $\chi_{\nu, \alpha/2}^2$  is the  $\alpha/2$  tabular value for the chi-square distribution with  $\nu$  degrees of freedom. (Point exceeded with probability  $\alpha/2$ .)

Since the variance of  $\tilde{f}(\omega)$  is a multiple of  $[f(\omega)]^2$ , the logarithm of  $\tilde{f}(\omega)$  for time series with  $f(\omega)$  strictly positive will have approximately constant variance,

$$\text{Var}\{\log \tilde{f}(\omega)\} \doteq [f(\omega)]^{-2} \text{Var}[f(\omega)] \doteq \frac{2}{\nu}.$$

Therefore,  $\log \tilde{f}(\omega)$  is often plotted as a function of  $\omega$ . Approximate confidence intervals for  $\log \tilde{f}(\omega)$  are given by

$$\log \tilde{f}(\omega) + \log\left(\frac{\nu}{\chi_{\nu, \alpha/2}^2}\right) \leq \log f(\omega) \leq \log \tilde{f}(\omega) + \log\left(\frac{\nu}{\chi_{\nu, 1-(\alpha/2)}^2}\right). \quad (7.2.4)$$

**Example 7.2.1.** As an illustration of spectral estimation, we consider an artificially generated time series. Table 7.2.1 contains 100 observations for the time series

$$X_t = 0.7 X_{t-1} + e_t,$$

where the  $e_t$  are computer generated normal independent  $(0, 1)$  random variables. The periodogram for this sample is given in Figure 7.2.1. We have connected the ordinates with straight lines. The approximate expected value of the periodogram ordinates is  $4\pi f(\omega) = 2(1.49 - 1.4 \cos \omega)^{-1}$  and has also been plotted in the figure. Both the average value and the variance are much larger for ordinates associated with the smaller frequencies.

We have labeled the frequency axis in radians. Thus, the fastest frequency we can observe is  $\pi$ , which is 0.50 cycles per time unit. This corresponds to a cycle with a period of two time units. In applications there are natural time units such as hours, months, or years, and one may choose to label the axis in terms of the frequency in these units.

In Figure 7.2.2 we display the smoothed periodogram where the smoothed value at  $\omega_k$  is

$$\bar{I}(\omega_k) = \frac{1}{5} \sum_{j=-2}^2 I_n(\omega_{k-j}).$$

The smoothed periodogram roughly assumes the shape of the spectral density. Because the standard error of the smoothed estimator for  $2 < k < 48$  is about 0.45 of the true value, there is considerable variability in the plot. The smoothing

**Table 7.2.1. One Hundred Observations from a First Order Autoregressive Time Series with  $\rho = 0.7$**

	First 25	Second 25	Third 25	Fourth 25
1	0.874	-0.613	-0.366	-0.955
2	0.850	0.110	-1.420	-0.948
3	2.345	0.113	-0.183	0.046
4	2.501	-0.308	-0.044	0.091
5	1.657	0.723	-0.391	0.254
6	1.649	-0.257	-0.095	2.750
7	2.498	1.051	-0.971	1.673
8	1.330	0.803	0.371	2.286
9	1.307	0.116	-1.622	1.220
10	3.404	-1.454	-2.941	-0.256
11	2.445	0.296	-2.814	0.252
12	2.805	1.501	-1.784	0.325
13	1.639	0.880	-2.471	-0.338
14	1.240	-0.672	-3.508	0.378
15	1.116	0.436	-2.979	0.127
16	0.448	0.930	-0.779	-2.006
17	0.377	1.168	0.869	-2.380
18	-0.488	1.999	1.786	-2.024
19	-0.960	1.376	0.123	-1.085
20	-0.579	1.613	0.093	1.037
21	-1.674	2.030	-0.731	-0.467
22	-0.366	0.616	-1.253	-0.794
23	-0.922	0.667	-2.213	-0.493
24	-1.174	0.707	-0.252	-0.157
25	-1.685	1.029	0.403	0.659

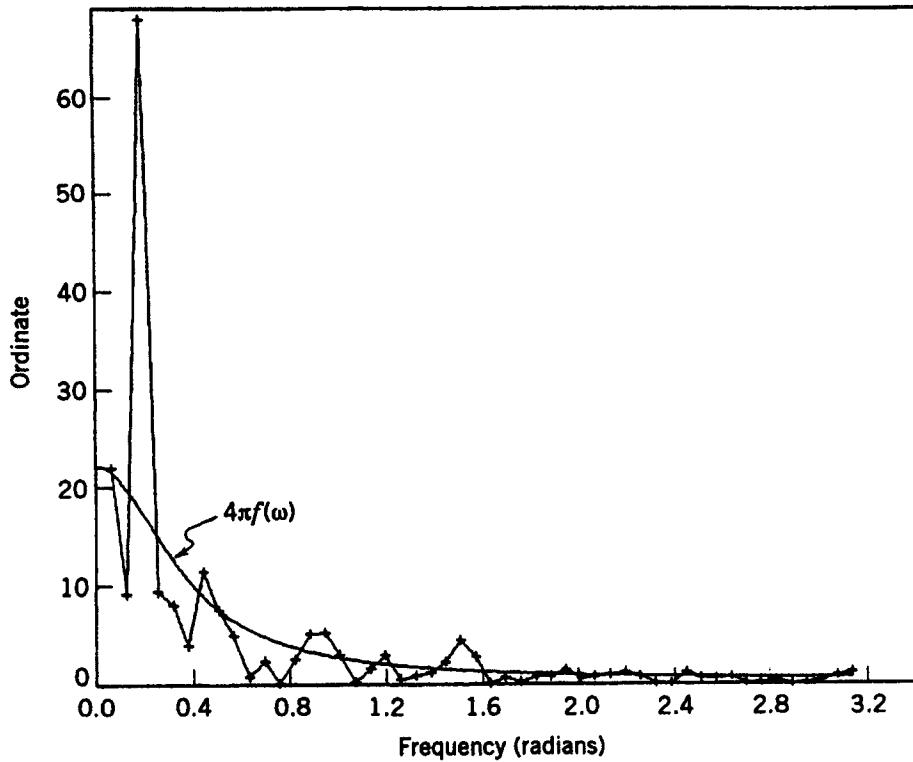
introduces a correlation between observations less than  $2d + 1$  units apart. In one sense this was the objective of the smoothing, since it produces a plot that more nearly approximates that of the spectral density. On the other hand, if the estimated spectral density is above the true density, it is apt to remain so for some distance.

To compute smoothed values near zero and  $\pi$ , the periodic nature of the function  $f(\omega)$  is used. In most applications the mean is estimated and therefore  $I_n(0)$  is not computed. We follow this practice in our example. To compute the smoothed value at zero we set  $I_n(\omega_0) = I_n(\omega_1)$  and compute

$$\bar{I}(\omega_0) = \frac{1}{3} [I_{100}(\omega_{-2}) + I_{100}(\omega_{-1}) + I_{100}(\omega_0) + I_{100}(\omega_1) + I_{100}(\omega_2)],$$

which, by the even periodic property of  $f(\omega)$ , is given by

$$\bar{I}(\omega_0) = \frac{1}{3} [I_{100}(\omega_0) + 2I_{100}(\omega_1) + 2I_{100}(\omega_2)].$$



**Figure 7.2.1.** Periodogram computed from the 100 autoregressive observations of Table 7.2.1 compared with  $4\pi f(\omega)$ .

In our example, replacing  $I_{100}(\omega_0)$  by  $I_{100}(\omega_1)$ ,

$$\bar{I}(\omega_0) = \frac{1}{3}[3(22.004) + 2(9.230)] = 16.894.$$

Similarly

$$\begin{aligned}\bar{I}(\omega_2) &= \frac{1}{3}[I_{100}(\omega_0) + I_{100}(\omega_1) + I_{100}(\omega_2) + I_{100}(\omega_3) + I_{100}(\omega_4)] \\ &= \frac{1}{3}[2(22.004) + (9.230) + (67.776) + (9.360)] = 26.075.\end{aligned}$$

As the sample size is even, there is a one-degree-of-freedom periodogram ordinate for  $\pi$ . The smoothed estimate at  $\pi$  is

$$\begin{aligned}\bar{I}(\omega_{50}) &= \frac{1}{3}[I_{100}(\omega_{50}) + 2I_{100}(\omega_{49}) + 2I_{100}(\omega_{48})] \\ &= \frac{1}{3}[(1.166) + 2(0.870) + 2(0.440)] = 0.757.\end{aligned}$$

In Figure 7.2.3 we plot the logarithm of the average of 11 periodogram ordinates ( $d = 5$ ). The 95% confidence intervals are also plotted in Figure 7.2.3.

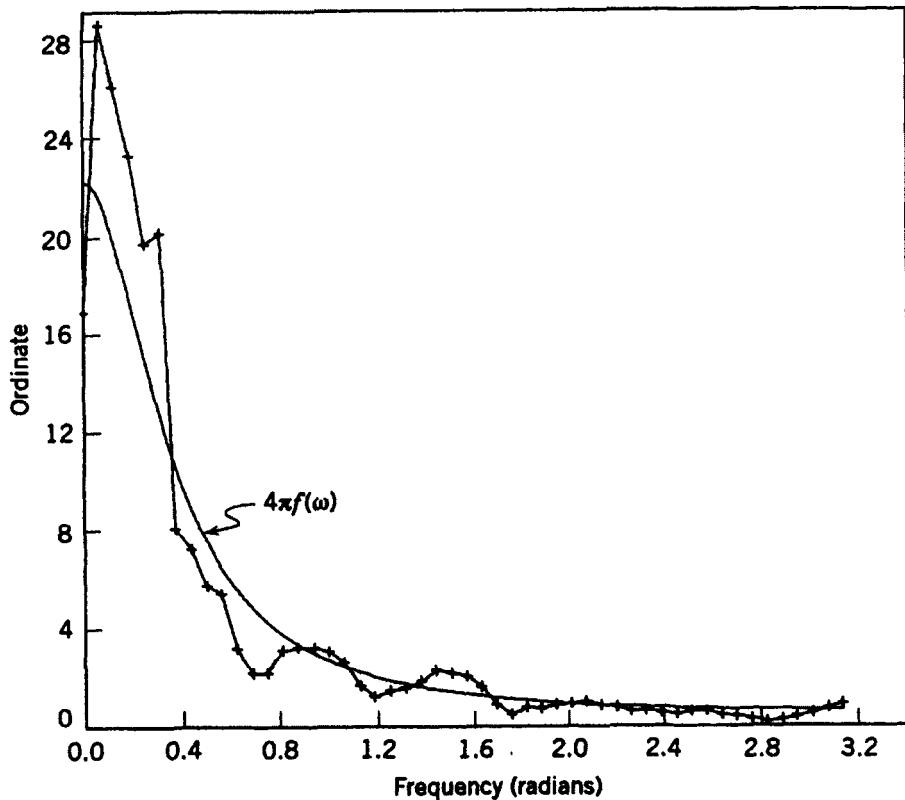


Figure 7.2.2. Smoothed periodogram ( $d = 2$ ) computed from 100 autoregressive observations of Table 7.2.1 compared with  $4\pi f(\omega)$ .

They were constructed using (7.2.4), so that the upper bound is

$$\log \bar{I}(\omega) + \log \left( \frac{22}{10.98} \right)$$

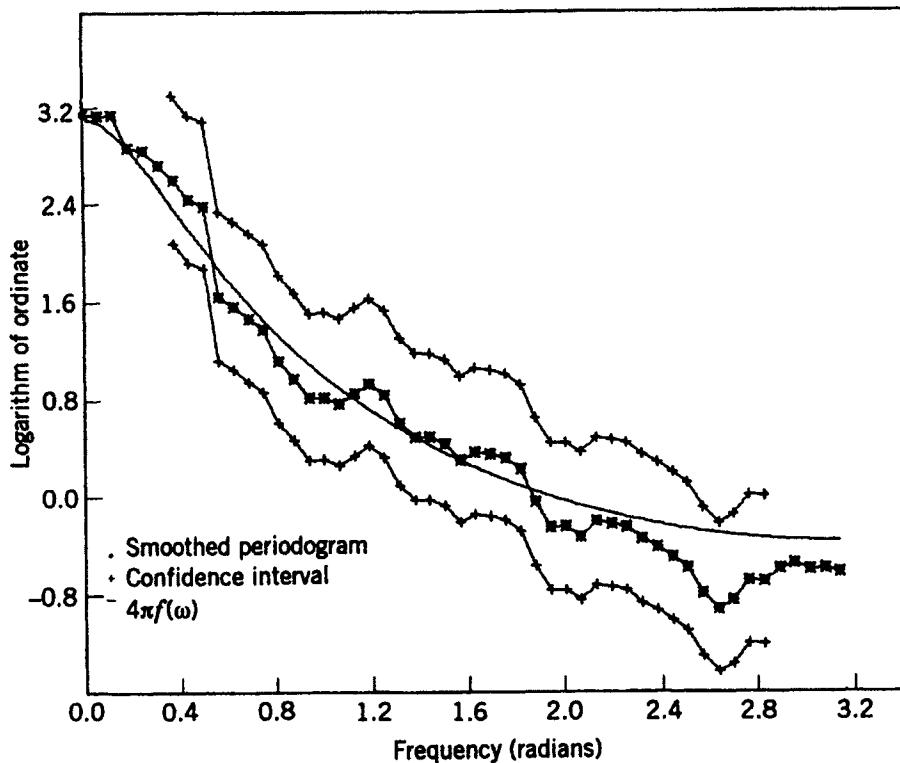
and the lower bound is

$$\log \bar{I}(\omega) + \log \left( \frac{22}{36.78} \right).$$

This interval is appropriate for  $6 \leq k \leq 44$  and adequate for  $k = 45$ . Confidence intervals for other values of  $k$  could be constructed using the variance of the estimator. For example, the smoothed value at zero is

$$\bar{I}(0) = \frac{1}{11} \left[ 3I_{100}(\omega_1) + 2 \sum_{j=2}^5 I_{100}(\omega_j) \right],$$

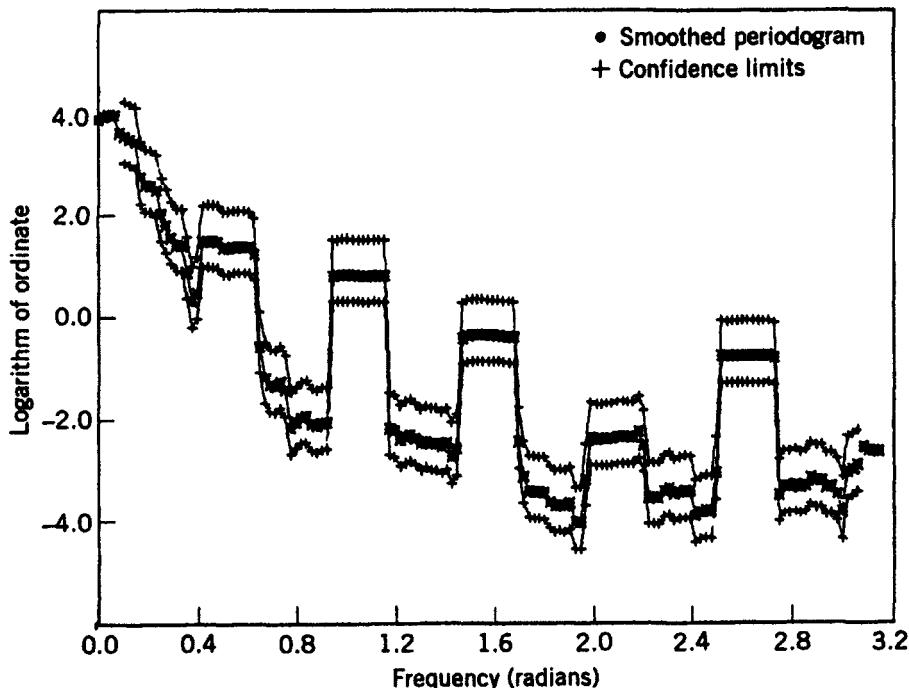
and the variance is approximately  $\frac{25}{121} [4\pi f(0)]^2 = 0.21[4\pi f(0)]^2$ . Since the



**Figure 7.2.3.** Logarithm of smoothed periodogram ( $d = 5$ ) and confidence interval for logarithm of  $4\pi f(\omega)$  computed from the 100 autoregressive observations of Table 7.2.1.

variance of a chi-square with 10 degrees of freedom divided by its degrees of freedom is 0.20, we can establish a confidence interval for  $4\pi f(0)$  using the critical points for a chi-square with 10 degrees of freedom. A similar approach can be used for  $1 \leq k \leq 5$  and  $46 \leq k \leq 50$ .

**Example 7.2.2.** As a second example of spectral estimation, we consider the United States monthly unemployment rate from October 1949 to September 1974. Figure 7.2.4 is a plot of the logarithm of the smoothed periodogram using the rectangular weights and  $d = 5$ . Also included in the figure are lines defining the 95% confidence interval. This plot displays characteristics typical of many economic time series. The smoothed periodogram is high at small frequencies, indicating a large positive autocorrelation for observations close together in time. Second, there are peaks at the seasonal frequencies  $\pi$ ,  $5\pi/6$ ,  $2\pi/3$ ,  $\pi/2$ ,  $\pi/3$ ,  $\pi/6$ , indicating the presence of a seasonal component in the time series. The periodogram has been smoothed using the simple average, and the flatness of the seasonal peaks indicates that they are being dominated by the center frequencies. That is, the shape of the peak is roughly that of the weights being used in the smoothing. Table 7.2.2 contains the ordinates at and near the seasonal frequencies.



**Figure 7.2.4.** Logarithm of smoothed periodogram for monthly United States unemployment rate for October 1949 through September 1974 ( $n = 300$ ,  $d = 5$ ).

With the exception of  $\pi$ , the ordinates at the seasonal frequencies are much larger than the other ordinates. Also, it seems that the ordinates adjacent to the seasonal frequencies are larger than those farther away. This suggests that the "seasonality" in the time series is not perfectly periodic. That is, more than the six seasonal frequencies are required to completely explain the peaks in the estimated spectral density.  $\blacktriangle \blacktriangle$

### 7.3. OTHER ESTIMATORS OF THE SPECTRUM

The smoothed periodogram is a weighted average of the Fourier transform of the sample autocovariance. An alternative method of obtaining an estimated spectral density is to apply weights to the estimated covariance function and then transform the "smoothed" covariance function. The impetus for this procedure came from a desire to reduce the computational costs of computing covariances for realizations with very large numbers of observations. Thus, the weight function has traditionally been chosen to be nonzero for the first few autocovariances and zero otherwise. The development of computer routines using the fast Fourier transform reduced the cost of computing finite Fourier transforms and reduced the emphasis on the use of the weighted autocovariance estimator of the spectrum.

**Table 7.2.2. Periodogram Ordinates Near Seasonal Frequencies**

Frequency (Radians)	Ordinate	Frequency (Radians)	Ordinate
0.461	0.127	2.032	0.025
0.482	0.607	2.052	0.021
0.503	5.328	2.073	0.279
0.523	31.584	2.094	0.575
0.545	2.904	2.115	0.012
0.565	0.557	2.136	0.011
0.586	0.069	2.157	0.008
0.984	0.053	2.555	0.028
1.005	0.223	2.576	0.010
1.026	0.092	2.597	0.286
1.047	23.347	2.618	4.556
1.068	0.310	2.639	0.038
1.089	0.253	2.660	0.017
1.110	0.027	2.681	0.022
1.508	0.191	3.079	0.008
1.529	0.041	3.100	0.003
1.550	0.177	3.120	0.294
1.571	6.402	3.142	0.064
1.592	0.478		
1.613	0.142		
1.634	0.012		

Let  $w(x)$  be a bounded even continuous function satisfying

$$\begin{aligned} w(0) &= 1, \\ w(x) &= 0, \quad |x| > 1, \\ |w(x)| &\leq 1 \quad \text{for all } x. \end{aligned} \tag{7.3.1}$$

Then a weighted estimator of the spectral density is

$$\tilde{f}(\omega) = \frac{1}{2\pi} \sum_{h=-g_n}^{g_n} w\left(\frac{h}{g_n}\right) \hat{\gamma}(h) e^{-i\omega h}, \tag{7.3.2}$$

where  $g_n \leq n$  is the chosen point of truncation and

$$\hat{\gamma}(h) = \hat{\gamma}(-h) = \frac{1}{h} \sum_{i=1}^{n-h} (X_i - \bar{x}_n)(X_{i+h} - \bar{x}_n), \quad h \geq 0.$$

From Section 3.4 we know that the transform of a convolution is the product of the transforms. Conversely, the transform of a product is the convolution of the

transforms. In the current context we define the following function (or transform):

$$\begin{aligned}\hat{f}(\omega) &= \frac{1}{2\pi} \sum_{h=-n+1}^{n-1} \hat{\gamma}(h) e^{-i\omega h} \\ &= \frac{1}{2\pi} \left[ \hat{\gamma}(0) + 2 \sum_{h=1}^{n-1} \hat{\gamma}(h) \cos \omega h \right].\end{aligned}$$

The function  $\hat{f}(\omega)$  is a continuous function of  $\omega$  and is an unweighted estimator of the continuous spectral density. By the uniqueness of Fourier transforms, we can write

$$\hat{\gamma}(h) = \int_{-\pi}^{\pi} \hat{f}(\omega) e^{i\omega h} d\omega, \quad h = 0, \pm 1, \pm 2, \dots, \pm(n-1).$$

We define the transform of  $w(x)$  similarly as

$$W(\omega) = \frac{1}{2\pi} \sum_{h=-n+1}^{n-1} w\left(\frac{h}{g_n}\right) e^{-i\omega h},$$

where  $W(\omega)$  is also a continuous function. It follows that

$$w\left(\frac{h}{g_n}\right) = \int_{-\pi}^{\pi} W(\omega) e^{i\omega h} d\omega, \quad h = 0, \pm 1, \pm 2, \dots, \pm(n-1).$$

Note that  $w(0) = 1$  means that  $\int_{-\pi}^{\pi} W(s) ds = 1$ . Then, by Exercise 3.14,

$$\begin{aligned}\tilde{f}(\omega) &= \frac{1}{2\pi} \sum_{h=-n+1}^{n-1} w\left(\frac{h}{g_n}\right) \hat{\gamma}(h) e^{-i\omega h} \\ &= \int_{-\pi}^{\pi} W(s) \hat{f}(\omega - s) ds.\end{aligned}$$

One should remember that both  $W(s)$  and  $\hat{f}(s)$  are even periodic functions. Thus, the estimated spectrum obtained from the weighted estimated covariances  $w(h/g_n) \hat{\gamma}(h)$  is a weighted average (convolution) of the spectrum estimated from the original covariances, where the weight function is the transform of weights applied to the covariances.

The function  $W(\omega)$  is called the *kernel* or *spectral window*. The weight function  $w(x)$  is often called the *lag window*.

**Theorem 7.3.1.** Let  $g_n$  be a sequence of positive integers such that

$$\begin{aligned}\lim_{n \rightarrow \infty} g_n &= \infty, \\ \lim_{n \rightarrow \infty} n^{-1} g_n &= 0;\end{aligned}$$

and let  $X_t$  be a time series defined by

$$X_t = \sum_{j=-\infty}^{\infty} \alpha_j e_{t-j},$$

where the  $e_t$  are independent  $(0, \sigma^2)$  random variables with fourth moment  $\eta\sigma^4$ , and  $\{\alpha_j\}$  is absolutely summable. Let

$$\tilde{\gamma}(h) = \tilde{\gamma}(-h) = \frac{1}{n-h} \sum_{t=1}^{n-h} X_t X_{t+h}, \quad h \geq 0.$$

and

$$\tilde{f}(\omega) = \frac{1}{2\pi} \sum_{h=-g_n}^{g_n} w\left(\frac{h}{g_n}\right) \left(\frac{n-|h|}{n}\right) \tilde{\gamma}(h) e^{-i\omega h},$$

where  $w(x)$  is a bounded even continuous function satisfying the conditions (7.2.5). Then

$$E\{\tilde{f}(\omega)\} = \frac{1}{2\pi} \sum_{h=-g_n}^{g_n} \frac{n-|h|}{n} w\left(\frac{h}{g_n}\right) \gamma(h) e^{-i\omega h}$$

and

$$\lim_{n \rightarrow \infty} \frac{n}{g_n} \text{Cov}\{\tilde{f}(\omega), \tilde{f}(\lambda)\} = \begin{cases} 2f^2(\omega) \int_{-1}^1 w^2(x) dx, & \omega = \lambda = 0, \pi, \\ f^2(\omega) \int_{-1}^1 w^2(x) dx, & \omega = \lambda \neq 0, \pi, \\ 0, & \omega \neq \lambda. \end{cases}$$

**Proof.** See, for example, Anderson (1971, Chapter 9) or Hannan (1970, Chapter 5).  $\blacktriangle$

The choice of a truncation point  $g_n$  for a particular sample of  $n$  observations is not determined by the asymptotic theory of Theorem 7.2.3. As in our discussion of the smoothed periodogram, the variance generally increases as  $g_n$  increases, but the bias in the estimator will typically decrease as  $g_n$  increases. It is possible to determine the order of the bias as a function of the properties of the weight function  $w(x)$  and the speed with which  $\gamma(h)$  approaches zero [see Parzen (1961)], but this still does not solve the problem for a given sample and unknown covariance structure.

Approximate confidence limits for the spectral density can be constructed in the same manner as that used for the smoothed periodogram estimator.

**Result 7.3.1.** Let  $X_t$  satisfy the assumptions of Theorem 7.3.1, let  $w(x)$  satisfy

(7.3.1), and let  $f(\omega) > 0$ . Then for  $\pi\nu/2n < \omega < \pi - \pi\nu/2n$ ,  $f^{-1}(\omega)\tilde{f}(\omega)$  is approximately distributed as a chi-square random variable divided by its degrees of freedom  $\nu$ , where

$$\nu = \frac{2n}{\int_{-1}^1 w^2(x) dx}.$$

Considerable research has been conducted on the weights  $w(x)$  to use in estimating the spectrum. One of the simplest windows is obtained by truncating the sequence of autocovariances at  $g_n$ . This procedure is equivalent to applying the window

$$w(x) = \begin{cases} 1, & |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases} \quad (7.3.3)$$

The function  $w(x)$  is sometimes called a *truncated* or *rectangular* window. While  $w(x)$  does not meet the conditions of Theorem 7.3.1, it can be shown that the conclusion holds. The spectral window for the function (7.3.3),

$$W(s) = \frac{1}{2\pi} \frac{\sin(g_n + \frac{1}{2})s}{\sin(s/2)},$$

takes on negative values, and it is possible that the weighted average  $\tilde{f}(\omega)$  will be negative for some  $\omega$ . This is generally considered an undesirable attribute, since  $f(\omega) \geq 0$  for all  $\omega$ .

Bartlett (1950) suggested splitting an observed time series of  $n$  observations into  $p$  groups of  $M$  observations each. The periodogram is then computed for each group, and the estimator for the ordinate associated with a particular frequency is taken to be the average of the  $p$  estimators; that is,

$$4\pi\tilde{f}(\omega_k) = \frac{1}{p} \sum_{s=0}^{p-1} I_{Ms}(\omega_k),$$

where  $I_{Ms}(\omega_k)$  is the estimator for the ordinate at frequency  $\omega_k$  obtained from the  $s$ th subsample.

Bartlett's estimator is closely related to the estimator with lag window

$$w(x) = \begin{cases} 1 - |x|, & |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

This window has been called *modified Bartlett* or *triangular*. Setting  $g_n = M$ , the spectral window is given by

$$\begin{aligned} W_B(\omega) &= \frac{1}{2\pi} \sum_{h=-M}^M \frac{M-|h|}{M} \cos \omega h \\ &= \frac{1}{2\pi M} \sum_{h=0}^{M-1} \sum_{j=-h}^h \cos j\omega. \end{aligned}$$

Using Lemma 3.1.2,

$$W_B(\omega) = \frac{\sin^2(M/2)\omega}{2\pi M \sin^2(\omega/2)}.$$

To evaluate the variance of the modified Bartlett estimator, we have

$$\sum_{h=-M}^M w^2\left(\frac{h}{M}\right) = \sum_{h=-M}^M \left(\frac{M-|h|}{M}\right)^2 \doteq \frac{2}{3} M.$$

Thus, for the modified Bartlett estimator with covariances truncated at  $M$ , the variance of the estimated spectrum is approximately

$$\text{Var}\{\tilde{f}(\omega)\} \doteq \frac{2M}{3n} f^2(\omega).$$

Blackman and Tukey (1959) suggested the weight function

$$w(x) = \begin{cases} 1 - 2a + 2a \cos \pi x, & |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

The use of the window with  $a = 0.23$  they called "hamming", and the use of the window with  $a = 0.25$ , "hanning." Parzen (1961) suggested the weight function

$$w(x) = \begin{cases} 1 - 6x^2 + 6|x|^3, & |x| \leq \frac{1}{2}, \\ 2(1 - |x|)^3, & \frac{1}{2} \leq |x| \leq 1, \\ 0 & \text{otherwise.} \end{cases}$$

This kernel will always produce nonnegative estimators. Brillinger (1975, Chapter 3) contains a discussion of these and other kernels.

Another method of estimating the spectral density is to estimate an autoregressive moving average model for the time series and then use the spectral density defined by that model as the estimator of the spectral density. Because of its simplicity, the pure autoregressive model is often used. The smoothness of the estimated spectral density is a function of the order of the model fit. A number of criteria for determining the order have been suggested. See Akaike (1969a), Parzen (1974, 1977), Marple (1987), Newton (1988), and the references cited in Section 8.4.

The Burg method of fitting the autoregressive process, discussed in Section 8.2.2, is often used in estimating it. The method has the advantage that the roots of the autoregressive process are always less than one and the computations are such that it is relatively easy to consider autoregressive models of different orders.

## 7.4. MULTIVARIATE SPECTRAL ESTIMATES

We now investigate estimators of the spectral parameters for vector time series. Let  $n$  observations be available on the bivariate time series  $\mathbf{Y}_t = (Y_{1t}, Y_{2t})'$ . The Fourier coefficients  $a_{1k}$ ,  $b_{1k}$ ,  $a_{2k}$ , and  $b_{2k}$  can be computed for the two time series

by the formulas following (7.1.1). Having studied estimators of  $f_{11}(\omega)$  and  $f_{22}(\omega)$  in the preceding sections, it remains only to investigate estimators of  $f_{12}(\omega)$  and of the associated quantities such as the phase spectrum and the squared coherency.

Recalling the transformation introduced in Section 4.2, which yields the normalized Fourier coefficients, we shall investigate the joint distributional properties of

$$\begin{aligned} 2^{-1}n^{1/2}(a_{1k} + i b_{1k}) &= n^{-1/2} \sum_{t=1}^n Y_{1t} e^{i\omega_k t}, \\ 2^{-1}n^{1/2}(a_{2k} + i b_{2k}) &= n^{-1/2} \sum_{t=1}^n Y_{2t} e^{i\omega_k t}, \end{aligned} \quad (7.4.1)$$

where  $\omega_k = 2\pi k/n$ ,  $k = 0, 1, \dots, n - 1$ . Define a transformation matrix  $\mathbf{H}$  by

$$\mathbf{H} = \begin{pmatrix} \mathbf{G} & \mathbf{O} \\ \mathbf{O} & \mathbf{G} \end{pmatrix}, \quad (7.4.2)$$

where  $\mathbf{G}$  is an  $n \times n$  matrix with rows given by

$$g_k = n^{-1/2}[1, e^{i2\pi k/n}, e^{i2\pi 2k/n}, \dots, e^{i2\pi(n-1)k/n}], \quad k = 0, 1, \dots, n - 1.$$

The matrix  $\mathbf{G}$  was introduced in (4.2.6), and is the matrix that will diagonalize a circular matrix. Let

$$\mathbf{V} = \begin{pmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{pmatrix} \quad (7.4.3)$$

be the  $2n \times 2n$  covariance matrix of  $\mathbf{y} = (\mathbf{y}'_1, \mathbf{y}'_2)'$ , where

$$\begin{aligned} \mathbf{y}'_1 &= (Y_{11}, Y_{12}, \dots, Y_{1n}), \\ \mathbf{y}'_2 &= (Y_{21}, Y_{22}, \dots, Y_{2n}), \end{aligned}$$

and

$$\mathbf{V}_{12} = \begin{pmatrix} \gamma_{12}(0) & \gamma_{12}(1) & \gamma_{12}(2) & \cdots & \gamma_{12}(n-1) \\ \gamma_{12}(-1) & \gamma_{12}(0) & \gamma_{12}(1) & \cdots & \gamma_{12}(n-2) \\ \gamma_{12}(-2) & \gamma_{12}(-1) & \gamma_{12}(0) & \cdots & \gamma_{12}(n-3) \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_{12}(-n+1) & \gamma_{12}(-n+2) & \gamma_{12}(-n+3) & \cdots & \gamma_{12}(0) \end{pmatrix}. \quad (7.4.4)$$

In Theorem 4.2.1, for time series with absolutely summable covariance function, we demonstrated that the elements of  $\mathbf{G}\mathbf{V}_{ii}\mathbf{G}^*$  converge to the elements of the diagonal matrix  $2\pi\mathbf{D}_{ii}$ , where the elements of  $\mathbf{D}_{ii}$  are  $f_{ii}(\omega)$  evaluated at  $\omega_k = 2\pi k/n$ ,  $k = 0, 1, 2, \dots, n - 1$ .

It remains to investigate the behavior of  $\mathbf{G}\mathbf{V}_{12}\mathbf{G}^*$ . To this end define the circular matrix

$$\mathbf{V}_{12c} = \begin{pmatrix} \gamma_{12}(0) & \gamma_{12}(1) & \gamma_{12}(2) & \cdots & \gamma_{12}(-1) \\ \gamma_{12}(-1) & \gamma_{12}(0) & \gamma_{12}(1) & \cdots & \gamma_{12}(-2) \\ \gamma_{12}(-2) & \gamma_{12}(-1) & \gamma_{12}(0) & \cdots & \gamma_{12}(-3) \\ \vdots & \vdots & \vdots & & \vdots \\ \gamma_{12}(1) & \gamma_{12}(2) & \gamma_{12}(3) & \cdots & \gamma_{12}(0) \end{pmatrix}. \quad (7.4.5)$$

Then  $\mathbf{G}\mathbf{V}_{12c}\mathbf{G}^*$  is a diagonal matrix with elements

$$\begin{aligned} \sum_{h=0}^M \gamma_{12}(h) e^{-\epsilon 2\pi kh/n} + \sum_{h=-M}^{-1} \gamma_{12}(h) e^{-\epsilon 2\pi k(n+h)/n} \\ = \sum_{h=-M}^M \gamma_{12}(h) e^{-\epsilon 2\pi kh/n}, \quad k = 0, 1, \dots, n-1, \end{aligned} \quad (7.4.6)$$

where we have assumed  $n$  is odd and set  $M = (n-1)/2$ . If  $n$  is even, the sum is from  $-M+1$  to  $M$ , where  $M = n/2$ . If  $\gamma_{12}(h)$  is absolutely summable, we obtain the following theorem.

**Theorem 7.4.1.** Let  $\mathbf{Y}_t$  be a stationary bivariate time series with absolutely summable autocovariance function. Let  $\mathbf{V}$  of (7.4.3) be the covariance matrix for  $n$  observations. Then, given  $\epsilon > 0$ , there exists an  $N$  such that for  $n > N$ , every element of the matrix

$$\mathbf{H}\mathbf{V}\mathbf{H}^* - 2\pi\mathbf{D}$$

is less than  $\epsilon$  in magnitude, where

$$\mathbf{D} = \begin{pmatrix} \mathbf{D}_{11} & \mathbf{D}_{12} \\ \mathbf{D}_{21} & \mathbf{D}_{22} \end{pmatrix},$$

$$\mathbf{D}_{ij} = \text{diag}\{f_{ij}(\omega_0), f_{ij}(\omega_1), \dots, f_{ij}(\omega_{n-1})\}, \quad i, j = 1, 2,$$

and  $\omega_k = 2\pi k/n$ ,  $k = 0, 1, \dots, n-1$ .

**Proof.** The result for  $\mathbf{D}_{11}$  and  $\mathbf{D}_{22}$  follows by Theorem 4.2.1. The result for  $\mathbf{D}_{12}$  is obtained by arguments completely analogous to those of Section 4.2, by showing that the elements of  $\mathbf{G}\mathbf{V}_{12}\mathbf{G}^* - \mathbf{G}\mathbf{V}_{12c}\mathbf{G}^*$  converge to zero as  $n$  increases. The details are reserved for the reader.  $\blacktriangle$

If we make a stronger assumption about the autocovariance function we obtain the stronger result parallel to Corollary 4.2.1.

**Corollary 7.4.1.1.** Let  $\mathbf{Y}_t$  be a stationary bivariate time series with an autocovariance function that satisfies

$$\sum_{h=-\infty}^{\infty} |h| |\gamma_{ij}(h)| < L < \infty, \quad i, j = 1, 2.$$

Let  $\mathbf{V}$  be as defined in (7.4.3),  $\mathbf{H}$  as defined in (7.4.2), and  $\mathbf{D}$  as defined in Theorem 7.4.1. Then every element of the matrix  $\mathbf{HVH}^* - 2\pi\mathbf{D}$  is less than  $3L/n$ .

**Proof.** The proof parallels that of Corollary 4.2.1 and is reserved for the reader.  $\blacktriangle$

By Theorem 7.4.1, the complex coefficients  $2^{-1}n^{1/2}(a_{1k} + ib_{1k})$  and  $2^{-1}n^{1/2}(a_{2j} + ib_{2j})$  are nearly uncorrelated in large samples if  $j \neq k$ . Since

$$a_{ik} - ib_{ik} = a_{i,n-k} + ib_{i,n-k}, \quad i = 1, 2, \quad k = 1, 2, \dots, n-1,$$

it follows that

$$\begin{aligned} \lim_{n \rightarrow \infty} E\{a_{1k}a_{2k} - b_{1k}b_{2k}\} &= 0, \\ \lim_{n \rightarrow \infty} E\{b_{1k}a_{2k} + b_{2k}a_{1k}\} &= 0. \end{aligned} \tag{7.4.7}$$

That is, the covariance between  $a_{1k}$  and  $a_{2k}$  is approximately equal to that between  $b_{1k}$  and  $b_{2k}$ , while the covariance between  $b_{1k}$  and  $a_{2k}$  is approximately the negative of the covariance between  $b_{2k}$  and  $a_{1k}$ .

We define the cross periodogram by

$$I_{12n}(\omega_k) = \begin{cases} \frac{n}{2} [a_{1k}a_{2k} + b_{1k}b_{2k} - i(a_{1k}b_{2k} - b_{1k}a_{2k})], & \omega_k \neq 0, \pi, \\ 2na_{1k}a_{2k}, & \omega_k = 0, \pi. \end{cases} \tag{7.4.8}$$

To obtain a function defined at all  $\omega$ , we recall the function

$$K(n, \omega) = k \quad \text{for} \quad \frac{\pi(2k-1)}{n} < \omega \leq \frac{\pi(2k+1)}{n} \tag{7.4.9}$$

and take  $I_{12n}(\omega) = I_{12n}(\omega_{K(n, \omega)})$ .

**Corollary 7.4.1.2.** Let  $\mathbf{Y}_t$  be a stationary bivariate time series with absolutely summable covariance function. Then

$$\lim_{n \rightarrow \infty} E\{I_{12n}(\omega)\} = 4\pi f_{12}(\omega).$$

**Proof.** The result is an immediate consequence of Theorem 7.4.1.  $\blacktriangle$

We now obtain some distributional results for spectral estimates. To simplify the presentation, we assume that  $\mathbf{Y}_t$  is a normal time series.

**Theorem 7.4.2.** Let  $\mathbf{Y}_t$  be a bivariate normal time series with covariance function that satisfies

$$\sum_{h=-\infty}^{\infty} |h| |\gamma_{ij}(h)| < L < \infty, \quad i, j = 1, 2.$$

Then  $\mathbf{r}_k = 2^{-1/2} n^{1/2} (a_{1k}, b_{1k}, a_{2k}, b_{2k})'$  is distributed as a multivariate normal random variable with zero mean and covariance matrix

$$E\{\mathbf{r}_k \mathbf{r}'_k\} = 2\pi \begin{pmatrix} f_{11}(\omega_k) & 0 & c_{12}(\omega_k) & q_{12}(\omega_k) \\ 0 & f_{11}(\omega_k) & -q_{12}(\omega_k) & c_{12}(\omega_k) \\ c_{12}(\omega_k) & -q_{12}(\omega_k) & f_{22}(\omega_k) & 0 \\ q_{12}(\omega_k) & c_{12}(\omega_k) & 0 & f_{22}(\omega_k) \end{pmatrix} + O(n^{-1})$$

for  $\omega_k \neq 0, \pi$ , where  $f_{12}(\omega_k) = c_{12}(\omega_k) - iq_{12}(\omega_k)$ . Also

$$E\{\mathbf{r}_k \mathbf{r}'_j\} = O(n^{-1}), \quad j \neq k.$$

It follows that

$$E\left\{\frac{n}{2} (a_{1k} a_{2k} + b_{1k} b_{2k})\right\} = 4\pi c_{12}(\omega_k) + O(n^{-1}),$$

$$E\left\{\frac{n}{2} (a_{1k} b_{2k} - a_{2k} b_{1k})\right\} = 4\pi q_{12}(\omega_k) + O(n^{-1}),$$

$$E\{I_{12n}(\omega_k)\} = 4\pi f_{12}(\omega_k) + O(n^{-1}),$$

and for  $\omega_k \neq 0, \pi$ ,

$$\text{Cov}\{I_{12n}(\omega_k), I_{12n}(\omega_j)\} = \begin{cases} (4\pi)^2 f_{11}(\omega_k) f_{22}(\omega_k) + O(n^{-1}), & j = k, \\ O(n^{-2}), & j \neq k. \end{cases}$$

**Proof.** Since the  $a_{ik}$  and  $b_{ik}$  are linear combinations of normal random variables, the normality is immediate. The moment properties follow from the moments of the normal distribution and from Corollary 7.4.1.1.  $\blacktriangle$

We may construct smoothed estimators of the cross spectral density in the same manner that we constructed smoothed estimators in Section 7.2. Let

$$\begin{aligned} \bar{f}_{12}(\omega_k) &= \sum_{j=-d}^d W_n(j) \hat{f}_{12}(\omega_{k+j}) \\ &= \frac{1}{4\pi} \bar{I}_{12}(\omega_k), \end{aligned} \tag{7.4.10}$$

where

$$\begin{aligned}\hat{f}_{12}(\omega_k) &= \frac{1}{4\pi} I_{12n}(\omega_k), \\ \bar{I}_{12}(\omega_k) &= \sum_{j=-d}^d W_n(j) I_{12n}(\omega_{k+j}),\end{aligned}$$

and  $W_n(j)$ ,  $j = 0, \pm 1, \pm 2, \dots, \pm d$ , is a weight function.

**Theorem 7.4.3.** Let  $\mathbf{Y}_t$  be a bivariate normal time series with covariance function that satisfies  $\sum_{h=-\infty}^{\infty} |h| |\gamma_{ij}(h)| < L < \infty$ ,  $i, j = 1, 2$ . Let  $d_n$  be a sequence of positive integers satisfying

$$\lim_{n \rightarrow \infty} d_n = \infty,$$

$$\lim_{n \rightarrow \infty} \frac{d_n}{n} = 0,$$

and let  $W_n(j)$ ,  $j = 0, \pm 1, \pm 2, \dots, \pm d_n$ , satisfy

$$\sum_{j=-d_n}^{d_n} W_n(j) = 1,$$

$$W_n(j) = W_n(-j),$$

$$\lim_{n \rightarrow \infty} \sum_{j=-d_n}^{d_n} W_n^2(j) = 0.$$

Then

$$\lim_{n \rightarrow \infty} E\{\bar{f}_{12}(\omega_{K(n,\omega)})\} = f_{12}(\omega),$$

$$\lim_{n \rightarrow \infty} \left( \sum_{j=-d_n}^{d_n} W_n^2(j) \right)^{-1} \text{Var}\{\bar{f}_{12}(\omega_{K(n,\omega)})\} = \begin{cases} f_{11}(\omega)f_{22}(\omega), & \omega \neq 0, \pi, \\ 2f_{11}(\omega)f_{22}(\omega), & \omega = 0, \pi, \end{cases}$$

where  $\bar{f}_{12}(\omega_{K(n,\omega)})$  is defined by (7.4.10) and  $K(n, \omega)$  is defined in (7.4.9).

**Proof.** Reserved for the reader. ▲

The properties of the other multivariate spectral estimators follow from Theorem 7.4.2 and Theorem 7.4.3. Table 7.4.1 will be used to illustrate the computations and emphasize the relationships to normal regression theory. Notice that the number of entries in a column of Table 7.4.1 is  $2(2d + 1)$ . We set

**Table 7.4.1.** Statistics Used in Computation of Cross Spectra at  $\omega_k = 2\pi k/n$ ,  $d = 2$ 

Fourier Coefficients for $Y_2$	Fourier Coefficients for $Y_1$	Signed Fourier Coefficients for $Y_1$
$a_{2,k-2}$	$a_{1,k-2}$	$-b_{1,k-2}$
$b_{2,k-2}$	$b_{1,k-2}$	$a_{1,k-2}$
$a_{2,k-1}$	$a_{1,k-1}$	$-b_{1,k-1}$
$b_{2,k-1}$	$b_{1,k-1}$	$a_{1,k-1}$
$a_{2k}$	$a_{1k}$	$-b_{1k}$
$b_{2k}$	$b_{1k}$	$a_{1k}$
$a_{2,k+1}$	$a_{1,k+1}$	$-b_{1,k+1}$
$b_{2,k+1}$	$b_{1,k+1}$	$a_{1,k+1}$
$a_{2,k+2}$	$a_{1,k+2}$	$-b_{1,k+2}$
$b_{2,k+2}$	$b_{1,k+2}$	$a_{1,k+2}$

$W_n(j) = (2d + 1)^{-1}$  for  $j = 0, \pm 1, \pm 2, \dots, \pm d$ . Then the cospectrum is estimated by

$$\bar{c}_{12}(\omega_k) = \frac{n}{8\pi(2d+1)} \sum_{j=-d}^d (a_{1,k-j}a_{2,k-j} + b_{1,k-j}b_{2,k-j}), \quad (7.4.11)$$

which is the mean of the cross products of the first two columns of Table 7.4.1 multiplied by  $n/4\pi$ . The quadrature spectrum is estimated by

$$\bar{q}_{12}(\omega_k) = \frac{n}{8\pi(2d+1)} \sum_{j=-d}^d (a_{1,k-j}b_{2,k-j} - a_{2,k-j}b_{1,k-j}), \quad (7.4.12)$$

which is the mean of the cross products of the first and third column of Table 7.4.1 multiplied by  $n/4\pi$ .

The estimator of the squared coherency for a bivariate time series computed from the smoothed periodogram estimator of  $\mathbf{f}(\omega)$ ,

$$\hat{\mathcal{K}}(\omega_k) = \frac{1}{2d+1} \sum_{j=-d}^d \begin{pmatrix} \hat{f}_{11}(\omega_{k-j}) & \hat{f}_{12}(\omega_{k-j}) \\ \hat{f}_{21}(\omega_{k-j}) & \hat{f}_{22}(\omega_{k-j}) \end{pmatrix},$$

is given by

$$\bar{\mathcal{K}}_{12}^2(\omega_k) = \frac{|\bar{f}_{12}(\omega_k)|^2}{\bar{f}_{11}(\omega_k)\bar{f}_{22}(\omega_k)} = \frac{[\bar{c}_{12}(\omega_k)]^2 + [\bar{q}_{12}(\omega_k)]^2}{\bar{f}_{11}(\omega_k)\bar{f}_{22}(\omega_k)}. \quad (7.4.13)$$

This quantity is recognizable as the multiple correlation coefficient of normal regression theory obtained by regressing the first column of Table 7.4.1 on the second and third columns. By construction, the second and third columns are orthogonal.

The estimation of the squared coherency generalizes immediately to higher

dimensions. If there is a second explanatory variable, the Fourier coefficients of this variable are added to Table 7.4.1 in the same form as the columns for  $Y_1$ . Then the multiple squared coherency is the multiple correlation coefficient associated with the regression of the column for  $Y_2$  on the four columns for the two explanatory variables.

An estimator of the error spectrum or residual spectrum of  $Y_2$  after  $Y_1$  is

$$\bar{f}_{zz}(\omega_k) = \bar{f}_{22}(\omega_k)[1 - \bar{\mathcal{K}}_{12}^2(\omega_k)] \frac{2d+1}{2d}. \quad (7.4.14)$$

This is the residual mean square for the regression of the first column of Table 7.4.1 on the second and third multiplied by  $n/4\pi$ . Many authors define the estimator of the error spectrum without the factor  $(2d+1)/2d$ . We include it to make the analogy to multiple regression complete. It also serves to remind us that  $\bar{\mathcal{K}}_{12}^2(\omega_k)$  is identically one if computed for  $d=0$ . A test of the hypothesis that  $\bar{\mathcal{K}}_{12}^2(\omega_k) = 0$  is given by the statistic

$$F_{4d}^2 = \frac{4d\bar{\mathcal{K}}_{12}^2(\omega_k)}{2[1 - \bar{\mathcal{K}}_{12}^2(\omega_k)]}, \quad (7.4.15)$$

which is approximately distributed as Snedecor's  $F$  with 2 and  $4d$  ( $d > 0$ ) degrees of freedom under the null hypothesis. This is the test of the hypothesis that the regression coefficients associated with columns two and three of Table 7.4.1 are zero.

If  $\bar{\mathcal{K}}_{12}^2(\omega_k) \neq 0$ , the distribution of  $\bar{\mathcal{K}}_{12}^2(\omega_k)$  is approximately that of the multiple correlation coefficient [see, for example, Anderson (1984, p. 134). Tables and graphs useful in constructing confidence intervals for  $\bar{\mathcal{K}}_{12}^2(\omega_k)$  have been given by Amos and Koopmans (1963)]. For many degrees of freedom and  $\bar{\mathcal{K}}_{12}^2(\omega_k) \neq 0$ ,  $\bar{\mathcal{K}}_{12}^2(\omega_k)$  is approximately normally distributed with variance

$$\text{Var}[\bar{\mathcal{K}}_{12}^2(\omega_k)] = \frac{4\bar{\mathcal{K}}_{12}^2(\omega_k)[1 - \bar{\mathcal{K}}_{12}^2(\omega_k)]^2}{4d+2}. \quad (7.4.16)$$

The estimated phase spectrum is

$$\bar{\varphi}_{12}(\omega_k) = \tan^{-1}[-\bar{q}_{12}(\omega_k)/\bar{c}_{12}(\omega_k)], \quad (7.4.17)$$

where it is understood that  $\bar{\varphi}_{12}(\omega_k)$  is the angle in  $(-\pi, \pi]$  between the positive half of the  $\bar{c}_{12}(\omega_k)$  axis and the ray from the origin through  $(\bar{c}_{12}(\omega_k), -\bar{q}_{12}(\omega_k))$ . The sample distribution of this quantity depends in a critical manner on the true coherency between the two time series. If  $\bar{\mathcal{K}}_{12}^2(\omega_k) = 0$ , then, conditional on  $\bar{f}_{11}(\omega_k)$ , the variables  $\bar{c}_{12}(\omega_k)/\bar{f}_{11}(\omega_k)$  and  $\bar{q}_{12}(\omega_k)/\bar{f}_{11}(\omega_k)$  are approximately distributed as independent normal  $(0, f_{zz}(\omega_k)[2(2d+1)\bar{f}_{11}(\omega_k)]^{-1})$  random variables. This is because  $\bar{c}_{12}(\omega_k)/\bar{f}_{11}(\omega_k)$  and  $\bar{q}_{12}(\omega_k)/\bar{f}_{11}(\omega_k)$  are the regression coefficients obtained by regressing column one of Table 7.4.1 on columns two and three. It is well known that the ratio of two independent normal random variables with zero mean and common variance has the Cauchy distribution, and that the arc

tangent of the ratio has a uniform distribution. Therefore, if  $\mathcal{K}_{12}^2(\omega_k) = 0$ , the principal value of  $\bar{\varphi}_{12}(\omega_k)$  will be approximately uniformly distributed on the interval  $(-\pi/2, \pi/2)$ .

If  $\mathcal{K}_{12}^2(\omega_k) \neq 0$ , then  $\bar{\varphi}_{12}(\omega)$  will converge in distribution to a normal random variable. While approximate confidence limits could be established on the basis of the limiting distribution, it seems preferable to set confidence limits using the normality of  $\bar{c}_{12}(\omega_k)/\bar{f}_{11}(\omega_k)$  and  $\bar{q}_{12}(\omega_k)/\bar{f}_{11}(\omega_k)$ .

Fieller's method [see Fieller (1954)] can be used to construct a confidence interval for  $\varphi_{12}(\omega)$ . This procedure follows from the fact that the statement  $-q_{12}(\omega)/c_{12}(\omega) = R_{12}(\omega)$  is equivalent to the statement  $c_{12}(\omega)R_{12}(\omega) + q_{12}(\omega) = 0$ . Therefore, the method of setting confidence intervals for the sum  $c_{12}(\omega)R_{12}(\omega) + q_{12}(\omega)$  can be used to determine a confidence interval for  $R_{12}(\omega) = \tan \varphi_{12}(\omega)$  and hence for  $\varphi_{12}(\omega)$ . The  $(1 - \alpha)$ -level confidence interval for the principal value of  $\varphi_{12}(\omega)$  is the set of  $\varphi_{12}(\omega)$  in  $[-\pi/2, \pi/2]$  such that

$$\sin^2[\varphi_{12}(\omega) - \bar{\varphi}_{12}(\omega)] \leq t_\alpha^2 [\bar{c}_{12}^2(\omega) + \bar{q}_{12}^2(\omega)]^{-1} \hat{\text{Var}}\{\bar{c}_{12}(\omega)\}, \quad (7.4.18)$$

where

$$\hat{\text{Var}}\{\bar{c}_{12}(\omega)\} = (4d + 2)^{-1} \bar{f}_{11}(\omega) \bar{f}_{zz}(\omega),$$

$t_\alpha$  is such that  $P\{|t| > t_\alpha\} = \alpha$ , and  $t$  is distributed as Student's  $t$  with  $4d$  degrees of freedom.

To obtain a confidence interval for  $\varphi_{12}(\omega)$  in the interval  $(-\pi, \pi]$  it is necessary to modify Fieller's method. We suggest the following procedure to establish an approximate  $(1 - \alpha)$ -level confidence interval. Let  $F_{4d}^2(\alpha)$  denote the  $\alpha\%$  point of the  $F$ -distribution with 2 and  $4d$  degrees of freedom. The possibilities for the interval fall into two categories.

1.  $\bar{c}_{12}^2(\omega) + \bar{q}_{12}^2(\omega) \leq 2F_{4d}^2(\alpha) \hat{\text{Var}}\{\bar{c}_{12}(\omega)\}$ . The confidence interval for  $\varphi_{12}(\omega)$  is  $(-\pi, \pi]$ .
2.  $\bar{c}_{12}^2(\omega) + \bar{q}_{12}^2(\omega) > 2F_{4d}^2(\alpha) \hat{\text{Var}}\{\bar{c}_{12}(\omega)\}$ . The confidence interval for  $\varphi_{12}(\omega)$  is  $[\bar{\varphi}_{12}(\omega) - \delta, \bar{\varphi}_{12}(\omega) + \delta]$ , where

$$\begin{aligned} \delta &= \sin^{-1}[t_\alpha^2 \{\bar{c}_{12}^2(\omega) + \bar{q}_{12}^2(\omega)\}^{-1} \hat{\text{Var}}\{\bar{c}_{12}(\omega)\}]^{1/2} \\ &= \sin^{-1}\left[t_\alpha^2 \frac{1 - \bar{\mathcal{K}}_{12}^2(\omega)}{4d \bar{\mathcal{K}}_{12}^2(\omega)}\right]^{1/2}. \end{aligned}$$

Note that the criterion for category 1 is satisfied when the  $F$ -statistic of (7.4.15) is less than  $F_{4d}^2(\alpha)$ . Assuming  $\bar{c}_{12}(\omega)$  and  $\bar{q}_{12}(\omega)$  to be normally distributed, it can be proven that the outlined procedure furnishes a confidence interval with probability at least  $1 - \alpha$  of covering the true  $\varphi(\omega)$ . If the true coherency is zero, the interval will have length  $2\pi$  with probability  $1 - \alpha$ .

Recall that the cross amplitude spectrum is

$$A_{12}(\omega) = [c_{12}^2(\omega) + q_{12}^2(\omega)]^{1/2} = |f_{12}(\omega)|$$

and the gain of  $X_{2t}$  over  $X_{1t}$  is

$$\psi_{12}(\omega) = \frac{A_{12}(\omega)}{f_{11}(\omega)}.$$

Estimator of these quantities are

$$\bar{A}_{12}(\omega) = [\bar{c}_{12}^2(\omega) + \bar{q}_{12}^2(\omega)]^{1/2}, \quad (7.4.19)$$

$$\bar{\psi}_{12}(\omega) = [\bar{f}_{11}(\omega)]^{-1} \bar{A}_{12}(\omega). \quad (7.4.20)$$

It is possible to establish approximate confidence intervals for these quantities using the approximate normality of  $\bar{c}_{12}(\omega)/\bar{f}_{11}(\omega)$  and  $\bar{q}_{12}(\omega)/\bar{f}_{11}(\omega)$ . As a consequence of this normality,

$$\frac{(2d+1)\{[\bar{c}_{12}(\omega) - c_{12}(\omega)]^2 + [\bar{q}_{12}(\omega) - q_{12}(\omega)]^2\}}{\bar{f}_{11}(\omega)\bar{f}_{zz}(\omega)} \quad (7.4.21)$$

has, approximately, the  $F$ -distribution with 2 and  $4d$  degrees of freedom. Therefore, those  $c_{12}(\omega)$  and  $q_{12}(\omega)$  for which (7.4.21) is less than the  $\alpha$  percentage tabular value of the  $F$ -distribution form a  $(1 - \alpha)$ -level confidence region. Let

$$\begin{aligned} A_U(\omega) &= \bar{A}_{12}(\omega) + [(2d+1)^{-1}\bar{f}_{11}(\omega)\bar{f}_{zz}(\omega)F_{4d}^2(\alpha)]^{1/2}, \\ A_L(\omega) &= \max\{0, \bar{A}_{12}(\omega) - [(2d+1)^{-1}\bar{f}_{11}(\omega)\bar{f}_{zz}(\omega)F_{4d}^2(\alpha)]^{1/2}\}. \end{aligned} \quad (7.4.22)$$

Assuming normal  $\bar{c}_{12}(\omega)$  and  $\bar{q}_{12}(\omega)$ , we conclude that  $[A_L(\omega), A_U(\omega)]$  is a confidence interval for  $A_{12}(\omega)$  of at least level  $1 - \alpha$ . The confidence interval for gain is that of  $A_{12}(\omega)$  divided by  $\bar{f}_{11}(\omega)$ .

**Example 7.4.1.** We use the data on the sediment in the Des Moines River discussed in Section 6.4 to illustrate some of the cross spectral computations. Table 7.4.2 contains the Fourier coefficients for the first 11 frequencies for the 205 observations. For  $d = 5$ , these are the statistics used to estimate  $f(\omega_6)$ , where  $\omega_6 = 0.0293$  cycles per day.

Using the rectangular weight function, we have

$$\bar{I}_{22, 205}(0.0293) = \frac{1}{11} \sum_{j=-5}^5 \frac{205}{2} (a_{2,6+j}^2 + b_{2,6+j}^2) = 4.0320,$$

$$\bar{I}_{11, 205}(0.0293) = \frac{1}{11} \sum_{j=-5}^5 \frac{205}{2} (a_{1,6+j}^2 + b_{1,6+j}^2) = 7.2999,$$

$$\bar{I}_{12, 205}(0.0293) = \frac{1}{11} \sum_{j=-5}^5 \frac{205}{2} (a_{1,6+j}a_{2,6+j} + b_{1,6+j}b_{2,6+j})$$

**Table 7.4.2. Statistics Used in Computing Smoothed Estimates of Cross Spectrum for Sediment in the Des Moines River at Boone and Saylorville for Frequency of 0.0293 Cycles per Day,  $d = 5$**

Index ( $k$ )	Frequency (Cycles per Day)	Period (Days)	Coefficients		
			Saylorville ( $a_{2k}, b_{2k}$ )	Boone ( $a_{1k}, b_{1k}$ )	Boone ( $-b_{1k}, a_{1k}$ )
1	0.0049	205.00	-0.194 0.305	-0.103 0.504	-0.504 -0.103
2	0.0098	102.50	0.175 0.161	0.432 0.196	-0.196 0.432
3	0.0146	68.33	0.011 0.013	0.047 -0.039	0.039 0.047
4	0.0195	51.25	0.058 0.147	0.055 0.088	-0.088 0.055
5	0.0244	41.00	0.067 0.042	0.202 0.116	-0.116 0.202
6	0.0293	34.17	0.021 0.061	0.133 0.072	-0.072 0.133
7	0.0341	29.29	0.322 0.059	0.234 -0.012	0.012 0.234
8	0.0390	25.63	0.065 -0.067	-0.017 -0.019	0.019 -0.017
9	0.0439	22.78	-0.053 0.019	-0.013 0.098	-0.098 -0.013
10	0.0488	20.50	0.281 0.037	0.332 -0.053	0.053 0.332
11	0.0537	18.64	0.081 -0.062	0.152 0.026	-0.026 0.152

$$= -\epsilon \cdot \frac{205}{22} \sum_{j=-5}^5 (b_{2,6+j}a_{1,6+j} - a_{2,6+j}b_{1,6+j}) \\ = 4.6768 - 1.4313\epsilon.$$

It follows that

$$\begin{aligned}\tilde{f}_{22}(0.0293) &= 0.3209, \\ \tilde{f}_{11}(0.0293) &= 0.5809, \\ \tilde{f}_{12}(0.0293) &= 0.3722 - 0.1138\epsilon.\end{aligned}$$

The error spectrum for Saylorville is estimated by the residual sum of squares obtained from the regression of the Saylorville column on the Boone columns

multiplied by  $205[2(10)(4\pi)]^{-1}$ . We have

$$\begin{aligned}\bar{f}_{zz}(0.0293) &= \frac{205}{80\pi} [0.4327 - 0.6407(0.5019) - (-0.1960)(-0.1536)] \\ &= 0.0661.\end{aligned}$$

The squared coherency is

$$\bar{\mathcal{H}}_{12}^2(0.0293) = \frac{|\bar{f}_{12}(0.0293)|^2}{\bar{f}_{11}(0.0293)\bar{f}_{22}(0.0293)} = 0.8128.$$

The  $F$ -statistic to test the hypothesis of zero coherency is

$$F_{20}^2 = \frac{20\bar{\mathcal{H}}_{12}^2(0.0293)}{2[1 - \bar{\mathcal{H}}_{12}^2(0.0293)]} = 43.42.$$

This is well beyond the 1% tabular value of 5.85 for 2 and 20 degrees of freedom, and it is clear that the two time series are not independent. The estimate of the phase spectrum is

$$\bar{\phi}_{12}(0.0293) = \tan^{-1}[-1.4313/4.6768] = -0.2970 \text{ radians}.$$

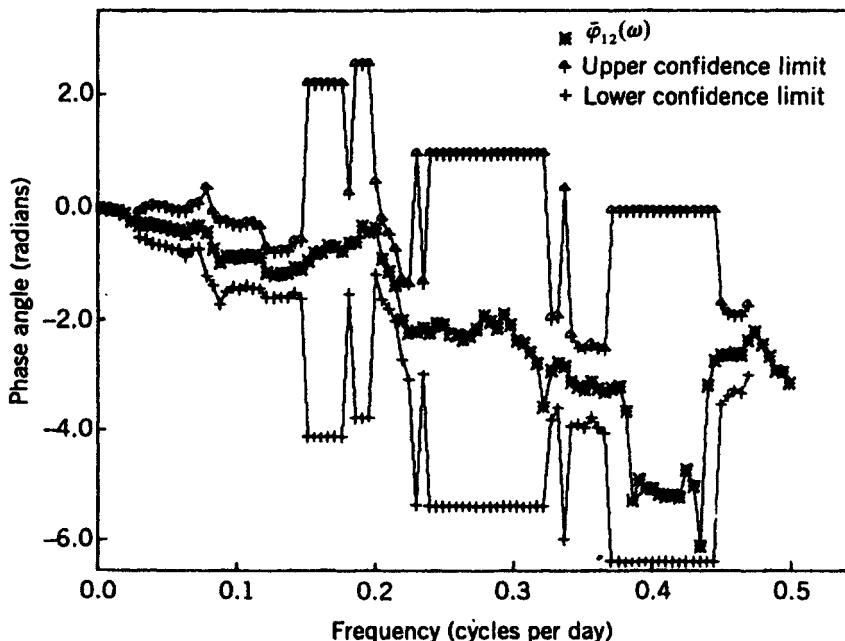


Figure 7.4.1. Estimated phase spectrum for Boone-Saylorville sediment data ( $d = 5$  and rectangular weight function).

Let us establish a 95% confidence interval for  $\varphi_{12}(0.0293)$ . Because the  $F$ -test rejects the hypothesis of zero coherency at the 5% level, the criterion of category 2 is satisfied. Consequently the confidence interval for  $\varphi_{12}(\omega_6)$  is  $[-0.5228, -0.0712]$ , where

$$\delta = \sin^{-1} \left[ (2.086)^2 \frac{0.1872}{20(0.8128)} \right]^{1/2} = 0.2258.$$

The estimated gain of  $X_{2t}$  over  $X_{1t}$  is

$$\bar{\psi}_{12}(\omega) = \frac{|\bar{f}_{12}(0.0293)|}{\bar{f}_{11}(0.0293)} = \frac{0.3892}{0.5809} = 0.6700.$$

A 95% confidence interval for  $\psi_{12}(\omega_6)$  is given by  $[\psi_L, \psi_U]$ , where

$$\psi_L = 0.6700 - [(11)^{-1}(0.5809)^{-1}(0.0661)(3.49)]^{1/2} = 0.4800,$$

$$\psi_U = 0.6700 + [(11)^{-1}(0.5809)^{-1}(0.0661)(3.49)]^{1/2} = 0.8600.$$

Figure 7.4.1 is a plot of  $\bar{\varphi}_{12}(\omega)$  and the confidence interval for  $\varphi_{12}(\omega)$  ( $d = 5$  and the rectangular weight function) for the Boone-Saylorville sediment data. Recall that  $q_{12}(\omega)$  is an odd function of  $\omega$  with  $q_{12}(0) = q_{12}(\pi) = 0$ . Therefore, we

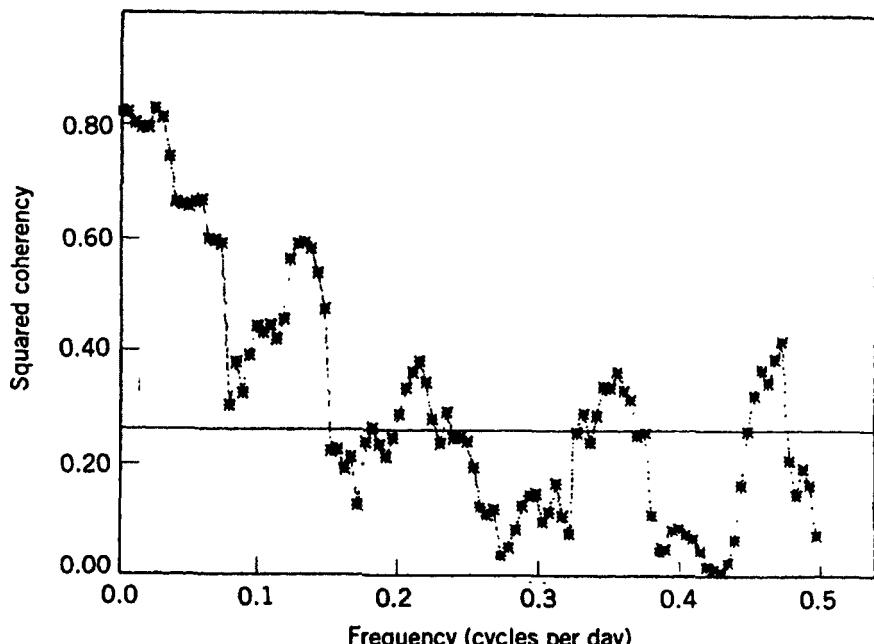


Figure 7.4.2. Squared coherency for the Boone-Saylorville sediment data ( $d = 5$  and rectangular weight function).

define the imaginary part of  $I_{12n}(\omega_{-k})$  to be the negative of the imaginary part of  $I_{12n}(\omega_k)$ . Likewise, the imaginary part of  $I_{12n}(\pi + \lambda)$  is set equal to the negative of the imaginary part of  $I_{12n}(\pi - \lambda)$ . If  $I_{12n}(0)$  or  $I_{12n}(\pi)$  are computed, the imaginary part is zero. As a result,  $\bar{q}_{12}(0) = 0$ , and  $\bar{q}_{12}(\pi) = 0$  if  $n$  is even. It follows that the estimated phase at zero is 0 if  $\bar{c}_{12}(0) > 0$  and is  $\pi$  if  $\bar{c}_{12}(0) < 0$ . In plotting the estimated phase a continuous function of  $\omega$  is desirable. Therefore, in creating Figure 7.4.1, that angle in the set

$$\{\bar{\varphi}_{12}(\omega_k) + j\pi: j = 0, \pm 2, \pm 4, \dots\}$$

that differed least from the angle previously chosen for  $\omega_{k-1}$ ,  $k = 2, 3, \dots, 102$ , was plotted.

The general downward slope of  $\bar{\varphi}_{12}(\omega)$  is associated with the fact that the readings at Saylorville lag behind those at Boone. If the relationship were a perfect one period lag,  $\bar{\varphi}_{12}(\omega)$  would be estimating a straight line with a negative slope of one radian per radian. The estimated function seems to differ enough from such a straight line to suggest a more complicated lag relationship.

Figure 7.4.2 contains a plot of squared coherency for the sediment data. The 5%

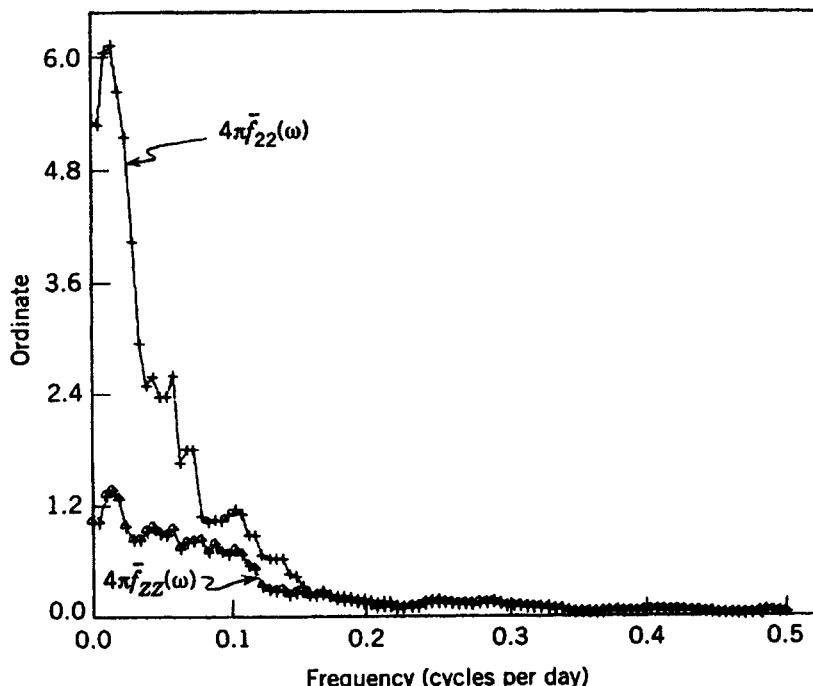


Figure 7.4.3. Plot of  $4\pi\bar{f}_{22}(\omega)$  and  $4\pi\bar{f}_{ZZ}(\omega)$  for Saylorville sediment ( $d = 5$  and rectangular weight function).

point for the  $F$ -distribution with 2 and 20 degrees of freedom is 3.49. On the basis of (7.4.15), any  $\bar{\mathcal{H}}_{12}^2(\omega)$  greater than 0.259 would be judged significant at that level. A line has been drawn at this height in the figure.

Similar information is contained in Figure 7.4.3, where the smoothed periodogram for Saylorville and  $4\pi\tilde{f}_{ZZ}(\omega)$  are plotted on the same graph. The estimated error spectrum lies well below the spectrum of the original time series for low frequencies, but the two nearly coincide for high frequencies. The estimated error spectrum is clearly not that of white noise, since it is considerably higher at low frequencies than at high. One might be led to consider a first or second order autoregressive process as a model for the error.

Figure 7.4.4 is a plot of the estimated gain of Saylorville over Boone. The 95% confidence interval plotted on the graph was computed using the limits (7.4.22) divided by  $\hat{f}_{11}(\omega)$ . Note that the lower confidence bound for gain is zero whenever the squared coherency falls below 0.259.

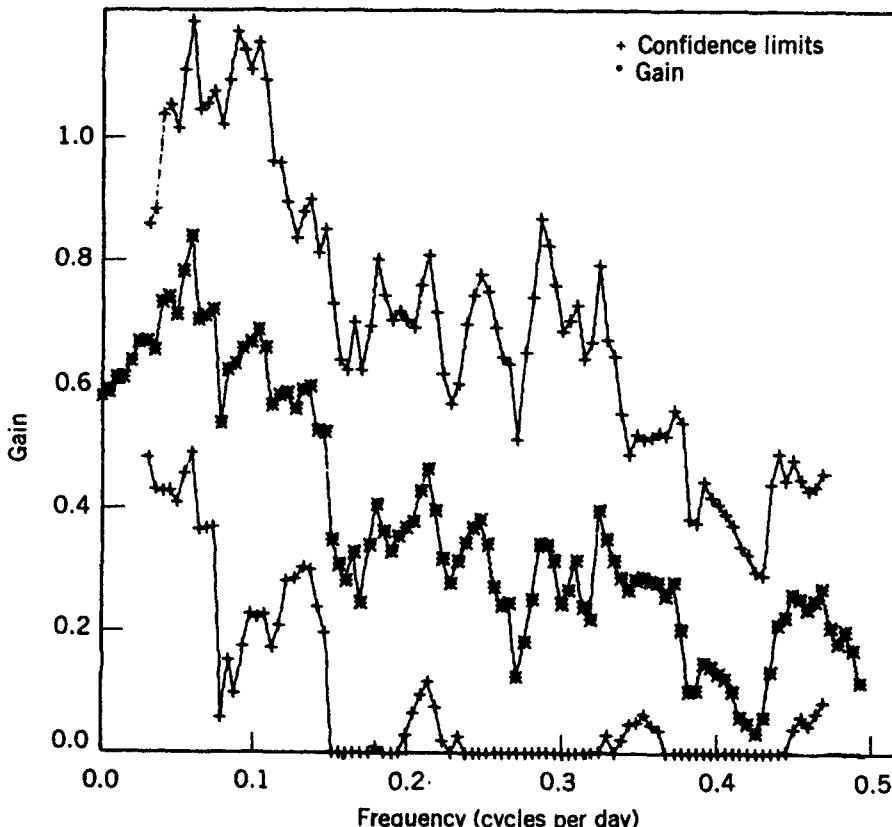


Figure 7.4.4. Gain of Saylorville over Boone ( $d = 5$  and rectangular weight function).

## REFERENCES

- Section 7.1.** Bartlett (1950, 1966), Birnbaum (1952), Davis (1941), Durbin (1967, 1968, 1969), Eisenhart, Hastay, and Wallis (1947), Fisher (1929), Grenander and Rosenblatt (1957), Hannan (1958, 1960), Harris (1967), Kendall and Stuart (1966), Koopmans (1974), Tintner (1952), Tukey (1961), Wilks (1962).
- Sections 7.2, 7.3.** Akaike (1969a), Anderson (1971), Bartlett (1950), Blackman and Tukey (1959), Bloomfield (1976), Box (1954), Brillinger (1975), Hannan (1960, 1970), Jenkins (1961), Jenkins and Watts (1968), Koopmans (1974), Marple (1987), Newton (1988), Otnes and Enochson (1972), Parzen (1957, 1961, 1974, 1977), Priestly (1981).
- Section 7.4.** Amos and Koopmans (1963), Anderson (1984), Brillinger (1975), Fieller (1954), Fishman (1969), Hannan (1970), Jenkins and Watts (1968), Koopmans (1974), Scheffé (1970), Wahba (1968).

## EXERCISES

1. Compute the periodogram for the data of Table 6.4.1. Calculate the smoothed periodogram using 2, 4 and 8 for  $d$  and rectangular weights. Plot the smoothed periodograms, and observe the differences in smoothness and in the height and width of the peak at zero. Compute a 95% confidence interval for  $4\pi f(\omega)$  using the smoothed periodogram with  $d = 4$ . Plot the logarithm of the smoothed periodogram and the confidence interval for  $\log 4\pi f(\omega)$ .
2. Given in the accompanying table is the quarterly gasoline consumption in California from 1960 to 1973 in millions of gallons.

Year	Quarter			
	I	II	III	IV
1960	1335	1443	1529	1447
1961	1363	1501	1576	1495
1962	1464	1450	1611	1612
1963	1516	1660	1738	1652
1964	1639	1754	1839	1736
1965	1699	1812	1901	1821
1966	1763	1937	2001	1894
1967	1829	1966	2068	1983
1968	1939	2099	2201	2081
1969	2008	2232	2299	2204
1970	2152	2313	2393	2278
1971	2191	2402	2450	2387
1972	2391	2549	2602	2529
1973	2454	2647	2689	2549

Source. U.S. Dept. of Transportation (1975), *Review and Analysis of Gasoline Consumption in the United States from 1960 to the Present*, and U.S. Dept. of Transportation, *News*, various issues.

Using these data:

- Compute the periodogram.
- Obtain the smoothed periodogram by computing the centered moving average

$$\bar{I}_n(\omega_k) = \frac{1}{5} \sum_{j=-2}^2 I_n(\omega_{k-j}) .$$

- Fit the regression model

$$Y_t = \alpha + \beta t + Z_t .$$

Repeat parts a and b for the regression residuals  $\hat{Z}_t$ .

- Compute the smoothed periodogram for  $\hat{Z}_t$  of part c using the symmetric weight function  $W_n(j)$  where

$$W_n(j) = \begin{cases} 0.3, & j = 0, \\ 0.2, & j = 1, \\ 0.1, & j = 2, \\ 0.05, & j = 3, \\ 0, & |j| \geq 4. \end{cases}$$

- Fit the regression model

$$Y_t = \sum_{j=1}^4 \alpha_j \delta_{ij} + \beta t + u_t ,$$

where

$$\delta_{ij} = \begin{cases} 1, & j\text{th quarter,} \\ 0 & \text{otherwise.} \end{cases}$$

Repeat parts a and b for the residuals from the fitted regression. Compute and plot a 95% confidence interval for the estimated spectral density.

3. Let the time series  $X_t$  be defined by

$$X_t = e_t + 0.6e_{t-1} ,$$

where  $\{e_t\}$  is a sequence of normal independent  $(0, 1)$  random variables. Given a sample of 10,000 observations from such a time series, what is the approximate joint distribution of the periodogram ordinates associated with  $\omega_{2500} = 2\pi(2500)/10,000$  and  $\omega_{1250} = 2\pi(1250)/10,000$ ?

4. Prove the sine result of Lemma 7.2.2.

5. Prove that the covariance function of a stationary finite autoregressive process

satisfies

$$\sum_{h=-n}^n |h| |\gamma(h)| = O(1).$$

6. Use the moment properties of the normal distribution to demonstrate the portion of Theorem 7.4.2 that states that

$$\text{Cov}\{I_{12n}(\omega_k), I_{12n}(\omega_j)\} = \begin{cases} (4\pi)^2 f_{11}(\omega_k) f_{22}(\omega_k) + O(n^{-1}), & j = k, \\ O(n^{-2}), & j \neq k. \end{cases}$$

7. Let  $X_{1t}$  denote the United States quarterly unemployment rate of Table 6.4.1, and let  $X_{2t}$  denote the weekly gross hours per production worker given in Exercise 13 of Chapter 6. Compute the periodogram quantities  $I_{11n}(\omega_k)$ ,  $I_{12n}(\omega_k)$ , and  $I_{22n}(\omega_k)$ . Compute the smoothed estimates using  $d = 5$  and the rectangular weight function. Compute and plot  $\hat{\chi}_{12}^2(\omega_k)$ ,  $\hat{\varphi}_{12}(\omega_k)$ . Obtain and plot a confidence interval for  $\varphi_{12}(\omega_k)$  and for the gain of  $X_{2t}$  over  $X_{1t}$ . Treating hours per week as the dependent variable, plot  $4\pi\hat{f}_{zz}(\omega)$ .

8. Show that

$$I_{XYn}(\omega_k) = 4\pi\hat{f}_{XY}(\omega_k), \quad k = 1, 2, \dots, m,$$

where

$$\hat{f}_{XY}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \hat{\gamma}_{XY}(h) e^{-i\omega h},$$

$$\hat{\gamma}_{XY}(h) = \begin{cases} \frac{1}{n} \sum_{t=1}^{n-h} (X_t - \bar{x}_n)(Y_{t+h} - \bar{y}_n), & h = 0, 1, \dots, n-1, \\ \frac{1}{n} \sum_{t=-h+1}^n (X_t - \bar{x}_n)(Y_{t+h} - \bar{y}_n), & h = -1, -2, \dots, -(n-1), \\ 0 & \text{otherwise.} \end{cases}$$

9. Let  $X_t$  be a time series defined by

$$X_t = \sum_{j=0}^{\infty} \alpha_j e_{t-j},$$

where  $\{e_t\}$  is a sequence of independent  $(0, \sigma^2)$  random variables with fourth moment  $\eta\sigma^4$  and

$$\sum_{j=0}^{\infty} j|\alpha_j| < \infty.$$

(a) Show that

$$\sum_{h=-\infty}^{\infty} |h| |\gamma(h)| < \infty$$

for such a time series.

(b) Let  $d_n$ ,  $W_n(j)$ , and  $\tilde{f}(\omega)$  be as defined in Theorem 7.2.2. Show that

$$E\{\tilde{f}(\omega)\} = f(\omega) + O(n^{-1}d_n).$$

10. Given  $X_1, X_2, \dots, X_n$ , let  $\hat{\gamma}_X(h)$  be defined by (7.1.10). Consider the augmented observations

$$Y_t = \begin{cases} X_t - \bar{x}_n, & 1 \leq t \leq n, \\ 0, & n+1 \leq t \leq 2n-1. \end{cases}$$

(a) Show that the periodogram of  $Y_t$  can be written

$$I_{Y,2n-1}(\omega_j) = (2n-1)^{-1} 2n \sum_{h=-n+1}^{n-1} \hat{\gamma}_X(h) e^{i\omega_j h},$$

where  $\omega_j = (2n-1)^{-1} 2\pi j$ .

(b) Show that

$$\hat{\gamma}_X(j) = (2n)^{-1} (2n-1)^{1/2} C_j,$$

where  $\nu_j = (2n-1)^{-1} 2\pi j$  and

$$C_j = (2n-1)^{-1/2} \sum_{r=0}^{2n-2} I_{Y,2n-1}(\omega_r) e^{-i\nu_j r}.$$

11. Suppose  $X_1, X_2, \dots, X_{p+1}$  are uncorrelated random variables. Show that

$$\text{Var}\left\{(p+1)^{-1} \sum_{i=1}^{p+1} X_i\right\} \leq \text{Var}\left\{p^{-1} \sum_{i=1}^p X_i\right\}$$

unless

$$\text{Var}\{X_{p+1}\} > (2p+1) \text{Var}\left\{p^{-1} \sum_{i=1}^p X_i\right\}.$$

## CHAPTER 8

## Parameter Estimation

## 8.1. FIRST ORDER AUTOREGRESSIVE TIME SERIES

The stationary time series defined by

$$Y_t - \mu = \rho(Y_{t-1} - \mu) + e_t, \quad (8.1.1)$$

where the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables and  $|\rho| < 1$ , is one of the simplest and most heavily used models in time series analysis. It is often a satisfactory representation of the error time series in economic models. This model also underlies many tests of the hypothesis that the observed time series is a sequence of independently and identically distributed random variables.

One estimator of  $\rho$  is the first order autocorrelation coefficient discussed in Chapter 6,

$$\hat{\chi}(1) = \frac{\hat{\chi}(1)}{\hat{\chi}(0)} = \frac{\sum_{t=1}^{n-1} (Y_t - \bar{y}_n)(Y_{t+1} - \bar{y}_n)}{\sum_{t=1}^n (Y_t - \bar{y}_n)^2}, \quad (8.1.2)$$

where  $\bar{y}_n = n^{-1} \sum_{t=1}^n Y_t$ . To introduce some other estimators of the parameter  $\rho$ , let us consider the distribution of the  $Y_t$  for the normal time series defined by (8.1.1).

The expected value of  $Y_t$  is  $\mu$ , and the expected value of  $Y_t - \rho Y_{t-1}$  is  $\lambda = (1 - \rho)\mu$ . For a sample of  $n$  observations we can write

$$\begin{aligned} Y_1 &= \mu + \nu_1, \\ Y_t - \mu &= \rho(Y_{t-1} - \mu) + e_t, \quad t = 2, 3, \dots, n, \end{aligned} \quad (8.1.3)$$

or

$$\begin{aligned} Y_1 &= \mu + \nu_1 \\ Y_t &= \lambda + \rho Y_{t-1} + e_t, \quad t = 2, 3, \dots, n, \end{aligned}$$

where the vector  $(\nu_1, e_2, e_3, \dots, e_n)$  is distributed as a multivariate normal with zero mean and covariance matrix

$$\Sigma = \text{diag}\{(1 - \rho^2)^{-1} \sigma^2, \sigma^2, \sigma^2, \dots, \sigma^2\}. \quad (8.1.4)$$

It follows that twice the logarithm of the likelihood of a sample of  $n$  observations is

$$\begin{aligned} 2 \log L(y: \mu, \rho, \sigma^2) &= -n \log 2\pi - n \log \sigma^2 + \log(1 - \rho^2) \\ &\quad - \sigma^{-2} \left\{ (Y_1 - \mu)^2 (1 - \rho^2) \right. \\ &\quad \left. + \sum_{t=2}^n [(Y_t - \mu) - \rho(Y_{t-1} - \mu)]^2 \right\}. \end{aligned} \quad (8.1.5)$$

The computation of the maximum likelihood estimators is greatly simplified if we treat  $Y_1$  as fixed and investigate the conditional likelihood. This is also an appropriate model in some experimental situations. For example, if we initiate an experiment at time 1 with an initial input of  $Y_1$ , it is very reasonable to condition on this initial input.

To construct the conditional likelihood, we consider the last  $n - 1$  equations of (8.1.3). Maximizing twice the logarithm of the likelihood,

$$\begin{aligned} 2 \log L(y: \mu, \rho, \sigma^2 | Y_1) &= -(n - 1) \log 2\pi - (n - 1) \log \sigma^2 \\ &\quad - \sigma^{-2} \sum_{t=2}^n (Y_t - \lambda - \rho Y_{t-1})^2, \end{aligned} \quad (8.1.6)$$

leads to the estimators

$$\begin{aligned} \hat{\rho} &= \left[ \sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})^2 \right]^{-1} \sum_{t=2}^n (Y_t - \bar{y}_{(0)})(Y_{t-1} - \bar{y}_{(-1)}), \\ \hat{\lambda} &= \bar{y}_{(0)} - \hat{\rho} \bar{y}_{(-1)}, \\ \hat{\sigma}^2 &= (n - 1)^{-1} \sum_{t=2}^n [(Y_t - \bar{y}_{(0)}) - \hat{\rho}(Y_{t-1} - \bar{y}_{(-1)})]^2, \end{aligned} \quad (8.1.7)$$

where  $(\bar{y}_{(-1)}, \bar{y}_{(0)}) = (n - 1)^{-1} \sum_{t=2}^n (Y_{t-1}, Y_t)$ . These estimators are, strictly speaking, not the maximum likelihood estimators for the model stated in (8.1.1). The estimator  $\hat{\rho}$  can take on values outside of  $(-1, 1)$ , while the maximum likelihood estimator is constrained to the parameter space.

The estimators of  $\lambda$  and  $\rho$  are those that would be obtained by applying least squares to the last  $n - 1$  equations. The least squares estimator of  $\sigma^2$ ,

$$s^2 = (n - 3)^{-1} \sum_{t=2}^n [(Y_t - \bar{y}_{(0)}) - \hat{\rho}(Y_{t-1} - \bar{y}_{(-1)})]^2, \quad (8.1.8)$$

is typically used in place of  $\hat{\sigma}^2$ .

The least squares estimator for  $\rho$  differs from (8.1.2) by terms whose order in probability is  $n^{-1}$ . Therefore, by (6.2.9) and Corollary 6.3.6.1,  $n^{1/2}(\hat{\rho} - \rho)$  is approximately normally distributed with mean zero and variance equal to  $1 - \rho^2$ . The limiting distribution is also derived in the next section. Note that the estimator

of  $\rho$  defined by (8.1.7) can be greater than one in absolute value, while that defined in (8.1.2) cannot.

Let us now return to a consideration of the unconditional likelihood as given by equation (8.1.5). Differentiating the log likelihood with respect to  $\mu$ ,  $\rho$ , and  $\sigma^2$  and setting the derivatives equal to zero, we obtain

$$\mu = [2 + (n - 2)(1 - \rho)]^{-1} \left[ Y_1 + (1 - \rho) \sum_{t=2}^{n-1} Y_t + Y_n \right],$$

$$[(Y_1 - \mu)^2 - (1 - \rho^2)^{-1} \sigma^2] \rho + \sum_{t=2}^n [(Y_t - \mu) - \rho(Y_{t-1} - \mu)](Y_{t-1} - \mu) = 0,$$

$$\sigma^2 = n^{-1} \left\{ (Y_1 - \mu)^2 (1 - \rho^2) + \sum_{t=2}^n [(Y_t - \mu) - \rho(Y_{t-1} - \mu)]^2 \right\}. \quad (8.1.9)$$

If  $\mu$  is known, Anderson (1971, p. 354) shows that the maximum likelihood estimator of  $\rho$  is a root of the cubic equation

$$f(\rho) = \rho^3 + c_1 \rho^2 + c_2 \rho + c_3 = 0, \quad (8.1.10)$$

where  $c_3 = -(n - 2)^{-1} n c_1$ ,

$$c_1 = -(n - 2)(n - 1)^{-1} \left[ \sum_{t=2}^{n-1} y_t^2 \right]^{-1} \sum_{t=2}^n y_t y_{t-1},$$

$$c_2 = -(n - 1)^{-1} \left[ n + \left[ \sum_{t=2}^{n-1} y_t^2 \right]^{-1} \sum_{t=1}^n y_t^2 \right],$$

and  $y_t = Y_t - \mu$ . Hasza (1980) gives explicit expressions for the three roots of (8.1.10) and shows that there is a root in each of the intervals  $(-\infty, -1)$ ,  $(-1, 1)$  and  $(1, \infty)$ . If  $y_t$  is stationary, then

$$[c_1, c_2] = -[\hat{\rho}_l, (n - 2)^{-1} n] + O_p(n^{-1}),$$

where  $\hat{\rho}_l = [\sum_{t=2}^n y_{t-1}^2]^{-1} \sum_{t=2}^n y_{t-1} y_t$  is the least squares estimator. We show in the next section that  $\hat{\rho}_l - \rho_0 = O_p(n^{-1/2})$ , where  $\rho_0$  is the true value. Hence,

$$f(\rho) \xrightarrow{P} \rho^3 - \rho_0 \rho^2 - \rho + \rho_0 = (\rho^2 - 1)(\rho - \rho_0).$$

It follows from the results of Section 5.8 that the three roots of  $f(\rho) = 0$  converge in probability to  $-1$ ,  $\rho_0$ , and  $1$ , respectively. Therefore, the root in  $(-1, 1)$  is consistent for  $\rho_0$ . We show in Section 8.4 that the least squares estimator (8.1.7) and the maximum likelihood estimator have the same limiting distribution for stationary processes.

If  $\mu$  is unknown, Gonzalez-Farias (1992) showed that the unconditional

maximum likelihood estimator is a solution of a fifth degree polynomial. A numerical solution can be obtained by iterating equation (8.1.10) and the estimator for  $\mu$  in (8.1.9), beginning with  $\hat{\mu} = \bar{y}_n$ .

## 8.2. HIGHER ORDER AUTOREGRESSIVE TIME SERIES

### 8.2.1. Least Squares Estimation for Univariate Processes

In this section we study the ordinary least squares estimators of the  $p$ th order autoregressive process. Consider the time series

$$Y_t + \sum_{i=1}^p \alpha_i Y_{t-i} = \theta_0 + e_t, \quad (8.2.1)$$

where the roots of

$$m^p + \sum_{i=1}^p \alpha_i m^{p-i} = 0 \quad (8.2.2)$$

are less than one in absolute value and the  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables with additional properties to be specified. Because the procedures of this section are closely related to multiple regression, it is convenient to write (8.2.1) as

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + \theta_2 Y_{t-2} + \cdots + \theta_p Y_{t-p} + e_t, \quad (8.2.3)$$

where  $\theta_i = -\alpha_i$ ,  $i = 1, 2, \dots, p$ . If the process is stationary, the expression

$$Y_t - \mu = \sum_{i=1}^p \theta_i (Y_{t-i} - \mu) + e_t, \quad (8.2.4)$$

where  $E\{Y_t\} = \mu = (1 + \sum_{i=1}^p \alpha_i)^{-1} \theta_0$ , is also useful.

The ordinary least squares estimator of  $\boldsymbol{\theta} = (\theta_0, \theta_1, \dots, \theta_p)'$  is

$$\hat{\boldsymbol{\theta}} = \left[ \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \right]^{-1} \sum_{t=p+1}^n \mathbf{X}'_t Y_t, \quad (8.2.5)$$

where  $\mathbf{X}_t = (1, Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$ . The error in this estimator is

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = \left[ (n-p)^{-1} \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \right]^{-1} (n-p)^{-1} \sum_{t=p+1}^n \mathbf{X}'_t e_t. \quad (8.2.6)$$

By the properties of autoregressive processes,  $E\{Y_{t-j} e_t\} = 0$  for  $j > 0$  and  $E\{\mathbf{X}'_t e_t\} = \mathbf{0}$ . Also, if the process is stationary with autocovariance function  $\gamma(h)$ , then

$$E\left\{(n-p)^{-1} \sum_{t=p+1}^n Y_{t-i} Y_{t-j}\right\} = \gamma(|i-j|) + \mu^2,$$

and under the conditions of Theorem 6.2.1,

$$(n-p)^{-1} \sum_{t=p+1}^n Y_{t-i} Y_{t-j} = \gamma(|i-j|) + \mu^2 + O_p(n^{-1/2}). \quad (8.2.7)$$

It follows that the error in  $\hat{\theta}$  is  $O_p(n^{-1/2})$ . Also, by (8.2.7) the least squares estimator is asymptotically equivalent to the estimator

$$\begin{pmatrix} \hat{\theta}_1^\dagger \\ \hat{\theta}_2^\dagger \\ \vdots \\ \hat{\theta}_p^\dagger \end{pmatrix} = \begin{pmatrix} \hat{\gamma}(0) & \hat{\gamma}(1) & \cdots & \hat{\gamma}(p-1) \\ \hat{\gamma}(1) & \hat{\gamma}(0) & \cdots & \hat{\gamma}(p-2) \\ \vdots & \vdots & & \vdots \\ \hat{\gamma}(p-1) & \hat{\gamma}(p-2) & \cdots & \hat{\gamma}(0) \end{pmatrix}^{-1} \begin{pmatrix} \hat{\gamma}(1) \\ \hat{\gamma}(2) \\ \vdots \\ \hat{\gamma}(p) \end{pmatrix}, \quad (8.2.8)$$

where  $\hat{\gamma}(h)$  is defined in (6.2.3). The estimator (8.2.8) is sometimes called the Yule-Walker estimator.

The least squares estimator of  $\sigma^2$  is

$$\hat{\sigma}^2 = (n - 2p - 1)^{-1} \sum_{t=p+1}^n \hat{e}_t^2, \quad (8.2.9)$$

where  $\hat{e}_t = Y_t - \mathbf{X}_t \hat{\theta}$ . The divisor for  $\hat{\sigma}^2$  is defined by analogy to ordinary regression theory. There are  $n - p$  observations in the regression, and  $p + 1$  parameters are estimated.

The asymptotic properties of  $\hat{\theta}$  and  $\hat{\sigma}^2$  are given in the following theorem. The theorem is stated and proven for martingale difference errors, but the result also holds for  $e_t$  that are iid  $(0, \sigma^2)$  random variables.

**Theorem 8.2.1.** Let  $Y_t$  satisfy

$$Y_t = \theta_0 + \sum_{j=1}^p \theta_j Y_{t-j} + e_t, \quad t = p + 1, p + 2, \dots \quad (8.2.10)$$

Assume  $(Y_1, Y_2, \dots, Y_p)$  is fixed, or assume  $(Y_1, Y_2, \dots, Y_p)$  is independent of  $e_t$  for  $t > p$  and  $E\{|Y_t|^{2+\nu}\} < \infty$  for  $i = 1, 2, \dots, p$  and some  $\nu > 0$ . Let the roots of

$$m^p - \sum_{j=1}^p \theta_j m^{p-j} = 0 \quad (8.2.11)$$

be less than one in absolute value. Suppose  $\{e_t\}_{t=1}^\infty$  is a sequence of  $(0, \sigma^2)$  random variables with

$$E\{(e_t, e_t^2) \mid \mathcal{A}_{t-1}\} = (0, \sigma^2) \quad \text{a.s.}$$

and

$$E\{|e_t|^{2+\nu} \mid \mathcal{A}_{t-1}\} < L < \infty \quad \text{a.s.}$$

for all  $t$  and some  $\nu > 0$ , where  $\mathcal{A}_{t-1}$  is the sigma-field generated by  $\{e_j : j \leq t-1\}$ . If  $(Y_1, Y_2, \dots, Y_p)$  is random, the sigma-field is generated by  $\{e_j : j \leq t-1\}$  and  $(Y_1, Y_2, \dots, Y_p)$ . Let

$$\hat{\theta} = \left[ \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \right]^{-1} \sum_{t=p+1}^n \mathbf{X}'_t Y_t,$$

where  $\mathbf{X}_t = (1, Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$ . Then

- (a)  $n^{1/2}(\hat{\theta} - \theta) \xrightarrow{d} N(\mathbf{0}, \mathbf{A}^{-1} \sigma^2)$ , and
- (b)  $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ ,

where

$$\mathbf{A} = \lim_{n \rightarrow \infty} n^{-1} \sum_{t=p+1}^n E(\mathbf{X}'_t \mathbf{X}_t),$$

and  $\hat{\sigma}^2$  is defined in (8.2.9).

**Proof.** We have

$$\hat{\theta} - \theta = \left[ \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \right]^{-1} \sum_{t=p+1}^n \mathbf{X}'_t e_t.$$

Given  $\epsilon > 0$ , there is some  $N_0$  such that  $\sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t$  is nonsingular with probability greater than  $1 - \epsilon$  for  $n > N_0$ . Let  $\eta$  be a column vector of arbitrary real numbers such that  $\eta' \eta \neq 0$ , and consider those samples for which  $\sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t$  is nonsingular. Let

$$n^{-1/2} \eta' \sum_{t=p+1}^n \mathbf{X}'_t e_t = \sum_{t=p+1}^n Z_{tn} = S_{nn},$$

where  $Z_{tn} = n^{-1/2} \eta' \mathbf{X}'_t e_t$ . Then  $E\{Z_{tn} \mid \mathcal{A}_{t-1}\} = 0$  a.s.,

$$\delta_m^2 = E\{Z_m^2 \mid \mathcal{A}_{t-1}\} = n^{-1} \eta' \mathbf{X}'_t \mathbf{X}_t \eta \sigma^2,$$

and

$$V_{nn}^2 = n^{-1} \sum_{t=p+1}^n \eta' \mathbf{X}'_t \mathbf{X}_t \eta \sigma^2.$$

By Theorem 6.3.5,

$$\operatorname{plim}_{n \rightarrow \infty} n^{-1} \sum_{t=p+1}^n \boldsymbol{\eta}' \mathbf{X}_t' \mathbf{X}_t \boldsymbol{\eta} \sigma^2 = \boldsymbol{\eta}' \mathbf{A} \boldsymbol{\eta} \sigma^2.$$

Now

$$\begin{aligned} E\{S_{nn}^2\} &= V\left\{n^{-1/2} \boldsymbol{\eta}' \sum_{t=p+1}^n \mathbf{X}_t' e_t\right\} \\ &= n^{-1} \sum_{t=p+1}^n E\{\boldsymbol{\eta}' \mathbf{X}_t' \mathbf{X}_t \boldsymbol{\eta}\} \sigma^2 \end{aligned}$$

and

$$\lim_{n \rightarrow \infty} s_{nn}^2 = \lim_{n \rightarrow \infty} E\{S_{nn}^2\} = \boldsymbol{\eta}' \mathbf{A} \boldsymbol{\eta} \sigma^2,$$

and hence condition ii of Theorem 5.3.4 is satisfied.

We now investigate

$$\begin{aligned} s_{nn}^{-2} \sum_{t=p+1}^n E\{Z_{tn}^2 I(|Z_{tn}| \geq \epsilon s_{nn})\} \\ &\leq s_{nn}^{-2} \sum_{t=p+1}^n E\{(\epsilon s_{nn})^{-\nu} |Z_{tn}|^{2+\nu} I(|Z_{tn}| \geq \epsilon s_{nn})\} \\ &\leq s_{nn}^{-2-\nu} \epsilon^{-\nu} \sum_{t=p+1}^n E\{|n^{-1/2} \boldsymbol{\eta}' \mathbf{X}_t' e_t|^{2+\nu}\} \\ &< s_{nn}^{-2-\nu} \epsilon^{-\nu} n^{-1-\nu/2} L \sum_{t=p+1}^n E\{|\boldsymbol{\eta}' \mathbf{X}_t'|^{2+\nu}\}. \end{aligned}$$

Let  $w_j$  be the weights of Theorem 2.6.1. By Holder's inequality,

$$\begin{aligned} \left| \sum_{j=0}^{t-1} w_j e_{t-j} \right|^{2+\nu} &\leq \left[ \sum_{j=0}^{t-1} |w_j| |e_{t-j}| \right]^{2+\nu} \\ &\leq \left[ \sum_{j=0}^{t-1} |w_j| \right]^{1+\nu} \sum_{j=0}^{t-1} |w_j| |e_{t-j}|^{2+\nu}. \end{aligned}$$

Therefore,  $E\{|\boldsymbol{\eta}' \mathbf{X}_t'|^{2+\nu}\}$  is bounded, condition iii of Theorem 5.3.4 is satisfied, and result a follows.

To prove part b, we observe that  $\hat{e}_t = e_t - \mathbf{X}_t(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$  and

$$\begin{aligned} \sum_{t=p+1}^n \hat{e}_t^2 &= \sum_{t=p+1}^n e_t^2 - (n-p)(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})' \mathbf{A}_n (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &= \sum_{t=p+1}^n e_t^2 + O_p(1) \end{aligned}$$

where  $\mathbf{A}_n = (n - p)^{-1} \sum_{i=p+1}^n \mathbf{X}'_i \mathbf{X}_i$ . The conclusion of part b follows because  $(n - 2p - 1)^{-1} \sum_{i=p+1}^n e_i^2$  converges to  $\sigma^2$  a.s. by Corollary 5.3.8. ▲

Because the matrix  $\mathbf{A}_n = (n - p)^{-1} \sum_{i=p+1}^n \mathbf{X}'_i \mathbf{X}_i$  converges to  $\mathbf{A}$  in probability, the usual regression distribution theory holds, approximately, for the autoregressive model. For example,  $v_{ii}^{-1/2}(\hat{\theta}_i - \theta_i)$ , where  $v_{ii}$  is the  $i$ th diagonal element of  $(\sum_{i=p+1}^n \mathbf{X}'_i \mathbf{X}_i)^{-1} \hat{\sigma}^2$ , is approximately a  $N(0, 1)$  random variable.

While we have obtained pleasant asymptotic results, the behavior of the estimators in small samples can deviate considerably from that based on asymptotic theory. If the roots of the autoregressive equation are near zero, the approach to normality is quite rapid. For example, if  $\rho = 0$  in the first order autoregressive process, the normal or  $t$ -distribution approximations will perform very well for  $n > 30$ . On the other hand, if the roots are near one in absolute value, very large samples may be required before the distribution is well approximated by the normal.

In Figure 8.2.1 we present the empirical density of the least squares estimator  $\hat{\rho}$

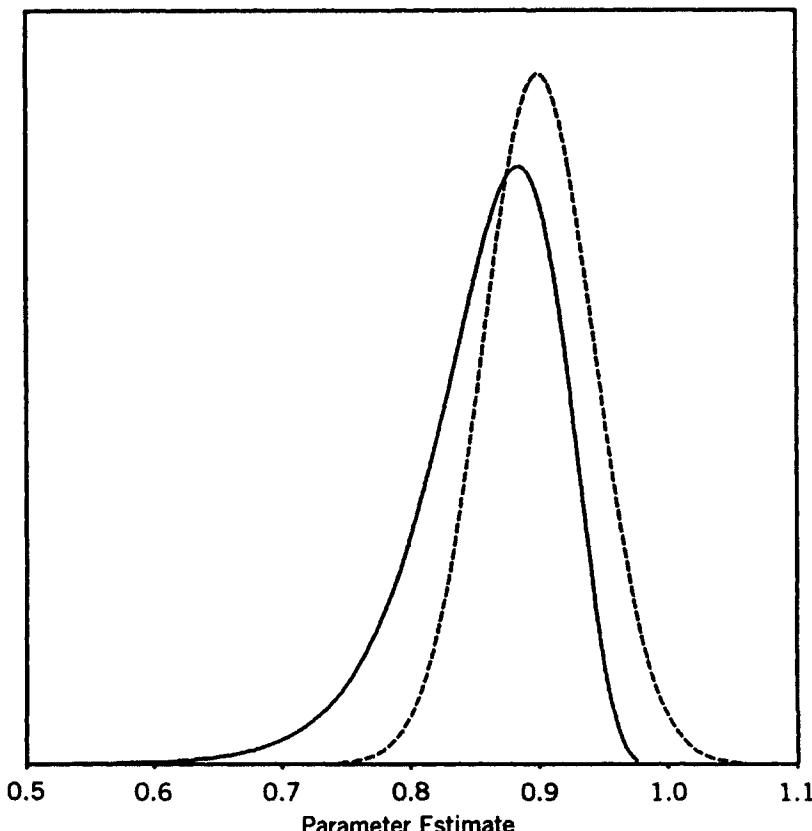


FIGURE 8.2.1. Estimated density of  $\hat{\rho}$  compared with normal approximation for  $\rho = 0.9$  and  $n = 100$ . (Dashed line is normal density.)

defined in (8.1.7). The density was estimated from 20,000 samples of size 100. The observations were generated by the autoregressive equation

$$Y_t = 0.9Y_{t-1} + e_t,$$

where the  $e_t$  are normal independent  $(0, 1)$  random variables. The empirical distribution displays a skewness similar to that which we would encounter in sampling from the binomial distribution. The mean of the empirical distribution is 0.861, and the variance is 0.0032. The distribution obtained from the normal approximation has a mean of 0.90 and a variance of 0.0019.

The mean of the empirical distribution agrees fairly well with the approximation obtained by the methods of Section 5.4. The bias approximated from a first order Taylor series is  $E\{\hat{\rho} - \rho\} \doteq -n^{-1}(1 + 3\rho)$ . This approximation to the expectation has been discussed by Mariott and Pope (1954) and Kendall (1954). Also see Pantula and Fuller (1985) and Shaman and Stine (1988).

Often the practitioner must determine the degree of autoregressive process as well as estimate the parameters. If it is possible to specify a maximum for the degree of the process, a process of that degree can first be estimated and high order terms discarded using the standard regression statistics. Anderson (1962) gives a procedure for this decision problem. Various model building methods based on regression theory can be used. Several such procedures are described in Draper and Smith (1981). In Section 8.4, we discuss other order determination procedures.

Often one inspects the residuals from the fit and perhaps computes the autocorrelations of these residuals. If the model is correct, the sample autocorrelations estimate zero with an error that is  $O_p(n^{-1/2})$ , but the variances of these estimators are generally smaller than the variances of estimators computed from a time series of independent random variables. See Box and Pierce (1970) and Ljung and Box (1978). Thus, while it is good practice to inspect the residuals, it is suggested that final tests of model adequacy be constructed by adding terms to the model and testing the hypothesis that the true value of the added coefficients is zero.

**Example 8.2.1.** To illustrate the regression estimation of the autoregressive process, we use the unemployment time series investigated in Section 6.3. The second order process, estimated by regressing  $Y_t - \bar{y}_n$  on  $Y_{t-1} - \bar{y}_n$  and  $Y_{t-2} - \bar{y}_n$ , is

$$\hat{Y}_t - 4.77 = 1.568(Y_{t-1} - 4.77) - 0.699(Y_{t-2} - 4.77), \\ (0.073) \quad (0.073)$$

where  $\bar{y}_{100} = 4.77$  and the numbers below the coefficients are the estimated standard errors obtained from the regression analysis. The residual mean square is 0.105.

If we condition the analysis on the first two observations and regress  $Y_t$  on  $Y_{t-1}$  and  $Y_{t-2}$  including an intercept term in the regression, we obtain

$$\hat{Y}_t = 0.63 + 1.568 Y_{t-1} - 0.699 Y_{t-2}. \\ (0.13) \quad (0.073) \quad (0.073)$$

The coefficients are slightly different from those in Section 6.4, since the coefficients in Section 6.4 were obtained from equation (8.2.8).

To check on the adequacy of the second order representation, we fit a fifth order process. The results are summarized in Table 8.2.1. The  $F$ -test for the hypothesis that the time series is second order autoregressive against the alternative that it is fifth order is

$$F_{89}^3 = 0.478[3(0.101)]^{-1} = 1.578.$$

The tabular 0.10 point for Snedecor's  $F$  with 3 and 89 degrees of freedom is 2.15, and so the null hypothesis is accepted at that level.  $\blacktriangle\blacktriangle$

### 8.2.2. Alternative Estimators for Autoregressive Time Series

The regression method of estimation is simple, easily understood, and asymptotically efficient for the parameters of stationary autoregressive processes. However, given the power of modern computing equipment, other procedures that are more efficient in small samples and (or) appropriate for certain models can be considered.

If the  $Y_t$  are normally distributed, the logarithm of the likelihood of a sample of  $n$  observations from a stationary  $p$ th order autoregressive process is the generalization of (8.1.5),

$$\log L(y; \theta) = -0.5n \log 2\pi - 0.5n \log |\Sigma_{yy}| - 0.5(\mathbf{Y} - \mathbf{J}\mu)' \Sigma_{yy}^{-1} (\mathbf{Y} - \mathbf{J}\mu), \quad (8.2.12)$$

where  $\mathbf{Y}' = (Y_1, Y_2, \dots, Y_n)$ ,  $\mathbf{J}' = (1, 1, \dots, 1)$ ,  $\Sigma_{yy}$  is the covariance matrix of  $\mathbf{Y}$  expressed as a function of  $(\sigma^2, \theta_1, \theta_2, \dots, \theta_p)$ , and

**Table 8.2.1. Analysis of Variance for Quarterly Seasonally Adjusted Unemployment rate, 1948 to 1972**

Source	Degrees of Freedom	Mean Square
$Y_{t-1}$	1	112.949
$Y_{t-2}$ after $Y_{t-1}$	1	9.481
$Y_{t-3}$ after $Y_{t-1}, Y_{t-2}$	1	0.305
$Y_{t-4}$ after $Y_{t-1}, Y_{t-2}, Y_{t-3}$	1	0.159
$Y_{t-5}$ after $Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4}$	1	0.014
Error	89	0.101

$$\mu = \left[ 1 - \sum_{i=1}^p \theta_i \right]^{-1} \theta_0. \quad (8.2.13)$$

Several computer packages contain algorithms that compute the maximum likelihood estimator. In Section 8.4, it is proven that the limiting distribution of the maximum likelihood estimator is the same as the limiting distribution of the ordinary least squares estimator.

By Corollary 2.6.1.3, a stationary autoregressive process can be given a forward representation

$$Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} = e_t$$

or a backward representation

$$Y_t + \sum_{j=1}^p \alpha_j Y_{t+j} = v_t,$$

where  $\{e_t\}$  and  $\{v_t\}$  are sequences of serially uncorrelated  $(0, \sigma^2)$  random variables. The ordinary least squares estimator of  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)$  is the value of  $\alpha$  that minimizes the sum of squares of the estimated  $e_t$ . One could also construct an estimator that minimizes the sum of squares of the estimated  $v_t$ . This suggests a class of estimators, where the estimator of  $\alpha$  is the  $\alpha$  that minimizes

$$Q(\alpha) = \sum_{t=p+1}^n w_t \left[ Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} \right]^2 + \sum_{t=1}^{n-p} (1 - w_{t+1}) \left[ Y_t + \sum_{j=1}^p \alpha_j Y_{t+j} \right]^2. \quad (8.2.14)$$

The ordinary least squares estimator is obtained by setting  $w_t = 1$ . The estimator obtained by setting  $w_t = 0.5$  was studied by Dickey, Hasza, and Fuller (1984). We call the estimator with  $w_t = 0.5$  the simple symmetric estimator. For the zero mean first order process, the simple symmetric estimator is

$$\hat{\alpha}_{1,s} = - \left[ \sum_{t=2}^{n-1} Y_t^2 + 0.5(Y_1^2 + Y_n^2) \right]^{-1} \sum_{t=2}^n Y_{t-1} Y_t.$$

Because  $|Y_{t-1} Y_t| \leq 0.5(Y_{t-1}^2 + Y_t^2)$ ,  $\hat{\alpha}_{1,s}$  for the first order process is always less than or equal to one in absolute value.

We call the estimator constructed with

$$w_t = \begin{cases} 0, & t = 1, 2, \dots, p, \\ (n - 2p + 2)^{-1}(t - p), & t = p + 1, p + 2, \dots, n - p + 1, \\ 1, & t = n - p + 2, n - p + 3, \dots, n, \end{cases} \quad (8.2.15)$$

the weighted symmetric estimator. The weights (8.2.15) assume  $n \geq 2p$ . The

weighted symmetric estimator is nearly identical to the maximum likelihood estimator unless one of the estimated roots is close to one. The roots of the weighted symmetric estimator with weights (8.2.15) are not restricted to be less than one in absolute value. For the zero mean first order process, the weighted symmetric estimator with weights (8.2.15) is

$$\hat{\alpha}_{1w} = - \left[ \sum_{t=2}^{n-1} Y_t^2 + n^{-1} \sum_{t=1}^n Y_t^2 \right]^{-1} \sum_{t=2}^n Y_{t-1} Y_t.$$

Thus, the least squares estimator,  $\hat{\alpha}_{1s}$ , and  $\hat{\alpha}_{1w}$  differ in the weights given to  $Y_1$  and  $Y_n$  in the divisor.

Table 8.2.2 contains the variables required for the symmetric estimation of the  $p$ th order process. The estimator is

$$\hat{\alpha} = -(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\mathbf{X}'\mathbf{W}\mathbf{Y},$$

where  $\mathbf{X}$  is the  $(2n - 2p) \times p$  matrix below the headings  $-\alpha_1, -\alpha_2, \dots, -\alpha_p$ ,  $\mathbf{Y}$  is the  $(2n - 2p)$ -dimensional column vector called the dependent variable, and  $\mathbf{W}$  is the  $(2n - 2p)$  diagonal matrix whose elements are given in the "Weight" column. An estimator of the covariance matrix of  $\hat{\alpha}$  is

$$(\mathbf{X}'\mathbf{W}\mathbf{X})^{-1}\hat{\sigma}^2,$$

where

$$\hat{\sigma}^2 = (n - p - 1)^{-1}(\mathbf{Y}'\mathbf{W}\mathbf{Y} - \hat{\alpha}'\mathbf{X}'\mathbf{W}\mathbf{Y}).$$

**Table 8.2.2. Data Arrangement for Regression Estimation of Autoregressive Parameters by the Weighted Symmetric Procedure**

Weight	Dependent Variable	Parameter			
		$-\alpha_1$	$-\alpha_2$	...	$-\alpha_p$
$w_{p+1}$	$Y_{p+1}$	$Y_p$	$Y_{p-1}$	...	$Y_1$
$w_{p+2}$	$Y_{p+2}$	$Y_{p+1}$	$Y_p$	...	$Y_2$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$w_n$	$Y_n$	$Y_{n-1}$	$Y_{n-2}$	...	$Y_{n-p}$
$1 - w_{n-p+1}$	$Y_{n-p}$	$Y_{n-p+1}$	$Y_{n-p+2}$	...	$Y_n$
$1 - w_{n-p}$	$Y_{n-p-1}$	$Y_{n-p}$	$Y_{n-p+1}$	...	$Y_{n-1}$
$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$
$1 - w_2$	$Y_1$	$Y_2$	$Y_3$	...	$Y_{p+1}$

If the mean is unknown, there are several ways to proceed. One option is to replace each  $Y_i$  in the table with  $Y_i - \bar{y}$ , where  $\bar{y} = n^{-1} \sum_{i=1}^n Y_i$ . The second is to replace the elements in each column of the table with  $Y_i - \bar{y}_{(i)}$ , where  $\bar{y}_{(i)}$  is the mean of the elements in the  $i$ th column. The third is to add a column of ones to the table and use the ordinary least squares formulas. The procedures are asymptotically equivalent, but the work of Park (1990) indicates that the use of separate means, or the column of ones, is preferred in small samples when the process has a root close to one in absolute value. If the estimator of the mean is of interest, the mean can be estimated by  $\bar{y}$ , or one can use estimated generalized least squares where the covariance matrix is based on the estimated autoregressive parameters.

If a regression program has a missing value option that omits an observation if any variable is missing, one can obtain the estimators by creating a data set composed of the original observations followed by a missing value, followed by the original data in reverse order. The created data set contains  $2n + 1$  "observations." Lagging the created vector  $p$  times gives the  $p$  explanatory variables required for the regression. Then calling the regression option that deletes observations with a missing value enables one to compute the simple symmetric estimator for any autoregression up to order  $p$ . The addition of weights is required to compute the weighted symmetric estimator.

The ideas of partial autocorrelation introduced in Section 1.4 can be used in a sequential estimation scheme for autoregressive models. Let a sample of  $n$  observations  $(Y_1, Y_2, \dots, Y_n)$  be available, and define  $X_i = Y_i - \bar{y}_n$ , where  $\bar{y}_n = n^{-1} \sum_{i=1}^n Y_i$ . Let

$$\hat{\theta}_{11} = \left[ \sum_{t=2}^n X_t^2 \sum_{t=2}^n X_{t-1}^2 \right]^{-1/2} \sum_{t=2}^n X_t X_{t-1} \quad (8.2.16)$$

be an estimator of the first autocorrelation. Then an estimator of the first order autoregressive equation is

$$\hat{Y}_t = \bar{y}_n (1 - \hat{\theta}_{11}) + \hat{\theta}_{11} Y_{t-1}, \quad (8.2.17)$$

and an estimator of the residual mean square for the autoregression is

$$\hat{\sigma}_{(11)}^2 = n(n-2)^{-1} (1 - \hat{\theta}_{11}^2) \hat{\gamma}(0), \quad (8.2.18)$$

where

$$\hat{\gamma}(0) = n^{-1} \sum_{t=1}^n (Y_t - \bar{y}_n)^2.$$

A test that the first order autocorrelation is zero under the assumption that higher order partial autocorrelations are zero is

$$t_1 = [(n-2)^{-1} (1 - \hat{\theta}_{11}^2)]^{-1/2} \hat{\theta}_{11}. \quad (8.2.19)$$

Under the null hypothesis, this statistic is approximately distributed as a  $N(0, 1)$  random variable in large samples.

Higher order partial autocorrelations, higher order autoregressions, higher order autocorrelations, and tests can be computed with the following formulas:

$$\begin{aligned}\hat{\theta}_{i+1,i+1} &= \left[ \sum_{t=i+2}^n W_t^2 \sum_{t=i+2}^n Z_{t-i-1,t}^2 \right]^{-1/2} \sum_{t=i+2}^n W_t Z_{t-i-1,t}, \\ \hat{\theta}_{j,i+1} &= \hat{\theta}_{ji} - \hat{\theta}_{i+1,i+1} \hat{\theta}_{i+1-j,i}, \quad j = 1, 2, \dots, i, \\ \hat{\sigma}_{(ii)}^2 &= n(n-1-i)^{-1} \hat{\rho}(0) [1 - \hat{R}_{(ii)}^2], \\ \hat{R}_{(ii)}^2 &= 1 - \prod_{j=1}^i (1 - \hat{\theta}_{jj}^2), \\ \hat{\rho}(i) &= \sum_{j=1}^i \hat{\theta}_{ji} \hat{\rho}(i-j), \\ \hat{\theta}_{0,i+1} &= \bar{y}_n \left[ 1 - \sum_{j=1}^{i+1} \hat{\theta}_{j,i+1} \right], \\ t_i &= [(n-1-i)^{-1} (1 - \hat{\theta}_{ii}^2)]^{-1/2} \hat{\theta}_{ii},\end{aligned}\tag{8.2.20}$$

where

$$(W_n, Z_n) = \left[ X_t - \sum_{j=1}^i \hat{\theta}_{jt} X_{t-j}, X_t - \sum_{j=1}^i \hat{\theta}_{jt} X_{t+j} \right],$$

and  $\hat{\rho}(0) = 1$ . The estimated partial autocorrelations  $\hat{\theta}_{ii}$  are defined for  $i = 1, 2, \dots, n-1$ , and  $t_i$  and  $\hat{\sigma}_{(ii)}^2$  are defined for  $i = 1, 2, \dots, n-2$ . A test that  $\theta_{ii} = 0$ , under the assumption that  $\theta_{jj} = 0$ ,  $j > i$ , is  $t_i$ , which is the generalization of  $t_1$  of (8.2.19).

If a  $p$ th order autoregression is selected as the representation for the time series, the covariance matrix of the vector of coefficients can be estimated with

$$\hat{V}\{\hat{\theta}_p\} = (n-1-p)^{-1} \hat{P}^{-1} \hat{\sigma}_{(pp)}^2, \tag{8.2.21}$$

where

$$\hat{P} = \begin{bmatrix} 1 & \hat{\rho}(1) & \hat{\rho}(2) & \cdots & \hat{\rho}(p-1) \\ \hat{\rho}(1) & 1 & \hat{\rho}(1) & \cdots & \hat{\rho}(p-2) \\ \vdots & \vdots & \vdots & & \vdots \\ \hat{\rho}(p-1) & \hat{\rho}(p-2) & \hat{\rho}(p-3) & \cdots & 1 \end{bmatrix},$$

$\hat{\theta}_p = (\hat{\theta}_{1p}, \hat{\theta}_{2p}, \dots, \hat{\theta}_{pp})'$ , and  $\hat{\rho}(h)$  is defined in (8.2.20). The matrix  $\hat{P}^{-1}$  is also given by

$$\mathbf{B}' \mathbf{S}^{-1} \mathbf{B}, \tag{8.2.22}$$

where

$$\mathbf{S} = \text{diag}\{1, 1 - \hat{R}_{(11)}^2, 1 - \hat{R}_{(22)}^2, \dots, 1 - \hat{R}_{(p-1,p-1)}^2\},$$

$$\mathbf{B} = \begin{bmatrix} 1 & 0 & 0 & \cdots & 0 \\ -\hat{\theta}_{11} & 1 & 0 & \cdots & 0 \\ -\hat{\theta}_{22} & -\hat{\theta}_{12} & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ -\hat{\theta}_{p-1,p-1} & -\hat{\theta}_{p-2,p-1} & -\hat{\theta}_{p-3,p-1} & \cdots & 1 \end{bmatrix}.$$

An alternative estimator of the partial autocorrelation is

$$\tilde{\theta}_{ii} = \left( \sum_{t=i+2}^n W_t^2 + \sum_{t=i+2}^n Z_{t-i,i-1}^2 \right)^{-1} 2 \sum_{t=i+2}^n W_t Z_{t-i,i-1}. \quad (8.2.23)$$

It can be verified that  $\tilde{\theta}_{ii}$  is also always less than one in absolute value. The estimator (8.2.23) in combination with (8.2.20) was suggested by Burg (1975).

The sequential method of computing the autoregression has the advantage for stationary time series that all roots of the estimated autoregression are less than one in absolute value. The estimators obtained by the sequential methods are very similar to the simple symmetric estimators. The sequential procedure also uses all available observations at each step. If regressions are only computed by regression in one direction, moving from an equation of order  $p - 1$  to an equation of order  $p$  involves dropping an observation and adding an explanatory variable. Hence, the maximum possible order for the one-direction regression procedure is  $\frac{1}{2}n$ . The sequential procedure defines the autoregression up to order  $n - 1$ .

The alternative estimators are asymptotically equivalent for stationary autoregressive processes.

**Theorem 8.2.2.** Let the assumptions of Theorem 8.2.1 hold. Then the limiting distribution of  $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ , where  $\hat{\boldsymbol{\theta}}$  is the maximum likelihood estimator, the simple symmetric estimator, the partial correlation estimator, or the weighted symmetric estimator, is the same as that given for the ordinary least squares estimator in Theorem 8.2.1.

**Proof.** Omitted. ▲

The maximum likelihood estimator is available in many computer packages and performs well in simulation studies for the correct model. The simple symmetric estimator, the weighted symmetric estimator, the Burg estimator, and the maximum likelihood estimator have similar efficiencies for stationary processes. The maximum likelihood estimator and the weighted symmetric estimator perform better than other estimators for processes with roots close to one in absolute value. The maximum normal likelihood estimator and the partial correlation methods

produce estimated equations such that the roots of the estimated characteristic equation are all less than one in absolute value. It is possible for the roots associated with ordinary least squares or with the weighted symmetric estimator to be greater than one in absolute value. The ordinary least squares estimator performs well for forward prediction and is recommended as a preliminary estimator if it is possible that the process is nonstationary with a root greater than one.

### 8.2.3. Multivariate Autoregressive Time Series

In this subsection, we extend the autoregressive estimation procedures to vector valued processes. Let  $\mathbf{Y}_t$  be a  $k$ -dimensional stationary process that satisfies the equation

$$\mathbf{Y}_t - \boldsymbol{\mu} + \sum_{i=1}^p \mathbf{A}_i (\mathbf{Y}_{t-i} - \boldsymbol{\mu}) = \mathbf{e}_t, \quad (8.2.24)$$

for  $t = p+1, p+2, \dots$ , where  $\mathbf{e}_t$  are independent  $(\mathbf{0}, \Sigma)$  random variables or martingale differences. If the process is stationary,  $E\{\mathbf{Y}_t\} = \boldsymbol{\mu}$ . The equation can also be written

$$\mathbf{Y}_t + \sum_{i=1}^p \mathbf{A}_i \mathbf{Y}_{t-i} = \boldsymbol{\theta}_0 + \mathbf{e}_t, \quad (8.2.25)$$

where  $\boldsymbol{\theta}_0$  is a  $k$ -dimensional column vector. Let  $\mathbf{Y}_1, \mathbf{Y}_2, \dots$  be observed. Discussion of estimation for vector autoregressive processes can proceed by analogy to the univariate case. Each of the equations in (8.2.25) can be considered as a regression equation. We write the  $i$ th equation of (8.2.24) as

$$Y_{it} = \theta_{0i} - \sum_{j=1}^p A_{ji} Y_{t-j} + e_{it}, \quad (8.2.26)$$

where  $Y_{it}$  is the  $i$ th element of  $\mathbf{Y}_t$ ,  $\theta_{0i}$  is the  $i$ th element of  $\boldsymbol{\theta}_0$ ,  $A_{ji}$  is the  $i$ th row of  $\mathbf{A}_j$ , and  $e_{it}$  is the  $i$ th element of  $\mathbf{e}_t$ .

Defining  $\boldsymbol{\theta}_i = (\theta_{0i}, -\mathbf{A}_{1i}, -\mathbf{A}_{2i}, \dots, -\mathbf{A}_{pi})'$ ,  $i = 1, 2, \dots, k$ , and  $\mathbf{X}_t = (1, \mathbf{Y}'_{t-1}, \mathbf{Y}'_{t-2}, \dots, \mathbf{Y}'_{t-p})$ , we can write equation (8.2.26) as

$$Y_{it} = \mathbf{X}_t \boldsymbol{\theta}_i + e_{it}. \quad (8.2.27)$$

On the basis of our scalar autoregressive results, we are led to consider the estimators

$$\hat{\boldsymbol{\theta}}_i = \left[ \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \right]^{-1} \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{Y}_{it}, \quad i = 1, 2, \dots, k. \quad (8.2.28)$$

If we let  $\mathbf{A}' = (-\boldsymbol{\theta}_0, \mathbf{A}_1, \mathbf{A}_2, \dots, \mathbf{A}_p)'$ , then the system (8.2.25) can be written

$$\mathbf{Y}'_t = -\mathbf{X}_t \mathbf{A}' + \mathbf{e}'_t, \quad t = p+1, p+2, \dots, n, \quad (8.2.29)$$

and the ordinary least squares estimator of  $\mathbf{A}'$  is

$$\hat{\mathbf{A}}' = - \left[ \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \right]^{-1} \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{Y}'_t, \quad (8.2.30)$$

where  $\hat{\theta}_i$  is the  $i$ th column of  $-\hat{\mathbf{A}}'$  and  $\theta_i$  is the  $i$ th column of  $-\mathbf{A}$ . The least squares estimator of  $\Sigma$  is

$$\hat{\Sigma} = [n - (k+1)p - 1]^{-1} \sum_{t=p+1}^n \hat{\mathbf{e}}_t \hat{\mathbf{e}}'_t, \quad (8.2.31)$$

where

$$\hat{\mathbf{e}}_t = \mathbf{Y}_t - \hat{\theta}_0 + \sum_{i=1}^p \hat{\mathbf{A}}_i \mathbf{Y}_{t-p}$$

and  $\hat{\mathbf{A}}$  is defined in (8.2.30).

An alternative method of computing the estimator of  $(\mathbf{A}_1, \dots, \mathbf{A}_p) = \mathbf{A}_{[2]}$  is as

$$\hat{\mathbf{A}}'_{[2]} = - \left[ \sum_{t=p+1}^n \mathbf{U}'_t \mathbf{U}_t \right]^{-1} \sum_{t=p+1}^n \mathbf{U}'_t (\mathbf{Y}_t - \bar{\mathbf{y}})', \quad (8.2.32)$$

where  $\mathbf{U}_t = (\mathbf{Y}'_{t-1} - \bar{\mathbf{y}}', \mathbf{Y}'_{t-2} - \bar{\mathbf{y}}', \dots, \mathbf{Y}'_{t-p} - \bar{\mathbf{y}}')$  and  $\bar{\mathbf{y}} = n^{-1} \sum_{t=1}^n \mathbf{Y}_t$ . The distributions of the estimators are the vector analogs of the distributions of Theorem 8.2.1.

**Theorem 8.2.3.** Let the vector time series  $\mathbf{Y}_t$  satisfy

$$\sum_{j=0}^p \mathbf{A}_j \mathbf{Y}_{t-j} = \theta_0 + \mathbf{e}_t$$

for  $t = p+1, p+2, \dots$ , where  $\{\mathbf{e}_t\}$  is a sequence of  $k$ -dimensional random variables and the  $\mathbf{A}_j$  are fixed  $k \times k$  matrices such that  $\mathbf{A}_0 = \mathbf{I}$  and the roots of

$$\left| \sum_{j=0}^p \mathbf{A}_j m^{p-j} \right| = 0$$

are less than one in absolute value. Suppose  $\{\mathbf{e}_t\}_{t=1}^\infty$  is a sequence of  $(\mathbf{0}, \Sigma)$  random vectors with

$$E\{(\mathbf{e}_t, \mathbf{e}_t \mathbf{e}'_t) | \mathcal{A}_{t-1}\} = (\mathbf{0}, \Sigma) \quad \text{a.s.}$$

and

$$E\{|\mathbf{e}_t|^{2+\nu} | \mathcal{A}_{t-1}\} < L < \infty \quad \text{a.s.}$$

for all  $t$  and some  $\nu > 0$ , where  $\mathcal{A}_{t-1}$  is the sigma-field generated by  $\{\mathbf{e}_j : j \leq t-1\}$ . Assume

$$\mathbf{G} = p\lim_{n \rightarrow \infty} (n-p)^{-1} \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t$$

is positive definite. If  $(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_p)$  are random, it is assumed that  $E\{\|\mathbf{Y}_i\|^{2+\nu}\} < \infty$  for  $i = 1, 2, \dots, p$  and the sigma-field is generated by  $\{\mathbf{e}_j : j \leq t-1\}$  and  $(\mathbf{Y}_1, \mathbf{Y}_2, \dots, \mathbf{Y}_p)$ . Then

$$n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{D}} N(\mathbf{0}, \boldsymbol{\Sigma} \otimes \mathbf{G}^{-1}),$$

and  $\hat{\boldsymbol{\Sigma}} \xrightarrow{P} \boldsymbol{\Sigma}$ , where  $\hat{\boldsymbol{\theta}}_i$  is defined in (8.2.28),  $\hat{\boldsymbol{\Sigma}}$  is defined in (8.2.31),

$$\boldsymbol{\theta}' = (\boldsymbol{\theta}'_1, \boldsymbol{\theta}'_2, \dots, \boldsymbol{\theta}'_k),$$

and  $\boldsymbol{\Sigma} \otimes \mathbf{G}^{-1}$  is the Kronecker product of the matrices  $\boldsymbol{\Sigma}$  and  $\mathbf{G}^{-1}$ .

**Proof.** We only outline the proof, because it differs in no substantive way from that of Theorem 8.2.1. Since  $\mathbf{Y}_i$  is converging to a stationary time series,  $p\lim(n-p)^{-1} \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t = \mathbf{G}$  by Corollary 6.3.5. Therefore, the limiting distribution of  $\hat{\boldsymbol{\theta}}$  follows from that of  $n^{1/2}(n-p)^{-1} \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{e}_i$ ,  $i = 1, 2, \dots, k$ . By arguments parallel to those of Theorem 8.2.1, we obtain the limiting distribution. If  $\boldsymbol{\Sigma}$  is singular, then the limiting distribution contains singular components.  $\blacktriangle$

An alternative estimator for the vector process is the maximum likelihood estimator which can be computed by numerical procedures. The limiting distribution of the maximum likelihood estimator is the same as that of the least squares estimator for stationary vector processes.

### 8.3. MOVING AVERAGE TIME SERIES

We have seen that ordinary least squares regression procedures can be used to obtain efficient estimators for the parameters of autoregressive time series. Unfortunately, the estimation for moving average processes is less simple.

By the results of Chapter 2 we know that there is a relationship between the correlation function of a moving average time series and the parameters of the time series. For example, for the first order process,  $\rho(1) = \beta(1 + \beta^2)^{-1}$ , where  $\beta$  is the parameter of the process. On this basis, we might be led to estimate  $\beta$  from an estimator of  $\rho(1)$ . Since we demonstrated in Chapter 6 that

$$\hat{\rho}(1) = \left[ \sum_{t=1}^n (Y_t - \bar{y}_n)^2 \right]^{-1} \sum_{t=2}^n (Y_t - \bar{y}_n)(Y_{t-1} - \bar{y}_n)$$

estimates  $\rho(1)$  with an error which is  $O_p(n^{-1/2})$ , it follows that

$$\hat{\beta}_r = \begin{cases} [2\hat{f}(1)]^{-1}\{1 - [1 - 4\hat{f}^2(1)]^{1/2}\}, & 0 < |\hat{f}(1)| \leq 0.5, \\ -1, & \hat{f}(1) < -0.5, \\ 1, & \hat{f}(1) > 0.5, \\ 0, & \hat{f}(1) = 0, \end{cases} \quad (8.3.1)$$

estimates  $\beta$  with an error of the same order. Obviously, if  $\hat{f}(1)$  lies outside the range  $(-0.5, 0.5)$  by a significant amount, the model of a first order moving average is suspect.

It follows from equation (6.2.8) that

$$\text{Var}\{\hat{f}(1)\} = n^{-1}(1 + \beta^2)^{-4}(1 + \beta^2 + 4\beta^4 + \beta^6 + \beta^8) + O(n^{-2}). \quad (8.3.2)$$

Because the derivative of  $\rho(1)$  with respect to  $\beta$  is  $(1 + \beta^2)^{-2}(1 - \beta^2)$ , the approximate variance of  $\hat{\beta}$ , is, for  $|\beta| < 1$ ,

$$\text{Var}\{\hat{\beta}\} = n^{-1}(1 - \beta^2)^{-2}(1 + \beta^2 + 4\beta^4 + \beta^6 + \beta^8). \quad (8.3.3)$$

While the estimator (8.3.1) is consistent, we shall see that it is inefficient for  $\beta \neq 0$ . Roughly, the other sample autocorrelations contain information about  $\beta$ .

More efficient estimators can be obtained by least squares or by likelihood methods. We first present an estimation procedure based on the Gauss–Newton method of estimating the parameters of a nonlinear function discussed in Chapter 5. One purpose for describing the procedure is to demonstrate the nature of the derivatives that define the limiting distribution of the estimator. In practice, there are a number of computer programs available to perform the numerical estimation. Maximum likelihood estimation is discussed in Section 8.4.

Consider the first order moving average

$$Y_t = e_t + \beta e_{t-1}, \quad (8.3.4)$$

where  $|\beta| < 1$  and the  $e_t$  are independent  $(0, \sigma^2)$  random variables. Equation (8.3.4) may also be written

$$e_t = -\beta e_{t-1} + Y_t, \quad (8.3.5)$$

and, using our previous difference equation results, we have

$$\begin{aligned} Y_t &= -\sum_{j=1}^{t-1} (-\beta)^j Y_{t-j} - (-\beta)^t e_0 + e_t \\ &= f_t(Y; \beta, e_0) + e_t, \end{aligned} \quad (8.3.6)$$

where

$$f_t(Y; \beta, e_0) = \beta e_0$$

and

$$f_t(Y; \beta, e_0) = - \sum_{j=1}^{t-1} (-\beta)^j Y_{t-j} - (-\beta)^t e_0 \quad (8.3.7)$$

for  $t > 1$ . The expression (8.3.6) places the estimation problem in the nonlinear estimation format of Section 5.5.2.

We assume initial estimators  $\tilde{\beta}$  and  $\tilde{e}_0$  satisfying  $(\tilde{\beta} - \beta) = o_p(n^{-1/4})$  and  $\tilde{e}_0 = O_p(1)$  are available. These requirements are satisfied if one uses  $\tilde{e}_0 = 0$  and the estimator  $\hat{\beta}_t$  of (8.3.1). The one-step Gauss-Newton estimator of  $\beta$  is obtained by regressing the deviations

$$e_t(Y; \tilde{\beta}) = Y_t - f_t(Y; \tilde{\beta}, \tilde{e}_0) = \sum_{j=0}^{t-1} (-\tilde{\beta})^j Y_{t-j} + (-\tilde{\beta})^t \tilde{e}_0 \quad (8.3.8)$$

on the first derivative of  $f_t(Y; \beta, e_0)$  evaluated at  $\beta = \tilde{\beta}$ ; that derivative is

$$W_t(Y; \tilde{\beta}) = \begin{cases} \tilde{e}_0, & t = 1, \\ \sum_{j=1}^{t-1} j(-\tilde{\beta})^{j-1} Y_{t-j} + t(-\tilde{\beta})^{t-1} \tilde{e}_0, & t = 2, 3, \dots, n. \end{cases} \quad (8.3.9)$$

We could also include  $e_0$  as a "random parameter" to be estimated. The inclusion of the derivative for a change in  $\tilde{e}_0$  does not affect the limiting distribution of the estimator of  $\beta$  for invertible moving averages. Therefore, we simplify our discussion by considering only the derivative with respect to  $\beta$ .

The computation of  $e_t(Y; \tilde{\beta})$  and  $W_t(Y; \tilde{\beta})$  is simplified by noting that both satisfy difference equations:

$$e_t(Y; \tilde{\beta}) = \begin{cases} Y_1 - \tilde{\beta} \tilde{e}_0, & t = 1, \\ Y_t - \tilde{\beta} e_{t-1}(Y; \tilde{\beta}), & t = 2, 3, \dots, n, \end{cases} \quad (8.3.10)$$

and

$$W_t(Y; \tilde{\beta}) = \begin{cases} \tilde{e}_0, & t = 1, \\ e_{t-1}(Y; \tilde{\beta}) - \tilde{\beta} W_{t-1}(Y; \tilde{\beta}), & t = 2, 3, \dots, n. \end{cases} \quad (8.3.11)$$

The difference equation for  $e_t(Y; \tilde{\beta})$  follows directly from (8.3.5). Equation (8.3.11) can be obtained by differentiating both sides of

$$e_t(Y; \beta) = Y_t - \beta e_{t-1}(Y; \beta)$$

with respect to  $\beta$  and evaluating the resulting expression at  $\beta = \tilde{\beta}$ .

Regressing  $e_t(Y; \tilde{\beta})$  on  $W_t(Y; \tilde{\beta})$ , we obtain an estimator of  $\beta - \tilde{\beta}$ . The improved estimator of  $\beta$  is then

$$\hat{\beta} = \tilde{\beta} + \Delta \hat{\beta},$$

where

$$\Delta \hat{\beta} = \left[ \sum_{t=1}^n [W_t(Y; \tilde{\beta})]^2 \right]^{-1} \sum_{t=1}^n e_t(Y; \tilde{\beta}) W_t(Y; \tilde{\beta}). \quad (8.3.12)$$

The asymptotic properties of  $\hat{\beta}$  are developed in Theorem 8.3.1. An interesting result is that the large sample behavior of the estimator of the moving average parameter  $\beta$  is the same as that of the estimator of the autoregressive process with parameter  $-\beta$ . The limiting distribution follows from the fact that the derivative (8.3.11) evaluated at the true  $\beta$  is an autoregressive process.

**Theorem 8.3.1.** Let  $Y_t$  satisfy (8.3.4), where  $|\beta^0| < 1$ ,  $\beta^0$  is the true value, and the  $e_t$  are independent  $(0, \sigma^2)$  random variables with  $E\{|e_t|^{2+\nu}\} < L < \infty$  for some  $\nu > 0$ . Let  $\tilde{e}_0$  and  $\tilde{\beta}$  be initial estimators satisfying  $\tilde{e}_0 = O_p(1)$ ,  $\tilde{\beta} - \beta = o_p(n^{-1/4})$ , and  $|\tilde{\beta}| < 1$ . Then

$$n^{1/2}(\hat{\beta} - \beta^0) \xrightarrow{\mathcal{L}} N[0, 1 - (\beta^0)^2],$$

where  $\hat{\beta}$  is defined in (8.3.12). Also,  $\hat{\sigma}^2 \xrightarrow{P} (\sigma^0)^2$ , where  $\sigma^0$  is the true value of  $\sigma$  and

$$\hat{\sigma}^2 = n^{-1} \sum_{t=1}^n e_t^2(Y; \hat{\beta}).$$

**Proof.** The model (8.3.6) is of the form (5.5.52) discussed in Section 5.5.2. The first derivative of  $f_t(Y; \beta)$  is given in (8.3.9). The next two derivatives are

$$\begin{aligned} \frac{\partial^2 f_t(Y; \beta)}{\partial \beta^2} &= H_t(Y; \beta) = - \sum_{j=2}^{t-1} j(j-1)(-\beta)^{j-2} Y_{t-j} - t(t-1)(-\beta)^{t-2} e_0, \\ \frac{\partial^3 f_t(Y; \beta)}{\partial \beta^3} &= G_t(Y; \beta) \\ &= \sum_{j=3}^{t-1} j(j-1)(j-2)(-\beta)^{j-3} Y_{t-j} + t(t-1)(t-2)(-\beta)^{t-3} e_0, \end{aligned}$$

where it is understood that the summation is defined as if  $Y_t = 0$  for  $t \leq 0$ . Let  $\bar{S}$  be a closed interval containing  $\beta$  as an interior point and such that  $\max_{\beta \in \bar{S}} |\beta| < \lambda < 1$ . By Corollary 2.2.2.3,  $f_t(Y; \beta)$  and the derivatives are moving averages with exponentially declining weights for all  $\beta$  in  $\bar{S}$ . Hence, for  $\beta$  in  $\bar{S}$  they converge to stationary infinite moving average time series, the effect of  $e_0$  being transient. It follows from Theorem 6.3.5 that the sample covariances and autocovariances of the four time series converge in probability for all  $\beta$  in  $\bar{S}$  and the limits are continuous functions of  $\beta$ . Convergence is uniform on the compact set  $\bar{S}$ . Hence, conditions 1 and 2 of Theorem 5.5.4 are satisfied and

$$\hat{\beta} - \beta = \left[ \sum_{t=1}^n W_t^2(Y; \beta^0) \right]^{-1} \sum_{t=1}^n W_t(Y; \beta^0) e_t + o_p(n^{-1/2}).$$

If  $\beta = \beta^0$ ,

$$\begin{aligned} W_t(Y; \beta^0) &= \sum_{j=1}^{t-1} j(-\beta^0)^{j-1}(e_{t-j} + \beta^0 e_{t-j-1}) + t(-\beta^0)^{t-1}\tilde{e}_0 \\ &= \sum_{j=1}^{t-1} (-\beta^0)^{j-1}e_{t-j} + (-\beta^0)^{t-1}[t\tilde{e}_0 - (t-1)e_0] \end{aligned}$$

and  $W_t(Y; \beta^0)$  is converging to a stationary first order autoregressive process with parameter  $-\beta^0$ . Hence,

$$n^{-1} \sum_{t=1}^n W_t^2(Y; \beta^0) \xrightarrow{P} [1 - (\beta^0)^2]^{-1} \sigma^2$$

by Theorem 6.3.5, and

$$n^{-1/2} \sum_{t=1}^n W_t(Y; \beta^0) e_t \xrightarrow{\mathcal{L}} N(0, [1 - (\beta^0)^2]^{-1} \sigma^4)$$

by the arguments used in the proof of Theorem 8.2.1. Thus, the limiting distribution for  $n^{1/2}(\hat{\beta} - \beta^0)$  is established.

Because  $\hat{\sigma}^2$  is a continuous function of  $\hat{\beta}$  and because  $n^{-1} \sum_{t=1}^n e_t^2(Y; \beta)$  converges uniformly on  $\bar{S}$ , it follows that  $\hat{\sigma}^2$  converges to  $(\sigma^0)^2$  in probability. ▲

Comparison of the result of Theorem 8.3.1 and equation (8.3.3) establishes the large sample inefficiency of the estimator constructed from the first order autocorrelation. By the results of Theorem 8.3.1, we can use the regular regression statistics as approximations when drawing inferences about  $\beta$ .

In our discussion we have assumed that the mean of the time series was known and taken to be zero. It can be demonstrated that the asymptotic results hold for  $Y_t$  replaced by  $Y_t - \bar{y}_n$ , where  $\bar{y}_n$  is the sample mean.

The procedure can be iterated using  $\hat{\beta}$  as the initial estimator. In the next section we discuss estimators that minimize a sum of squares criterion or maximize a likelihood criterion. The asymptotic distribution of those estimators is the same as that obtained in Theorem 8.3.1. However, the estimators of the next section generally perform better in small samples.

A method of obtaining initial estimators that is applicable to higher order processes is an estimation procedure suggested by Durbin (1959). By Theorem 2.6.2, any  $q$ th order moving average

$$Y_t = \sum_{s=1}^q \beta_s e_{t-s} + e_t, \quad (8.3.13)$$

for which the roots of the characteristic equation are less than one can be represented in the form

$$Y_t = - \sum_{j=1}^{\infty} c_j Y_{t-j} + e_t,$$

where the weights satisfy the difference equation  $c_1 = -\beta_1, c_2 = -\beta_2 - \beta_1 c_1, \dots$

$$\begin{aligned} c_3 &= -\beta_3 - \beta_1 c_2 - \beta_2 c_1, \\ &\vdots \\ c_q &= -\beta_q - \beta_1 c_{q-1} - \beta_2 c_{q-2} - \cdots - \beta_{q-1} c_1, \\ c_j &= - \sum_{m=1}^q \beta_m c_{j-m}, \quad j = q+1, q+2, \dots \end{aligned} \quad (8.3.14)$$

Since the weights  $c_j$  are sums of powers of the roots, they decline in absolute value, and one can terminate the sum at a convenient finite number, say  $k$ . Then we can write

$$Y_t = - \sum_{j=1}^k c_j Y_{t-j} + e_t. \quad (8.3.15)$$

On the basis of this approximation, we treat  $Y_t$  as a finite autoregressive process and estimate  $c_j, j = 1, 2, \dots, k$ , by the regression procedures of Section 8.2. As the true weights satisfy equation (8.3.14), we treat the estimated  $c_j$ 's as a finite autoregressive process satisfying (8.3.14) and estimate the  $\beta$ 's. That is, we treat

$$\hat{c}_j = - \sum_{s=1}^q \hat{c}_{j-s} \beta_s \quad (8.3.16)$$

as a regression equation and estimate the  $\beta$ 's by regressing  $-\hat{c}_j$  on  $\hat{c}_{j-1}, \hat{c}_{j-2}, \dots, \hat{c}_{j-q}$ , where the appropriate modifications must be made for  $j = 1, 2, \dots, q$ , as per (8.3.14).

If we let  $\{k_n\}$  be a sequence such that  $k_n = o(n^{1/3})$  and  $k_n \rightarrow \infty$  as  $n \rightarrow \infty$ , it is possible to use the results of Berk (1974) to demonstrate that

$$\sum_{j=1}^{k_n} (\hat{c}_j - c_j)^2 = O_p(k_n n^{-1}).$$

It follows that the preliminary estimators constructed from the  $\hat{c}_j$  will have an error that is  $o_p(n^{-1/3})$ .

In carrying out the Gauss–Newton procedure, initial values  $\tilde{e}_{1-q}, \tilde{e}_{2-q}, \dots, \tilde{e}_0$  are required. The simplest procedure is to set them equal to zero. Alternatively, the autoregressive equation (8.3.15) can be used to estimate the  $Y$ -values preceding the sample period. Recall that a stationary autoregressive process can be written in either a forward or a backward manner (Corollary 2.6.1.2) and, as a result, the extrapolation formula for  $Y_0$  is of the same form as that for  $Y_{n+1}$ . If the true process is a  $q$ th order moving average, one uses the autoregressive equation to

estimate  $q$  values, since the best predictors for  $Y_t$ ,  $t \leq -q$ , are zero. Thus, one predicts  $Y_0, Y_{-1}, \dots, Y_{-q}$ , sets  $Y_t = 0$  for  $t < 1 - q$ , and uses equation (8.3.13) and the predicted  $Y$ -values to estimate  $e_{1-q}, e_{2-1}, \dots, e_0$ .

**Example 8.3.1.** We illustrate the Gauss–Newton procedure by fitting a first order moving average to an artificially generated time series. Table 8.3.1 contains 100 observations on  $X_t$  defined by  $X_t = 0.7e_{t-1} + e_t$ , where the  $e_t$  are computer generated normal independent  $(0, 1)$  random variables. We assume that we know the mean of the time series is zero.

As the first step in the analysis, we fit a seventh order autoregressive model to the data. This yields

$$\begin{aligned}\hat{Y}_t = & 0.685 Y_{t-1} - 0.584 Y_{t-2} + 0.400 Y_{t-3} - 0.198 Y_{t-4} \\ & (0.108) \quad (0.131) \quad (0.149) \quad (0.152) \\ & + 0.020 Y_{t-5} + 0.014 Y_{t-6} + 0.017 Y_{t-7}, \\ & (0.149) \quad (0.133) \quad (0.108)\end{aligned}$$

**Table 8.3.1. One Hundred Observations from a First Order Moving Average Time Series with  $\beta = 0.7$**

	First 25	Second 25	Third 25	Fourth 25
1	1.432	1.176	-1.311	2.607
2	-0.343	0.846	-0.105	1.572
3	-1.759	0.079	0.313	-0.261
4	-2.537	0.815	-0.890	-0.686
5	-0.295	2.566	-1.778	-2.079
6	0.689	1.675	-0.202	-2.569
7	-0.633	0.933	0.450	-0.524
8	-0.662	0.284	-0.127	0.044
9	-0.229	0.568	-0.463	-0.088
10	-0.851	0.515	0.344	-1.333
11	-3.361	-0.436	-1.412	-1.977
12	-0.912	0.567	-1.525	0.120
13	1.594	1.040	-0.017	1.558
14	1.618	0.064	-0.525	0.904
15	-1.260	-1.051	-2.689	-1.437
16	0.288	-1.845	-0.211	0.427
17	0.858	0.281	2.145	0.061
18	-1.752	-0.136	0.787	0.120
19	-0.960	-0.992	-0.452	1.460
20	1.738	0.321	1.267	-0.493
21	-1.008	2.621	2.316	-0.888
22	-1.589	2.804	0.258	-0.530
23	0.289	2.174	-1.645	-2.757
24	-0.580	1.897	-1.552	-1.452
25	1.213	-0.781	-0.213	0.158

where the numbers in parentheses are the estimated standard errors obtained from a standard regression program. The residual mean square for the regression is 1.22. The first few regression coefficients are declining in absolute magnitude with alternating signs. Since the third coefficient exceeds twice its standard error, a second order autoregressive process would be judged an inadequate representation for this realization. Thus, even if one did not know the nature of the process generating the data, the moving average representation would be suggested as a possibility by the regression coefficients. The regression coefficients estimate the negatives of the  $c_j$ ,  $j \geq 1$ , of Theorem 2.6.2. By that theorem,  $\beta = -c_1$ , and  $c_j = -\beta c_{j-1}$ ,  $j = 2, 3, \dots$ . Therefore, we arrange the regression coefficients as in Table 8.3.2. The initial estimator of  $\beta$  is obtained by regressing the first row on the second row of that table. This regression yields  $\tilde{\beta} = 0.697$ .

Our initial estimator for  $e_0$  is

$$\begin{aligned}\tilde{e}_0 &= 0.685Y_1 - 0.584Y_2 + 0.400Y_3 - 0.198Y_4 + 0.020Y_5 + 0.014Y_6 + 0.017Y_7 \\ &= 0.974.\end{aligned}$$

The values of  $e_t(Y; 0.697)$  are calculated using (8.3.10), and the values of  $W_t(Y; 0.697)$  are calculated using (8.3.11). We have

$$e_t(Y; 0.697)$$

$$= \begin{cases} Y_1 - 0.697\tilde{e}_0 = 1.432 - 0.697(0.974) = 0.753, & t = 1, \\ Y_t - 0.697e_{t-1}(Y; 0.697), & t = 2, 3, \dots, 100, \end{cases}$$

$$W_t(Y; 0.697)$$

$$= \begin{cases} \tilde{e}_0 = 0.974, & t = 1, \\ e_{t-1}(Y; 0.697) - 0.697W_{t-1}(Y; 0.697), & t = 2, 3, \dots, 100. \end{cases}$$

The first five observations are displayed in Table 8.3.3. Regressing  $e_t(Y; \tilde{\beta})$  on  $W_t(Y; \tilde{\beta})$  gives a coefficient of 0.037 and an estimate of  $\hat{\beta} = 0.734$ . The estimated standard error is 0.076, and the residual mean square is 1.16.

Table 8.3.2. Observations for Regression Estimation of an Initial Estimate for  $\beta$

$j$	Regression Coefficient $-c_j$	Multiplier of $\beta$
1	0.685	1
2	-0.584	-0.685
3	0.400	0.584
4	-0.198	-0.400
5	0.020	0.198
6	0.014	-0.020
7	0.017	-0.014

**Table 8.3.3. First Five Observations Used in Gauss-Newton Computation**

<i>t</i>	$e_i(Y; 0.697)$	$W_i(Y; 0.697)$
1	0.753	0.974
2	-0.868	0.074
3	-1.154	-0.920
4	-1.732	-0.513
5	0.913	-1.375

Maximum likelihood estimation of the parameters of moving average processes is discussed in Section 8.4. Most computer programs offer the user the option of specifying initial values for the likelihood maximization routine or of permitting the program to find initial values. Because of our preliminary analysis, we use 0.697 as our initial value in the likelihood computations. The maximum likelihood estimator computed in SAS/ETS® is 0.725 with an estimated standard error of 0.071. The maximum likelihood estimator of the error variance adjusted for degrees of freedom is  $\hat{\sigma}^2 = 1.17$ . The similarity of the two sets of estimates is a reflection of the fact that the estimators have the same limiting distribution. ▲▲

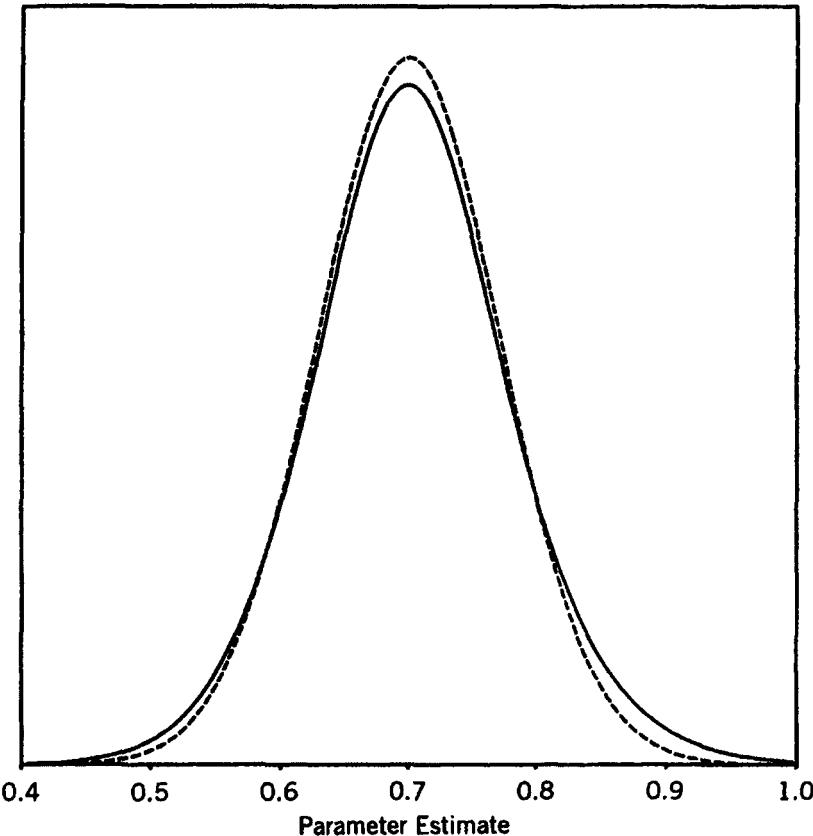
The theory we have presented is all for large samples. There is some evidence that samples must be fairly large before the results are applicable. For example, Macpherson (1981) and Nelson (1974) have conducted Monte Carlo studies in which the empirical variances for the estimated first order moving average parameter are about 1.1 to 1.2 times that based on large sample theory for samples of size 100 and  $0 \leq \beta \leq 0.7$ . Unlike the estimator for the autoregressive process, the distribution of the nonlinear estimator of  $\beta$  for  $\beta$  near zero differs considerably from that suggested by asymptotic theory for  $n$  as large as 100. In the Macpherson study the variance of  $\hat{\beta}$  for  $\beta = 0.1$  and  $n = 100$  was 1.17 times that suggested by asymptotic theory.

In Figure 8.3.1 we compare an estimate of the density of  $\hat{\beta}$  for  $\beta = 0.7$  with the normal density suggested by the asymptotic theory. The empirical density is based on 15,000 samples of size 100. The estimator of  $\beta$  is the maximum likelihood estimator determined by a grid search procedure.

The estimated density for  $\hat{\beta}$  is fairly symmetric about 0.7 with a mean of 0.706. However, the empirical density is considerably flatter than the normal approximation. The variance of the empirical distribution is 0.0064, compared to 0.0051 for the normal approximation.

## 8.4. AUTOREGRESSIVE MOVING AVERAGE TIME SERIES

In this section, we treat estimation of the parameters of time series with representation



**FIGURE 8.3.1.** Estimated density of maximum likelihood estimator of  $\beta$  compared with normal approximation for  $\beta = 0.7$  and  $n = 1.00$ . (Dashed line is normal density.)

$$Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} = e_t + \sum_{i=1}^q \beta_i e_{t-i}, \quad (8.4.1)$$

where the  $e_t$  are independent  $(0, \sigma^2)$  random variables, and the roots of

$$A(m; \alpha) = m^p + \sum_{j=1}^p \alpha_j m^{p-j} = 0, \quad (8.4.2)$$

and of

$$B(s; \beta) = s^q + \sum_{i=1}^q \beta_i s^{q-i} = 0 \quad (8.4.3)$$

are less than one in absolute value.

Estimators of the parameters of the process can be defined in a number of different ways. We consider three estimators obtained by minimizing three

different functions. If  $Y_t$  is a stationary normal time series, the logarithm of the likelihood is

$$L_n(\zeta) = -0.5n \log 2\pi - 0.5 \log |\Sigma_{YY}(\zeta)| - 0.5 \mathbf{Y}' \Sigma_{YY}^{-1}(\zeta) \mathbf{Y} \quad (8.4.4)$$

where  $\zeta' = (\theta', \sigma^2)$ ,  $\theta' = (\alpha_1, \alpha_2, \dots, \alpha_p, \beta_1, \beta_2, \dots, \beta_q)$ ,  $\mathbf{Y}' = (Y_1, Y_2, \dots, Y_n)$ , and  $\Sigma_{YY} = \Sigma_{YY}(\zeta) = E\{\mathbf{Y}\mathbf{Y}'\}$ .

The estimator that maximizes  $L_n(\zeta)$  is often called the maximum likelihood estimator or Gaussian likelihood estimator even if  $Y_t$  is not normal. Let  $\sigma^{-2} \Sigma_{YY}(\zeta) = \mathbf{M}_{YY}(\theta)$ . Then the maximum likelihood estimator of  $\theta$  can be obtained by minimizing

$$l_n(\theta) = n^{-1} |\mathbf{M}_{YY}(\theta)|^{1/n} \mathbf{Y}' \mathbf{M}_{YY}^{-1}(\theta) \mathbf{Y}. \quad (8.4.5)$$

The estimator of  $\theta$  obtained by minimizing

$$Q_{1n}(\theta) = n^{-1} \mathbf{Y}' \mathbf{M}_{YY}^{-1}(\theta) \mathbf{Y} \quad (8.4.6)$$

is called the least squares estimator or the unconditional least squares estimator. An approximation to the least squares estimator is the estimator that minimizes

$$Q_{2n}(\theta) = n^{-1} \sum_{t=p+1}^n \left[ Y_t + \sum_{j=1}^{t-1} d_j(\theta) Y_{t-j} \right]^2, \quad (8.4.7)$$

where the  $d_j(\theta)$  are defined in Theorem 2.7.2. Observe that  $Q_{2n}(\theta)$  is the average of the squares obtained by truncating the infinite autoregressive representation for  $e_t$ . In Corollary 8.4.1 of this section, we show that the estimators that minimize  $l_n(\theta)$ ,  $Q_{1n}(\theta)$ , and  $Q_{2n}(\theta)$  have the same limiting distribution for stationary invertible time series.

To obtain the partial derivatives associated with the estimation, we express (8.4.1) as

$$e_t(\mathbf{Y}; \theta) = Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} - \sum_{i=1}^q \beta_i e_{t-i}(\mathbf{Y}; \theta). \quad (8.4.8)$$

Differentiating both sides of (8.4.8), we have

$$\begin{aligned} -\frac{\partial e_t(\mathbf{Y}; \theta)}{\partial \alpha_j} &= W_{\alpha_j, t}(\mathbf{Y}; \theta) \\ &= -Y_{t-j} - \sum_{s=1}^q \beta_s W_{\alpha_j, t-s}(\mathbf{Y}; \theta), \quad j = 1, 2, \dots, p, \\ -\frac{\partial e_t(\mathbf{Y}; \theta)}{\partial \beta_i} &= W_{\beta_i, t}(\mathbf{Y}; \theta) \\ &= e_{t-i}(\mathbf{Y}; \theta) - \sum_{s=1}^q \alpha_s W_{\beta_i, t-s}(\mathbf{Y}; \theta), \quad i = 1, 2, \dots, q. \end{aligned} \quad (8.4.9)$$

Using the initial conditions

$$W_{\alpha_j,t}(\mathbf{Y}; \boldsymbol{\theta}) = 0, \quad j = 1, 2, \dots, p,$$

$$W_{\beta_i,t}(\mathbf{Y}; \boldsymbol{\theta}) = 0, \quad i = 1, 2, \dots, q,$$

for  $t \leq p$  and  $e_{t-i}(\mathbf{Y}; \boldsymbol{\theta}) = 0$  for  $t - i \leq p$ , the derivatives of  $Q_{2n}(\boldsymbol{\theta})$  are defined recursively. The  $W_{\alpha_j,t}(\mathbf{Y}; \boldsymbol{\theta})$  and  $W_{\beta_i,t}(\mathbf{Y}; \boldsymbol{\theta})$  are autoregressive moving averages of  $Y_t$ . Therefore, if the roots of (8.4.2) and (8.4.3) associated with the  $\boldsymbol{\theta}$  at which the derivatives are evaluated are less than one in absolute value, the effect of the initial conditions dies out and the derivatives converge to autoregressive moving average time series as  $t$  increases.

The large sample properties of the estimator associated with  $Q_{2n}(\boldsymbol{\theta})$  are given in Theorem 8.4.1. We prove the theorem for iid( $0, \sigma^2$ ) errors, but the result also holds for martingale difference errors satisfying the conditions of Theorem 8.2.1.

**Theorem 8.4.1.** Let the stationary time series  $Y_t$  satisfy (8.4.1) where the  $e_t$  are iid( $0, \sigma^2$ ) random variables. The parameter space  $\Theta$  is such that all roots of (8.4.2) and (8.4.3) are less than one in absolute value, and (8.4.2) and (8.4.3) have no common roots. Let  $\boldsymbol{\theta}^0$  denote the true parameter, which is in the interior of the parameter space  $\Theta$ . Let  $\hat{\boldsymbol{\theta}}_2$  be the value of  $\boldsymbol{\theta}$  in the closure of  $\Theta$  that minimizes  $Q_{2n}(\boldsymbol{\theta})$ . Then

$$[\hat{\boldsymbol{\theta}}'_2, Q_{2n}(\hat{\boldsymbol{\theta}}_2)] \xrightarrow{P} [\boldsymbol{\theta}^{0'}, (\sigma^0)^2]$$

and

$$n^{1/2}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} N[0, \mathbf{V}_{00}^{-1}(\sigma^0)^2], \quad (8.4.10)$$

where

$$\mathbf{V}_{00} = \lim_{t \rightarrow \infty} E\{\mathbf{W}'_{0,t} \mathbf{W}_{0,t}\},$$

$$\begin{aligned} \mathbf{W}_{0,t} &= [W_{\alpha_1,t}(\mathbf{Y}; \boldsymbol{\theta}^0), W_{\alpha_2,t}(\mathbf{Y}; \boldsymbol{\theta}^0), \dots, W_{\alpha_p,t}(\mathbf{Y}; \boldsymbol{\theta}^0), \\ &\quad W_{\beta_1,t}(\mathbf{Y}; \boldsymbol{\theta}^0), W_{\beta_2,t}(\mathbf{Y}; \boldsymbol{\theta}^0), \dots, W_{\beta_q,t}(\mathbf{Y}; \boldsymbol{\theta}^0)], \end{aligned}$$

the elements of  $\mathbf{W}_{0,t}$  are defined in (8.4.9), and  $\mathbf{W}_{0,t}$  is  $\mathbf{W}_{0,t}$  evaluated at  $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ .

**Proof.** Let  $\Theta_\epsilon$  be a compact space such that, for any  $\boldsymbol{\theta} \in \Theta_\epsilon$ , the roots of (8.4.2) are less than or equal to one in absolute value, the roots of (8.4.3) are less than or equal to  $1 - \epsilon$  in absolute value for some  $\epsilon > 0$ , and  $\boldsymbol{\theta}^0$  is in the interior of  $\Theta_\epsilon$ . For any  $\boldsymbol{\theta}$  in  $\Theta_\epsilon$ ,

$$Q_{2n}(\boldsymbol{\theta}) \xrightarrow{P} V\{Z_t(\boldsymbol{\theta})\},$$

where

$$\begin{aligned} Z_t(\theta) &= [\mathcal{B}^p A(\mathcal{B}^{-1}; \alpha)][\mathcal{B}^q B(\mathcal{B}^{-1}; \beta)]^{-1} Y_t \\ &= \frac{[\mathcal{B}^p A(\mathcal{B}^{-1}; \alpha)][\mathcal{B}^q B(\mathcal{B}^{-1}; \beta^0)]}{[\mathcal{B}^q B(\mathcal{B}^{-1}; \beta)][\mathcal{B}^p A(\mathcal{B}^{-1}; \alpha^0)]} e_t, \end{aligned}$$

and the polynomials are defined in (8.4.2) and (8.4.3). The time series  $Z_t(\theta)$  is a stationary autoregressive moving average because the roots of  $B(s; \beta)$  are less than one in absolute value. See Corollary 2.2.2.3, equation (2.7.14), and Theorem 6.3.5. The convergence is uniform by Lemma 5.5.5, because  $Z_t(\theta)$  and its derivatives are infinite moving averages with exponentially declining coefficients.

Now

$$[\mathcal{B}^p A(\mathcal{B}^{-1}; \alpha^0)][\mathcal{B}^q B(\mathcal{B}^{-1}; \beta^0)]^{-1} Y_t = e_t,$$

defines the unique minimum variance prediction error for the predictor of  $Y_t$  based on  $Y_{t-1}, Y_{t-2}, \dots$ . See Section 2.9. Therefore, if  $\theta \neq \theta^0$  and  $\theta \in \Theta_\epsilon$ ,

$$V\{Z_t(\theta)\} > V\{Z_t(\theta^0)\}.$$

If any of the roots of  $B(\mathcal{B}; \beta)$  are equal to one in absolute value,  $Q_{2n}(\theta)$  increases without bound as  $n \rightarrow \infty$ . It follows that the condition (5.5.5) of Lemma 5.5.1 is satisfied for  $\hat{\theta}_2$  defined by the minimum of  $Q_{2n}(\theta)$  over the closure of  $\Theta$ . Hence,  $\hat{\theta}_2$  converges to  $\theta^0$  in probability as  $n \rightarrow \infty$ . Because  $Q_{2n}(\theta)$  is a continuous function of  $\theta$  that converges uniformly to  $V\{Z_t(\theta)\}$  on  $\Theta_\epsilon$ , and  $\hat{\theta}_2 \xrightarrow{P} \theta^0$ , it follows that  $Q_{2n}(\hat{\theta}_2) \xrightarrow{P} (\sigma^0)^2$ .

The first derivatives of  $e_t(\mathbf{Y}; \theta)$  are defined in (8.4.9). The second derivatives can be defined in a similar manner. For example,

$$\frac{\partial^2 e_t(\mathbf{Y}; \theta)}{\partial \alpha_j \partial \alpha_i} = - \frac{\partial W_{\alpha_j, i}(\mathbf{Y}; \theta)}{\partial \alpha_i} = \sum_{s=1}^q \beta_s \frac{\partial W_{\alpha_j, t-s}(\mathbf{Y}; \theta)}{\partial \alpha_i}.$$

Therefore, the second derivatives are also autoregressive moving averages. It follows that the matrix

$$\mathbf{B}_n(\theta) = 0.5 \frac{\partial^2 Q_{2n}(\theta)}{\partial \theta \partial \theta'} = 0.5 n^{-1} \sum_{t=1}^n \left[ e_t(\mathbf{Y}; \theta) \frac{\partial^2 e_t(\mathbf{Y}; \theta)}{\partial \theta \partial \theta'} + \mathbf{W}'_{\theta t} \mathbf{W}_{\theta t} \right]$$

converges uniformly to  $\mathbf{B}(\theta) = \lim_{n \rightarrow \infty} 0.5 E\{\mathbf{W}'_{\theta t} \mathbf{W}_{\theta t}\}$ , which is a continuous function of  $\theta$  on some convex compact neighborhood  $S$  of  $\theta^0$  containing  $\theta^0$  as an interior point.

By the Taylor series arguments used in the proof of Theorem 5.5.1,

$$\hat{\theta}_2 - \theta^0 = \mathbf{B}^{-1}(\theta^0) n^{-1} \sum_{t=1}^n \mathbf{W}'_{\theta t}(\mathbf{Y}; \theta^0) \left[ Y_t + \sum_{j=1}^{t-1} d_j(\theta^0) Y_{t-j} \right] + \mathbf{r}_n,$$

where  $\mathbf{B}(\theta^0)$  is defined in that theorem and  $\mathbf{r}_n$  is of smaller order than  $\hat{\theta}_2 - \theta^0$ .

Now,

$$n^{-1/2} \sum_{t=1}^n \mathbf{W}'_{\theta_t}(\mathbf{Y}; \boldsymbol{\theta}^0) \left\{ e_t - \left[ Y_t + \sum_{j=1}^{t-1} d_j(\boldsymbol{\theta}^0) Y_{t-j} \right] \right\} \xrightarrow{P} 0,$$

because the  $d_j(\boldsymbol{\theta}^0)$  decline exponentially. The vector  $\mathbf{W}'_{\theta_t}(\mathbf{Y}; \boldsymbol{\theta}^0)$  is a function of  $(Y_1, Y_2, \dots, Y_{t-1})$  and hence is independent of  $e_t$ . Following the arguments in the proof of Theorem 5.5.1, we obtain the asymptotic normality of

$$n^{-1/2} \sum_{t=1}^n \mathbf{W}'_{\theta_t}(\mathbf{Y}; \boldsymbol{\theta}^0) e_t,$$

and hence of  $n^{1/2}(\hat{\boldsymbol{\theta}}_2 - \boldsymbol{\theta}^0)$ . Theorem 5.5.1 does not apply directly, because  $Y_t + \sum_{j=1}^{t-1} d_j(\boldsymbol{\theta}) Y_{t-j}$  are not independent and not identically distributed. However, as  $t$  increases, every element of  $\mathbf{W}'_{\theta_t}(\mathbf{Y}; \boldsymbol{\theta}^0)$  converges to an autoregressive moving average time series in the  $e_t$ , and we obtain  $\mathbf{V}_{00}^{-1}(\sigma^0)^2$  as the covariance matrix of the limiting distribution.  $\blacktriangle$

The three estimators defined by (8.4.5), (8.4.6), and (8.4.7) have the same limiting behavior.

**Theorem 8.4.2.** Let  $\hat{\boldsymbol{\theta}}_{ml}$  and  $\hat{\boldsymbol{\theta}}_1$  be the estimators obtained by minimizing (8.4.5) and (8.4.6), respectively. Then under the assumptions of Theorem 8.4.1, the limiting distribution of  $n^{1/2}(\hat{\boldsymbol{\theta}}_{ml} - \boldsymbol{\theta}^0)$  is equal to the limiting distribution of  $n^{1/2}(\hat{\boldsymbol{\theta}}_1 - \boldsymbol{\theta}^0)$  and is that given in (8.4.10).

**Proof.** Let  $\hat{\mathbf{Y}} = \hat{\mathbf{Y}}(\boldsymbol{\theta})$  be an autoregressive moving average  $(p, q)$  with parameter  $\boldsymbol{\theta} \in \Theta_\epsilon$  and uncorrelated  $(0, \sigma^2)$  errors  $\hat{e}_t$ , where  $\Theta_\epsilon$  is defined in the proof of Theorem 8.4.1. Let

$$\hat{\mathbf{v}} = (\hat{Y}_{1-p}, \hat{Y}_{2-p}, \dots, \hat{Y}_0, \hat{e}_{1-p}, \hat{e}_{2-p}, \dots, \hat{e}_0),$$

$\hat{\mathbf{Y}}'_n = (\hat{Y}_1, \hat{Y}_2, \dots, \hat{Y}_n)$ , and  $\hat{\mathbf{e}}'_n = (\hat{e}_1, \hat{e}_2, \dots, \hat{e}_n)$ . Then

$$\begin{bmatrix} \hat{\mathbf{v}} \\ \hat{\mathbf{e}}_n \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ \mathbf{D} \end{bmatrix} \hat{\mathbf{Y}}_n + \begin{pmatrix} \mathbf{I} \\ \mathbf{K} \end{pmatrix} \hat{\mathbf{v}},$$

where  $\mathbf{D} = \mathbf{D}(\boldsymbol{\theta})$  is the  $n \times n$  lower triangular matrix with  $d_{ij} = d_{ij}(\boldsymbol{\theta}) = d_{|i-j|}(\boldsymbol{\theta})$ ,  $d_j(\boldsymbol{\theta})$  is defined in Theorem 2.7.2,  $\mathbf{0}$  is a  $(p+q) \times n$  matrix of zeros,  $\mathbf{I}$  is  $(p+q) \times (p+q)$ ,  $\mathbf{K} = (\mathbf{I}, \mathbf{K}')'$  is a matrix whose elements satisfy

$$\hat{k}_{ij} = \begin{cases} \delta_{ij}, & 1 \leq i \leq p+q, \\ -\sum_{r=1}^q \beta_r \hat{k}_{i-r,j} - \alpha_{i-j-q}, & p+q+1 \leq i \leq n+p+q, \quad 1 \leq j \leq p, \\ -\sum_{r=1}^q \beta_r \hat{k}_{i-r,j}, & p+q+1 \leq i \leq n+p+q, \quad p+1 \leq j \leq p+q, \end{cases}$$

and  $\delta_{ij}$  is the Kronecker delta. Note that the  $\hat{k}_{ij} = \hat{k}_{ij}(\boldsymbol{\theta})$  satisfy the same difference

equation as the  $d_{ij}(\theta)$ , but with different initial values. Galbraith and Galbraith (1974) give the expressions

$$\mathbf{M}_{YY}^{-1} = \mathbf{D}'\mathbf{D} - \mathbf{D}'\mathbf{K}(\Lambda^{-1} + \mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{D} \quad (8.4.11)$$

and

$$|\mathbf{M}_{YY}| = |\Lambda| |\Lambda^{-1} + \mathbf{K}'\mathbf{K}|,$$

where

$$\Lambda = \sigma^{-2} \mathbf{V} \{ (\hat{Y}_{1-p}, \hat{Y}_{2-p}, \dots, \hat{Y}_0, \hat{e}_{1-q}, \hat{e}_{2-q}, \dots, \hat{e}_0)' \}.$$

The dependence of all matrices on  $\theta$  has been suppressed to simplify the expressions. Observe that  $Q_{2n} = n^{-1} \mathbf{Y}'\mathbf{D}'\mathbf{D}\mathbf{Y}$  and

$$Q_{2n} - Q_{1n} = n^{-1} \mathbf{Y}'\mathbf{D}'\mathbf{K}(\Lambda^{-1} + \mathbf{K}'\mathbf{K})^{-1}\mathbf{K}'\mathbf{D}\mathbf{Y}.$$

To show that  $Q_{2n} - Q_{1n}$  converges to zero, we look at the difference on the set  $\Theta_\epsilon$ . Now

$$\begin{aligned} \sup_{\theta \in \Theta_\epsilon} |\mathbf{Y}'\mathbf{D}'\mathbf{K}|^2 &= \sup_{\theta \in \Theta_\epsilon} \sum_{j=1}^{p+q} \left( \sum_{i=1}^n Y_i \sum_{r=0}^{n-i} d_r k_{i+r,j} \right)^2 \\ &\leq \sum_{j=1}^{p+q} \left( \sum_{i=1}^n |Y_i| M \lambda^i \right)^2 \\ &= (p+q) M^2 \left( \sum_{i=1}^n |Y_i| \lambda^i \right)^2 \end{aligned}$$

for some  $M < \infty$  and some  $0 < \lambda < 1$ , because the  $d_r(\theta)$  and the elements of  $K$  decline exponentially. Also,  $\Lambda$  is a positive definite matrix for  $\theta \in \Theta_\epsilon$ , and the determinant is uniformly bounded above and below by positive numbers for all  $n$  and all  $\theta \in \Theta$ . It follows that

$$\sup_{n, \theta \in \Theta_\epsilon} |Q_{2n}(\theta) - Q_{1n}(\theta)| = O(n^{-1}) \quad \text{a.s.}$$

We now investigate the ratio  $l_n(\theta)[Q_1(\theta)]^{-1}$ . Let the  $n \times n$  lower triangular matrix  $\mathbf{T} = \mathbf{T}(\theta)$  define the prediction error made in using the best predictor based on  $(Y_{t-1}, Y_{t-2}, \dots, Y_t)$  to predict  $Y_t$ . See Theorem 2.9.1 and Theorem 2.10.1. Then,

$$E\{(\mathbf{T}\mathbf{Y})(\mathbf{T}\mathbf{Y})'\} = \mathbf{H} = \text{diag}(h_{11}, h_{22}, \dots, h_{nn})$$

and

$$\mathbf{M}_{YY}^{-1} = \mathbf{T}'\mathbf{H}^{-1}\mathbf{T}\sigma^2,$$

where  $h_{ii} \geq \sigma^2$  and  $h_{ii} \rightarrow \sigma^2$  as  $i \rightarrow \infty$ . Also, letting  $\sum_{j=1}^{i-1} b_j Y_{i-j}$  be the best predictor of  $Y_i$ ,

$$\begin{aligned} h_{ii} &= V\left\{Y_i - \sum_{j=1}^{i-1} b_j Y_{i-j}\right\} \leq V\left\{Y_i - \sum_{j=1}^{i-1} d_j Y_{i-j}\right\} \\ &= V\left\{e_i - \sum_{j=i}^{\infty} d_j Y_{i-j}\right\} \leq \sigma^2 + \left[\sum_{j=i}^{\infty} |d_j|\right]^2 V\{Y_i\} \\ &\leq \sigma^2 + M^2(1-\lambda)^{-2} \lambda^{2i} V\{Y_i\} \end{aligned}$$

for some  $M < \infty$  and  $0 < \lambda < 1$ . Thus,

$$|M_{YY}| \leq \prod_{i=1}^n (1 + \overset{*}{M} \lambda^{2i}) \leq \exp\left\{\overset{*}{M} \sum_{i=1}^n \lambda^{2i}\right\} \leq \overset{*}{k}$$

for some  $\overset{*}{M} < \infty$  and  $\overset{*}{k} < \infty$ . It follows that

$$l_n(\theta)[Q_{1n}(\theta)]^{-1} = |M_{YY}(\theta)|^{1/n} < \overset{*}{k}^{1/n} \rightarrow 1.$$

Let the  $(p+q)$ -vector  $\overset{*}{Z} = \overset{*}{Z}(\theta) = K'DY$ , where  $\overset{*}{Z}_j(\theta) = \sum_{i=1}^n Y_i g_{ij}(\theta)$  and

$$g_{ij}(\theta) = \sum_{r=0}^{n-i} d_r k_{i+r,j}.$$

Now the  $g_{ij}(\theta)$ , their first derivatives, and their second derivatives are exponentially declining in  $i$  for  $\theta \in \Theta_\epsilon$ . Also, the first and second derivatives of  $k_{ij}(\theta)$  are exponentially declining in  $i$ . The first and second derivatives of  $\Lambda$  are bounded on  $\Theta_\epsilon$ . Therefore,  $l_n(\theta) - Q_{1n}(\theta)$ ,  $Q_{1n}(\theta) - Q_{2n}(\theta)$ , and  $Q_{2n}(\theta) - l_n(\theta)$  and their first and second derivatives converge uniformly to zero in probability.

By the stated derivative properties of  $K(\theta)$  and  $\Lambda(\theta)$ , the first and second derivatives of  $n^{-1} \log|M_{YY}(\theta)|$  converge uniformly to zero in probability for  $\theta \in \Theta_\epsilon$ . Therefore, the limits of  $l_n(\theta)$ , of  $Q_{1n}(\theta)$ , and of  $Q_{2n}(\theta)$  are the same, and the limits of the first and second derivatives of the three quantities are also the same. It follows that the limiting distributions of the three estimators are the same.  $\blacktriangle$

We derived the limiting distribution of the estimators for the time series with known mean. Because the limiting behavior of sample autocovariances computed with mean adjusted data is the same as that for autocovariances computed with known mean, the results extend to the unknown mean case.

The mean squares and products of the estimated derivatives converge to the mean squares and products of the derivatives based on the true  $\theta^0$ . Therefore, the usual nonlinear least squares estimated covariance matrix can be used for inference.

There are a number of computer programs available that compute the Gaussian

maximum likelihood estimates or approximations to them. It is good practice to use the autoregressive representation technique introduced in Section 8.3 to obtain initial estimates for these programs. By Theorem 2.7.2, we can represent the invertible autoregressive moving average by the infinite autoregressive process

$$Y_t = - \sum_{j=1}^{\infty} d_j Y_{t-j} + e_t, \quad (8.4.12)$$

where the  $d_j$  are defined in Theorem 2.7.2. Hence, by terminating the sum at a convenient finite number, say  $k$ , and estimating the autoregressive parameters,  $d_1, d_2, \dots, d_k$ , we can use the definitions of the  $d_j$  to obtain initial estimates of the autoregressive moving average parameters.

In using a nonlinear method such as the Gauss–Newton procedure, one must beware of certain degeneracies that can occur. For example, consider the autoregressive moving average (1, 1) time series. If we specify zero initial estimates for both parameters, the derivative with respect to  $\alpha_1$  evaluated at  $\alpha_1 = \beta_1 = 0$  is  $-Y_{t-1}$ . Likewise, the derivative with respect to  $\beta_1$  evaluated at  $\alpha_1 = \beta_1 = 0$  is  $Y_{t-1}$ , and the matrix of partial derivatives to be inverted is clearly singular. At second glance, this is not a particularly startling result. It means that a first order autoregressive process with small  $\alpha_1$  behaves very much like a first order moving average process with small  $\beta_1$  and that both behave much like an autoregressive moving average (1, 1) time series where both  $\alpha_1$  and  $\beta_1$  are small. Therefore, one should consider the autoregressive moving average (1, 1) representation only if at least one of the trial parameter values is well away from zero. Because the autoregressive moving average (1, 1) time series with  $\alpha_1 = \beta_1$  is a sequence of uncorrelated random variables, the singularity occurs whenever the initial values are taken to be equal.

In developing our estimation theory for autoregressive moving averages, we have assumed that estimation is carried out for a correctly specified model. In practice, one often is involved in specifying the model at the same time that one is constructing estimates. A number of criteria have been suggested for use in model selection, some developed with time series applications in mind. Because the estimated model parameters are asymptotically normally distributed and because the estimation procedures are closely related to regression, model selection procedures developed for regression models are also applicable to the autoregressive moving average problem. A test based on such statistics was used in Example 8.2.1.

Another selection procedure is based on the variance of regression prediction. See Mallows (1973) and Akaike (1969a). Assume one fits the regression model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{e}, \quad (8.4.13)$$

where  $\mathbf{y}$  is  $n \times 1$ ,  $\mathbf{X}$  is  $n \times k$ ,  $\boldsymbol{\beta}$  is  $k \times 1$ , and  $\mathbf{e} \sim (\mathbf{0}, \sigma^2 \mathbf{I})$ . If one predicts the  $y$ -value for each of the observed  $\mathbf{X}$ , rows of  $\mathbf{X}$ , the average of the  $n$  prediction variances is

$$n^{-1} \sum_{t=1}^n V\{\hat{Y}_t - Y_t\} = n^{-1} \sigma^2 \sum_{t=1}^n [1 + \mathbf{X}_t' (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}_t'] = n^{-1} (n+k) \sigma^2 , \quad (8.4.14)$$

where  $\hat{Y}_t = \mathbf{X}_t \hat{\beta}$ ,  $\hat{\beta} = (\mathbf{X}' \mathbf{X})^{-1} \mathbf{X}' \mathbf{y}$ , and  $k$  is the dimension of  $\mathbf{X}_t$ . Thus, one might choose the model that minimizes an estimator of the mean square prediction error. To estimate (8.4.14), one requires an estimator of  $\sigma^2$ . The estimator of  $\sigma^2$  that is most often suggested in the regression literature is

$$\hat{\sigma}^2 = (n - k_M)^{-1} (\mathbf{y}' \mathbf{y} - \hat{\beta}_M' \mathbf{X}_M' \mathbf{X}_M \hat{\beta}_M) , \quad (8.4.15)$$

where  $\mathbf{X}_M$  is the model of highest dimension and that dimension is  $k_M$ . Thus, the model is chosen to minimize

$$\text{MPE} = n^{-1} (n+k) \hat{\sigma}^2 . \quad (8.4.16)$$

The use of the regression residual mean square for the particular set of regression variables being evaluated as the estimator of  $\sigma^2$  generally leads to the same model selection.

A criterion closely related to the mean square prediction error is the criterion called AIC, introduced by Akaike (1973). This criterion is

$$\text{AIC} = -2 \log L(\hat{\theta}) + 2k , \quad (8.4.17)$$

where  $L(\hat{\theta})$  is the likelihood function evaluated at the maximum likelihood estimator  $\hat{\theta}$ , and  $k$  is the dimension of  $\theta$ . For normal autoregressive moving average models,

$$-2 \log L(\hat{\theta}) = \log |\Sigma_{YY}(\hat{\theta})| + n + n \log 2\pi ,$$

where  $\Sigma_{YY}(\hat{\theta})$  is the covariance matrix of  $(Y_1, Y_2, \dots, Y_n)$  evaluated at  $\theta = \hat{\theta}$ . The determinant of  $\Sigma_{YY}(\theta)$  can be expressed as a product of the prediction error variances (see Theorem 2.9.3). Also, the variances of the prediction errors converge to the error variance  $\sigma^2$  of the process as  $t$  increases. Therefore,  $\log |\Sigma_{YY}(\hat{\theta})|$  is close to  $n \log \hat{\sigma}_m^2$ , where  $\hat{\sigma}_m^2 = \sigma_m^2(\hat{\theta})$  is the maximum likelihood estimator of  $\sigma^2$ . It follows that AIC and MPE are closely related, because

$$n^{-1} \log |\Sigma_{YY}(\hat{\theta})| + 2n^{-1} k \approx \log [(1 + 2n^{-1} k) \hat{\sigma}_m^2] \approx \log \text{MPE} ,$$

when the  $\hat{\sigma}^2$  of MPE is close to  $(n-k)^{-1} n \hat{\sigma}_m^2$ . The AIC criterion is widely used, although it is known that this criterion tends to select rather high order models and will overestimate the true order of a finite order autoregressive model. Because of the tendency to overestimate the order of the model, a number of related criteria have been developed, and considerable research has been conducted on the use of model selection criteria. See Parzen (1974, 1977), Shibata (1976, 1986), Hannan

and Quinn (1979), Hannan (1980), Hannan and Rissanen (1982), Bhansali (1991), Findley (1985), Schwartz (1978), and Findley and Wei (1989).

**Example 8.4.1.** The data of Table 10.B.2 of Appendix 10.B are artificially created data generated by an autoregressive moving average model. As a first step in the analysis, we fit a pure autoregressive model of order 10 by ordinary least squares. The fitted model is

$$\begin{aligned}\hat{Y}_t = & -0.099 + 2.153 Y_{t-1} - 2.032 Y_{t-2} + 0.778 Y_{t-3} \\& (0.104) \quad (0.105) \quad (0.250) \quad (0.332) \\& + 0.072 Y_{t-4} - 0.440 Y_{t-5} + 0.363 Y_{t-6} \\& (0.352) \quad (0.361) \quad (0.366) \\& - 0.100 Y_{t-7} - 0.144 Y_{t-8} + 0.150 Y_{t-9} - 0.082 Y_{t-10}, \\& (0.368) \quad (0.355) \quad (0.268) \quad (0.113)\end{aligned}$$

where the numbers in parentheses are the estimated standard errors of an ordinary regression program. We added seven observations equal to the sample mean to the beginning of the data set and then created ten lags of the data. This is a compromise between the estimators (8.2.5) and (8.2.8). Also, we keep the same number of observations when we fit reduced autoregressive models. There are 97 observations in the regression. The residual mean square with 86 degrees of freedom is 1.020. The fourth order autoregressive process estimated by ordinary least squares based on 97 observations is

$$\hat{Y}_t = -0.086 + 2.170 Y_{t-1} - 2.074 Y_{t-2} + 0.913 Y_{t-3} - 0.209 Y_{t-4} \\(0.102) \quad (0.099) \quad (0.224) \quad (0.225) \quad (0.100)$$

with a residual mean square of 0.996. This model might be judged acceptable if one were restricting oneself to autoregressive processes because the test for the fourth order versus the tenth order gives  $F = 0.63$ , where the distribution of  $F$  can be approximated by Snedecor's  $F$  with 6 and 86 degrees of freedom.

We consider several alternative models estimated by Gaussian maximum likelihood in Table 8.4.1. All calculations were done in SAS/ETS®. The estimate

Table 8.4.1. Comparison of Alternative Models Estimated by Maximum Likelihood

Model	$\hat{\sigma}^2$	AIC	MPE
AR(10)	1.005	301.4	1.115
AR(4)	0.996	294.5	1.046
ARMA(2,1)	1.032	296.9	1.074
ARMA(2,2)	0.991	294.0	1.041
ARMA(2,3)	0.990	295.0	1.050
ARMA(3,1)	1.010	295.8	1.061
ARMA(3,2)	0.992	295.2	1.052

of  $\sigma^2$  given in the table is the maximum likelihood estimator adjusted for degrees of freedom,

$$\hat{\sigma}^2 = (n - r)^{-1} \text{tr}\{\mathbf{Y}'\hat{\mathbf{M}}_{YY}^{-1}\mathbf{Y}\},$$

where  $\mathbf{Y}' = (Y_1, Y_2, \dots, Y_n)$ ,  $\hat{\mathbf{M}}_{YY}$  is the estimate of  $\mathbf{M}_{YY}$ ,  $\Sigma_{YY} = \sigma^2 \mathbf{M}_{YY}$ ,  $\Sigma_{YY}$  is the  $n \times n$  covariance matrix of  $\mathbf{Y}$ , and  $r$  is the total number of parameters estimated. The mean prediction error was calculated as

$$\text{MPE} = n^{-1}(n + r)\hat{\sigma}^2,$$

where  $n = 100$ , and the AIC is defined in (8.4.17).

Several of the models have similar properties. On the basis of the AIC criterion, one would choose the autoregressive moving average (2, 2). For this model, the estimated mean is  $-0.26$  with a standard error of 0.56, and the other estimated parameters are

$$(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2) = (-1.502, 0.843, 0.715, 0.287), \\ (0.109) (0.110) (0.061) (0.059)$$

▲▲

**Example 8.4.2.** As an example of estimation for a process containing several parameters, we fit an autoregressive moving average to the United States monthly unemployment rate from October 1949 to September 1974 (300 observations). The periodogram of this time series was discussed in Section 7.2. Because of the very large contribution to the total sum of squares from the seasonal frequencies, it seems reasonable to treat the time series as if there were a different mean for each month. Therefore, we analyze the deviations from monthly means, which we denote by  $Y_t$ . The fact that the periodogram ordinates close to the seasonal frequencies are large relative to those separated from the seasonal frequencies leads us to expect a seasonal component in a representation for  $Y_t$ .

As the first step in the analysis, we regress  $Y_t$  on  $Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-12}, Y_{t-13}, Y_{t-14}, Y_{t-15}, Y_{t-24}, Y_{t-25}, Y_{t-26}, Y_{t-27}, Y_{t-36}, Y_{t-37}, Y_{t-38}, Y_{t-39}, Y_{t-48}, Y_{t-49}, Y_{t-50}$ , and  $Y_{t-51}$ . Notice that we are anticipating a model of the "component" or "multiplicative" type, so that when we include a variable of lag 12, we also include the next three lags. That is, we are anticipating a model of the form

$$(1 - \theta_1 \mathcal{B} - \theta_2 \mathcal{B}^2 - \theta_3 \mathcal{B}^3)(1 - \theta_4 \mathcal{B}^{12} - \theta_5 \mathcal{B}^{24} - \theta_6 \mathcal{B}^{36} - \theta_7 \mathcal{B}^{48})Y_t \\ = Y_t - \theta_1 Y_{t-1} - \theta_2 Y_{t-2} - \theta_3 Y_{t-3} - \theta_4 Y_{t-12} \\ + \theta_1 \theta_4 Y_{t-13} + \theta_2 \theta_4 Y_{t-14} + \cdots + \theta_3 \theta_7 Y_{t-51} = e_t.$$

In calculating the regression equation we added 36 zeros to the beginning of the data set, lagged  $Y_t$  the requisite number of times, and used the last 285 observations in the regression. This is a compromise between the forms (8.2.5) and (8.2.8) for the estimation of the autoregressive parameters. The regression vectors for the explanatory variables  $Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-12}, Y_{t-13}, Y_{t-14}$ , and  $Y_{t-15}$

**Table 8.4.2. Regression Coefficients Obtained in Preliminary Autoregressive Fit to United States Monthly Unemployment Rate**

Variable	Coefficient	Standard Error of Coefficient
$Y_{t-1}$	1.08	0.061
$Y_{t-2}$	0.06	0.091
$Y_{t-3}$	-0.16	0.063
$Y_{t-12}$	0.14	0.060
$Y_{t-13}$	-0.22	0.086
$Y_{t-14}$	0.00	0.086
$Y_{t-15}$	0.06	0.059
$Y_{t-24}$	0.18	0.053
$Y_{t-25}$	-0.17	0.075
$Y_{t-26}$	-0.09	0.075
$Y_{t-27}$	0.09	0.054
$Y_{t-36}$	0.08	0.054
$Y_{t-37}$	-0.13	0.075
$Y_{t-38}$	0.04	0.075
$Y_{t-39}$	0.00	0.054
$Y_{t-48}$	0.11	0.055
$Y_{t-49}$	-0.01	0.075
$Y_{t-50}$	-0.01	0.075
$Y_{t-51}$	-0.09	0.053

contain all observed values, but the vectors for longer lags contain zeros for some of the initial observations.

The regression coefficients and standard errors are given in Table 8.4.2. The data seem to be consistent with the component type of model. The coefficients for  $Y_{t-12i-1}$  are approximately the negatives of the coefficients on  $Y_{t-12i}$  for  $i = 1, 2, 3$ . Even more consistently, the sum of the three coefficients for  $Y_{t-12i-j}$  for  $j = 1, 2, 3$  is approximately the negative of the coefficient for  $Y_{t-12i}$ ,  $i = 1, 2, 3, 4$ . The individual coefficients show variation about the anticipated relationships, but they give us no reason to reject the component model. The residual mean square for this regression is 0.0634 with 254 degrees of freedom. There are 285 observations in the regression and 19 regression variables. We deduct an additional 12 degrees of freedom for the 12 means previously estimated. We also fit the model with  $Y_{t-1}$ ,  $Y_{t-2}$ , and the corresponding lags of 12 as well as the model with  $Y_{t-1}$ ,  $Y_{t-2}$ ,  $Y_{t-3}$ ,  $Y_{t-4}$ ,  $Y_{t-5}$ , and the corresponding lags of 12. Since the coefficient on  $Y_{t-3}$  is almost twice its standard error, while the coefficients on  $Y_{t-4}$  and  $Y_{t-5}$  were small, we take the third order autoregressive process as our tentative model for the nonseasonal component.

The coefficients for  $Y_{t-12}$ ,  $Y_{t-24}$ ,  $Y_{t-36}$ ,  $Y_{t-48}$  are of the same sign, are of small magnitude relative to one, and are declining slowly. The autoregressive co-

efficients for an autoregressive moving average can display this behavior. Therefore, we consider

$$(1 - \theta_1 \mathcal{B} - \theta_2 \mathcal{B}^2 - \theta_3 \mathcal{B}^3)(1 - \delta \mathcal{B}^{12})Y_t = e_t + \beta e_{t-12}, \quad (8.4.18)$$

as a potential model. On the basis of Theorem 2.7.2, the regression coefficients for lags of multiples of 12 should satisfy, approximately, the relations of Table 8.4.3. Regressing the first column of that table on the second two columns, we obtain the initial estimates  $\hat{\delta} = 0.97$ ,  $\hat{\beta} = -0.83$ . The estimate of  $\beta$  is of fairly large absolute value to be estimated from only four coefficients, but we are only interested in obtaining crude values that can be used as start values for the nonlinear estimation. The maximum likelihood estimate of the parameter vector of model (8.4.18) using (1.08, 0.06, -0.16, 0.97, -0.83) as the initial vector is

$$(\hat{\theta}_1, \hat{\theta}_2, \hat{\theta}_3, \hat{\delta}, \hat{\beta}) = (1.152, -0.002, -0.195, 0.817, -0.651)$$

with estimated standard errors of (0.055, 0.085, 0.053, 0.067, 0.093). The residual mean square error is 0.0626 with 268 degrees of freedom. Since this residual mean square is smaller than that associated with the previous regressions, the hypothesis that the restrictions associated with the autoregressive moving average representation are valid is easily accepted. As a check on model adequacy beyond that given by our initial regressions, we estimated four alternative models with the additional terms  $Y_{t-4}$ ,  $e_{t-1}$ ,  $Y_{t-24}$ ,  $e_{t-24}$ . In no case was the added term significant at the 5% level using the approximate tests based on the regression statistics.

One interpretation of our final model is of some interest. Define  $X_t = Y_t - 1.152Y_{t-1} + 0.002Y_{t-2} + 0.195Y_{t-3}$ . Then  $X_t$  has the autoregressive moving average representation

$$X_t = 0.817X_{t-12} - 0.651e_{t-12} + e_t,$$

where the  $e_t$  are uncorrelated (0, 0.0626) random variables. Now  $X_t$  would have this representation if it were the sum of two independent time series  $X_t = S_t + v_t$ , where  $v_t$  is a sequence of uncorrelated (0, 0.0499) random variables,

$$S_t = 0.817S_{t-12} + u_t,$$

Table 8.4.3. Calculation of Initial Estimates of  $\delta$  and  $\beta$

$j$	Regression Coefficients $-d_j$	Multipliers for $\delta$	Multipliers for $\beta$
1	0.14	1	1
2	0.18	0	-0.14
3	0.08	0	-0.18
4	0.11	0	-0.08

and  $u_i$  is a sequence of  $(0, 0.0059)$  random variables. In such a representation,  $S_i$  can be viewed as the "seasonal component" and the methods of Section 4.5 could be used to construct a filter to estimate  $S_i$ .  $\blacktriangle \blacktriangle$

## 8.5. PREDICTION WITH ESTIMATED PARAMETERS

We now investigate the use of the estimated parameters of autoregressive moving average time series in prediction. Prediction was introduced in Section 2.9 assuming the parameters to be known. The estimators of the parameters of stationary finite order autoregressive invertible moving average time series discussed in this chapter possess errors whose order in probability is  $n^{-1/2}$ . For such time series, the use of the estimated parameters in prediction increases the prediction error by a quantity of  $O_p(n^{-1/2})$ .

Let the time series  $Y_t$  be defined by

$$Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} = \sum_{i=1}^q \beta_i e_{t-i} + e_t, \quad (8.5.1)$$

where the roots of

$$m^p + \sum_{j=1}^p \alpha_j m^{p-j} = 0$$

and of

$$r^q + \sum_{i=1}^q \beta_i r^{q-i} = 0$$

are less than one in absolute value and the  $e_t$  are independent  $(0, \sigma^2)$  random variables with  $E\{e_t^4\} = \eta\sigma^4$ . Let

$$\boldsymbol{\theta}' = (-\alpha_1, -\alpha_2, \dots, -\alpha_p, \beta_1, \beta_2, \dots, \beta_q)$$

denote the vector of parameters of the process.

When  $\boldsymbol{\theta}$  is known, the best one-period-ahead predictor for  $Y_t$  is given by Theorems 2.9.1 and 2.9.3. The predictor for large  $n$  is given in (2.9.25). The large  $n$  predictor obtained by replacing  $\alpha_i$  and  $\beta_i$  in (2.9.25) by  $\hat{\alpha}_i$  and  $\hat{\beta}_i$  is

$$\hat{Y}_{n+1}(Y_1, \dots, Y_n) = - \sum_{j=1}^p \hat{\alpha}_j Y_{n+1-j} + \sum_{i=1}^q \hat{\beta}_i \tilde{e}_{n+1-i}(Y; \hat{\boldsymbol{\theta}}), \quad (8.5.2)$$

where

$$\tilde{e}_t(Y; \boldsymbol{\theta}) = \begin{cases} 0, & t = p - q + 1, p - q + 2, \dots, p \\ Y_t + \sum_{j=1}^p \hat{\alpha}_j Y_{t-j} - \sum_{i=1}^q \hat{\beta}_i \tilde{e}_{t-i}(Y; \boldsymbol{\theta}), & t = p + 1, p + 2, \dots, n. \end{cases} \quad (8.5.3)$$

**Theorem 8.5.1.** Let  $Y_t$  be the time series defined in (8.5.1). Let  $\hat{\boldsymbol{\theta}}$  be an estimator of  $\boldsymbol{\theta} = (-\alpha_1, \dots, -\alpha_p, \beta_1, \dots, \beta_q)'$  such that  $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(n^{-1/2})$ . Then

$$\hat{Y}_{n+1}(Y_1, \dots, Y_n) - \tilde{Y}_{n+1}(Y_1, \dots, Y_n) = O_p(n^{-1/2}),$$

where  $\hat{Y}_{n+1}(Y_1, \dots, Y_n)$  is defined in (2.9.25) and  $\tilde{Y}_{n+1}(Y_1, \dots, Y_n)$  in (8.5.2).

**Proof.** We write

$$\begin{aligned} \tilde{Y}_{n+1}(Y_1, \dots, Y_n) &= - \sum_{j=1}^p \alpha_j Y_{n+1-j} - \sum_{j=1}^p (\hat{\alpha}_j - \alpha_j) Y_{n+1-j} \\ &\quad + \sum_{i=1}^q \beta_i \tilde{e}_{n+1-i}(Y; \boldsymbol{\theta}) \\ &\quad + \sum_{k=1}^{p+q} \frac{\partial \sum_{i=1}^q \beta_i^\top \tilde{e}_{n+1-i}(Y; \boldsymbol{\theta}^\dagger)}{\partial \theta_k} (\theta_k - \hat{\theta}_k), \end{aligned}$$

where  $\boldsymbol{\theta}^\dagger$  is between  $\hat{\boldsymbol{\theta}}$  and  $\boldsymbol{\theta}$  and, for example,

$$\frac{\partial \beta_1^\top \tilde{e}_n(Y; \boldsymbol{\theta}^\dagger)}{\partial \beta_1} = \tilde{e}_n(Y; \boldsymbol{\theta}^\dagger) - \beta_1^\top \left[ \tilde{e}_{n-1}(Y; \boldsymbol{\theta}^\dagger) + \sum_{i=1}^q \beta_i^\top \frac{\partial \tilde{e}_{n-i}(Y; \boldsymbol{\theta}^\dagger)}{\partial \beta_1} \right]$$

and

$$\tilde{e}_n(Y; \boldsymbol{\theta}^\dagger) = Y_n + \sum_{j=1}^p \alpha_j^\top Y_{n-j} - \sum_{i=1}^q \beta_i^\top \tilde{e}_{n-i}(Y; \boldsymbol{\theta}^\dagger).$$

For  $\boldsymbol{\theta}$  such that the roots of the characteristic equations are less than one in absolute value, the derivatives multiplying  $\theta_k - \hat{\theta}_k$  are, except for the initial effects, stationary time series. Since  $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta} = O_p(n^{-1/2})$ , the result follows.  $\blacktriangle$

Theorem 8.5.1 generalizes immediately to predictions  $s$  periods ahead. On the basis of this result, the prediction variance formulas of Section 2.9 can be used as approximations for the predictor  $\tilde{Y}_{n+s}(Y_1, \dots, Y_n)$ .

Additional results are available for the error in predictions for the stationary  $p$ -th order autoregressive process. Let  $Y_t$  be the stationary process satisfying

$$Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} = \alpha_0 + e_t, \quad t = 1, 2, \dots, \quad (8.5.4)$$

where the roots of

$$m^p + \sum_{j=1}^p \alpha_j m^{p-j} = 0 \quad (8.5.5)$$

are less than one in absolute value and the  $e_t$  are independent  $(0, \sigma^2)$  random variables. We can also write

$$\mathbf{Y}_t = \mathbf{A}\mathbf{Y}_{t-1} + \mathbf{e}_t, \quad (8.5.6)$$

where  $\mathbf{Y}_t = (Y_t, Y_{t-1}, \dots, Y_{t-p+1}, 1)'$ ,  $\mathbf{e}_t = (e_t, 0, \dots, 0)'$ , and

$$\mathbf{A} = \begin{bmatrix} -\alpha_1 & -\alpha_2 & -\alpha_3 & \cdots & -\alpha_{p-1} & -\alpha_p & \alpha_0 \\ 1 & 0 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 1 & 0 & \cdots & 0 & 0 & 0 \\ 0 & 0 & 1 & \cdots & 0 & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & 0 & 0 \\ 0 & 0 & 0 & \cdots & 0 & 0 & 1 \end{bmatrix}.$$

The least squares estimator of  $\alpha = (-\alpha_1, -\alpha_2, \dots, -\alpha_p, \alpha_0)'$  is

$$\hat{\alpha} = \left[ \sum_{t=p+1}^n \mathbf{Y}_{t-1} \mathbf{Y}'_{t-1} \right]^{-1} \sum_{t=p+1}^n \mathbf{Y}_{t-1} Y_t, \quad (8.5.7)$$

and the estimator of  $\mathbf{A}$ , denoted by  $\hat{\mathbf{A}}$ , is obtained from  $\mathbf{A}$  by replacing the  $\alpha_i$  with the estimator  $\hat{\alpha}_i$ . Let the least squares predictor of  $\mathbf{Y}_{n+s}$ , based upon (8.5.7), be

$$\hat{\mathbf{Y}}_{n+s} = \hat{\mathbf{A}}^s \mathbf{Y}_n \quad (8.5.8)$$

for  $s \geq 1$ .

**Theorem 8.5.2.** Let the model (8.5.4) hold with the  $e_t$  independent identically distributed random variables with a symmetric distribution function. Let  $Y_1, Y_2, \dots, Y_p$  be symmetrically distributed with finite variance, independent of  $e_{p+1}, e_{p+2}, \dots$ . Assume the distributions are such that

$$E\{|\hat{\mathbf{A}}^s \mathbf{Y}_n|\} < \infty$$

for  $n > p$ . Then, for  $n > p$ ,

$$E\{\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s}\} = \mathbf{0}.$$

**Proof.** The least squares estimators  $\hat{\alpha}_i$ ,  $i = 1, 2, \dots, p$ , can be expressed as functions of

$$\sum_{t=p+1}^n Y_{t-i} Y_{t-j} - n \bar{y}_{(-i)} \bar{y}_{(-j)}, \quad i, j = 0, 1, 2, \dots, p,$$

where  $\bar{y}_{(-i)} = (n-p)^{-1} \sum_{t=p+1}^n Y_{t-i}$ . Also,

$$\hat{\alpha}_0 = \bar{y}_{(0)} + \sum_{i=1}^p \hat{\alpha}_i \bar{y}_{(-i)}.$$

Therefore, the  $\hat{\alpha}_i$ ,  $i = 1, 2, \dots, p$ , are even functions of  $(Y_1, Y_2, \dots, Y_n)$ , and  $\hat{\alpha}_0$  is an odd function of  $(Y_1, Y_2, \dots, Y_n)$ . Note that adding a constant to each  $Y_t$  will change the predictor by the same amount. Therefore, there is no loss of generality in the assumption that  $\alpha_0 = 0$ . The prediction error is

$$\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s} = \sum_{i=0}^{s-1} \mathbf{A}^i \mathbf{e}_{n+s-i} + (\mathbf{A}^s - \hat{\mathbf{A}}^s) \mathbf{Y}_n.$$

The last element of the last column of  $\hat{\mathbf{A}}^s$  is one. All other entries of the last column are multiples of  $\hat{\alpha}_0$ , where the multiples are functions of  $\hat{\alpha}_1, \hat{\alpha}_2, \dots, \hat{\alpha}_p$  and the functions may be zero for small  $s$ . Let  $(Y_1, Y_2, \dots, Y_n) = \mathcal{S}$  be a realization, and let  $(-Y_1, -Y_2, \dots, -Y_n) = \mathcal{S}'$ . Then the value of  $(\mathbf{A}^s - \hat{\mathbf{A}}^s) \mathbf{Y}_n$  for  $\mathcal{S}'$  is the negative of the value for  $\mathcal{S}$ . The result follows because  $E\{\mathbf{e}_i\} = 0$ , and because the distribution of  $(Y_1, Y_2, \dots, Y_n)$  is symmetric. ▲

We now obtain an approximation for the variance of the prediction error, conditional on  $(\mathbf{Y}_n, \mathbf{Y}_{n-1}, \dots, \mathbf{Y}_{n-p+1})$ .

**Theorem 8.5.3.** Let  $Y_t$  be a stationary normal time series satisfying the model (8.5.4) with the roots of (8.5.5) less than one in absolute value. Then the mean square error of the predictor  $\hat{\mathbf{Y}}_{n+s}$  of (8.5.8), conditional on  $\mathbf{Y}_n$ , is

$$\begin{aligned} & E\{(\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s})(\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s})' | \mathbf{Y}_n\} \\ &= \sigma^2 \sum_{j=0}^{s-1} \mathbf{A}' \mathbf{M} \mathbf{A}'^j + n^{-1} \sigma^2 \sum_{j=0}^{s-1} \sum_{k=0}^{s-1} \hat{\mathbf{Y}}_{n+s-j-1}' \Gamma^{-1} \hat{\mathbf{Y}}_{n+s-k-1} \mathbf{A}' \mathbf{M} \mathbf{A}'^k + \mathbf{R}_n, \\ & E\{\|\mathbf{R}_n\| | \mathbf{Y}_n\} = O_p(n^{-3/2}), \\ & E\{\|\mathbf{R}_n\|\} = O(n^{-3/2}), \end{aligned} \tag{8.5.9}$$

where  $\Gamma = E\{\mathbf{Y}_n \mathbf{Y}_n'\}$ ,  $\|\mathbf{R}_n\|^2 = \text{tr } \mathbf{R}_n' \mathbf{R}_n$ ,  $\hat{\mathbf{Y}}_{n+j} = \mathbf{A}' \mathbf{Y}_n$  for  $j = 0, 1, \dots, s$ , and  $\mathbf{M}$  is a matrix with one as the upper left element and zeros elsewhere.

**Proof.** We have

$$\mathbf{Y}_{n+s} = \mathbf{A}^s \mathbf{Y}_n + \sum_{j=0}^{s-1} \mathbf{A}^j \mathbf{e}_{n+s-j},$$

where  $\mathbf{A}^0 = \mathbf{I}$ . Therefore, the error in the predictor is

$$\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s} = \sum_{j=0}^{s-1} \mathbf{A}^j \mathbf{e}_{n+s-j} - (\hat{\mathbf{A}}^s - \mathbf{A}^s) \mathbf{Y}_n.$$

It follows that

$$\begin{aligned} E\{(\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s})(\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s})' | \mathbf{Y}_n\} \\ = \sigma^2 \sum_{j=0}^{s-1} \mathbf{A}^j \mathbf{M} \mathbf{A}'^j + E\{(\hat{\mathbf{A}}^s - \mathbf{A}^s) \mathbf{Y}_n \mathbf{Y}_n' (\hat{\mathbf{A}}^s - \mathbf{A}^s)' | \mathbf{Y}_n\}. \end{aligned}$$

Because  $\hat{\mathbf{A}} = \mathbf{A} + O_p(n^{-1/2})$ , we may expand  $\hat{\mathbf{A}}^s$  in a first order Taylor series about  $\mathbf{A}$  to obtain

$$\hat{\mathbf{A}}^s = \mathbf{A}^s + \sum_{j=0}^{s-1} \mathbf{A}^j (\hat{\mathbf{A}} - \mathbf{A}) \mathbf{A}^{s-j-1} + O_p(n^{-1}).$$

By the normality,  $E\{|\hat{\alpha} - \alpha|^8\} = O(n^{-4})$ , and we obtain

$$\begin{aligned} E\{(\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s})(\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s})' | \mathbf{Y}_n\} &= \sigma^2 \sum_{j=0}^{s-1} \mathbf{A}^j \mathbf{M} \mathbf{A}'^j \\ &+ E\left\{ \left[ \sum_{j=0}^{s-1} \mathbf{A}^j (\hat{\mathbf{A}} - \mathbf{A}) \bar{\mathbf{Y}}_{n+s-j-1} \right] \left[ \sum_{j=0}^{s-1} \mathbf{A}^j (\hat{\mathbf{A}} - \mathbf{A}) \bar{\mathbf{Y}}_{n+s-j-1} \right]' \middle| \mathbf{Y}_n \right\} + \mathbf{R}_{1n}, \end{aligned} \quad (8.5.10)$$

where  $E\{\|\mathbf{R}_{1n}\|\} = O(n^{-3/2})$  and  $E\{\|\mathbf{R}_{1n}\| | \mathbf{Y}_n\} = O_p(n^{-3/2})$ .

Note that for  $r > n$ ,  $q > n$  the upper left element of  $(\hat{\mathbf{A}} - \mathbf{A}) \bar{\mathbf{Y}}_r \bar{\mathbf{Y}}_q' (\hat{\mathbf{A}} - \mathbf{A})'$  is  $(\hat{\alpha} - \alpha)' \bar{\mathbf{Y}}_r \bar{\mathbf{Y}}_q' (\hat{\alpha} - \alpha)$  and all other elements are zero. Therefore, because  $\bar{\mathbf{Y}}_q$  and  $\bar{\mathbf{Y}}_r$  are functions of  $\mathbf{Y}_n$ ,

$$E\{\bar{\mathbf{Y}}_q' (\hat{\alpha} - \alpha) (\hat{\alpha} - \alpha)' \bar{\mathbf{Y}}_r | \mathbf{Y}_n\} = n^{-1} \sigma^2 \bar{\mathbf{Y}}_q' \Gamma^{-1} \bar{\mathbf{Y}}_r + \mathbf{R}_{2n},$$

where

$$\mathbf{R}_{2n} = \bar{\mathbf{Y}}_q' E\{[(\hat{\alpha} - \alpha) (\hat{\alpha} - \alpha)' - n^{-1} \sigma^2 \Gamma^{-1}] | \mathbf{Y}_n\} \bar{\mathbf{Y}}_r.$$

Under our assumptions,

$$E\{[(\hat{\alpha} - \alpha) (\hat{\alpha} - \alpha)' - n^{-1} \Gamma^{-1} \sigma^2] | \mathbf{Y}_n\} = O_p(n^{-3/2})$$

and

$$E\{\|\mathbf{R}_{2n}\|\} = O(n^{-3/2}). \quad \blacktriangle$$

Note that the upper left element of  $\mathbf{A}^j \mathbf{M} \mathbf{A}'^j$  is  $w_j^2$ , where the  $w_j$  satisfy the

difference equation (2.6.4). For the first order process ( $p = 1$ ), the expression for the conditional mean square error of  $\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s}$ , reduces to

$$\sigma^2 \sum_{j=0}^{s-1} \alpha_1^{2j} + n^{-1} \sigma^2 \left[ s^2 \alpha_1^{2s-2} + \left( \sum_{j=0}^{s-1} (-\alpha_1)^j \right)^2 \gamma_Y^{-1}(0) (Y_n - \mu)^2 \right] + R_n,$$

where  $\mu = E\{Y_t\}$  and  $\gamma_Y(0) = E\{(Y_t - \mu)^2\}$ . If the mean is known to be zero and not estimated, we obtain for  $p = 1$

$$E\{(Y_{n+1} - \hat{Y}_{n+1})^2 | \mathbf{Y}_n\} = \sigma^2 + Y_n^2 \gamma_Y^{-1}(0) + R_n.$$

The unconditional mean square error of  $\hat{\mathbf{Y}}_{n+s}$ , is the expected value of (8.5.9).

**Corollary 8.5.3.** Let the assumptions of Theorem 8.5.3 hold. Then

$$\begin{aligned} & E\{(\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s})(\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s})'\} \\ &= \sigma^2 \sum_{j=0}^{s-1} \mathbf{A}' \mathbf{M} \mathbf{A}'^j + n^{-1} \sigma^2 \sum_{j=0}^{s-1} \sum_{k=0}^{s-1} \mathbf{A}' \mathbf{M} \mathbf{A}'^k \\ & \quad \times \text{tr}[(\mathbf{A}'^{s-j-1} \boldsymbol{\Gamma})' (\boldsymbol{\Gamma}^{-1} \mathbf{A}^{s-k-1})] + O(n^{-3/2}). \end{aligned} \quad (8.5.11)$$

**Proof.** Omitted

▲

The approximate mean square error of the predictor given in (8.5.9) can be expressed in scalar form as

$$E\{(Y_{n+s} - \hat{Y}_{n+s})^2 | \mathbf{Y}_n\} \doteq \sigma^2 \left\{ \sum_{j=0}^{s-1} w_j^2 + n^{-1} \sum_{j=0}^{s-1} \sum_{k=0}^{s-1} w_j w_k \hat{\mathbf{Y}}_{n+s-j-1}' \boldsymbol{\Gamma}^{-1} \hat{\mathbf{Y}}_{n+s-k-1} \right\}, \quad (8.5.12)$$

where the  $w_j$  are defined in Theorem 2.6.1. For  $s = 1$ , we have

$$E\{(Y_{n+1} - \hat{Y}_{n+1})^2 | \mathbf{Y}_n\} \doteq \sigma^2 + n^{-1} \sigma^2 \mathbf{Y}_n' \boldsymbol{\Gamma}^{-1} \mathbf{Y}_n$$

and

$$E\{(Y_{n+1} - \hat{Y}_{n+1})^2\} \doteq \sigma^2 [1 + n^{-1}(p + 1)], \quad (8.5.13)$$

an approximation that is often used.

Let

$$\hat{\sigma}^2 = (n - p - 1)^{-1} \sum_{t=1}^n (Y_t - \hat{\boldsymbol{\alpha}}' \mathbf{Y}_{t-1})^2$$

and  $\hat{\boldsymbol{\Gamma}} = n^{-1} \sum_{t=1}^n \mathbf{Y}_{t-1} \mathbf{Y}_{t-1}'$ . An estimator,  $\hat{\mathbf{V}}(\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s})$ , of the mean square

error of prediction may be obtained from (8.5.9) by ignoring the remainder term and replacing  $\mathbf{A}$ ,  $\Gamma$ , and  $\sigma^2$  by their estimators  $\hat{\mathbf{A}}$ ,  $\hat{\Gamma}$ , and  $\hat{\sigma}^2$ , respectively. Let

$$\begin{aligned}\hat{V}(\mathbf{Y}_{n+s} - \hat{\mathbf{Y}}_{n+s}) &= \hat{\sigma}^2 \sum_{j=0}^{s-1} \hat{\mathbf{A}}^j \mathbf{M} \hat{\mathbf{A}}'^j \\ &\quad + n^{-1} \hat{\sigma}^2 \sum_{j=0}^{s-1} \sum_{k=0}^{s-1} [(\hat{\mathbf{A}}^{n+s-j-1} \mathbf{Y}_n)' \hat{\Gamma}^{-1} \hat{\mathbf{A}}^{n+s-k-1} \mathbf{Y}_n] \hat{\mathbf{A}}^j \mathbf{M} \hat{\mathbf{A}}'^k.\end{aligned}\tag{8.5.14}$$

For  $s = 1$ , equation (8.5.14) reduces to the estimated variance of the error in predicting  $\mathbf{Y}_{n+1}$  obtained by treating the problem as a fixed- $X$  regression prediction of  $\mathbf{Y}_{n+1}$  given  $\mathbf{X}_{n+1} = \mathbf{Y}_n$ . Predictions more than one period ahead are nonlinear functions of the estimated parameters.

**Example 8.5.1.** A number of computer programs are available to compute predictions. Procedure ARIMA of SAS/ETS® was used to estimate models in Example 8.4.1. That program computes predictions that are essentially the estimated minimum mean square error predictor. The estimated prediction variance for the  $s$ -period prediction computed by the program is the formula that ignores the estimation error,

$$\hat{V}\{\hat{Y}_{n+s} - Y_{n+s}\} = \hat{\sigma}^2 \sum_{j=0}^{s-1} \hat{v}_j^2,\tag{8.5.15}$$

where  $\hat{v}_j$  are estimators of the coefficients of Theorem 2.7.1. Predictions for the next three periods using the estimated autoregressive moving average (2, 2) of Example 8.4.1 are  $-7.96$ ,  $-5.17$ , and  $-1.14$  for one, two, and three periods, respectively. The standard errors output by the program are  $1.00$ ,  $2.42$ , and  $3.67$  for one, two, and three periods, respectively.

If we use Gaussian maximum likelihood to estimate the fourth order autoregressive process, we obtain the model

$$\begin{aligned}\hat{Y}_t - \hat{\mu} &= 2.188 (Y_{t-1} - \hat{\mu}) - 2.130 (Y_{t-2} - \hat{\mu}) \\ &\quad (0.098) \qquad (0.220) \\ &\quad + 0.977 (Y_{t-3} - \hat{\mu}) - 0.232 (Y_{t-4} - \hat{\mu}), \\ &\quad (0.220) \qquad (0.098)\end{aligned}$$

where  $\hat{\mu} = -0.270$  (0.493). The predictions for the next three periods based on the fourth order autoregressive model are  $-8.11$ ,  $-5.37$ , and  $-1.30$  for one, two, and three periods, respectively. The standard errors that do not contain a component for estimation error, as output by the program, are  $1.00$ ,  $2.40$ , and  $3.58$  for one, two, and three periods, respectively.

To illustrate the effect on the estimated standard errors of the component due to estimating parameters, we use the ordinary least squares estimates of the fourth

Table 8.5.1. Observations for Nonlinear Model to Construct Predictions

$t$	Dependent Variable		Intercept	$X_{t1}$	$X_{t2}$	$X_{t3}$	$X_{t4}$	$X_{t5}$	$X_{t6}$	$X_{t7}$
	$Y_4$	$Y_5$	1	$Y_3$	$Y_2$	$Y_1$	$\bar{y}$	0	0	0
4	$Y_4$		1	$Y_3$	$Y_2$	$Y_1$	$\bar{y}$	0	0	0
5	$Y_5$		1	$Y_4$	$Y_3$	$Y_2$	$Y_1$	0	0	0
$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$	$\vdots$
$n$	$Y_n$		1	$Y_{n-1}$	$Y_{n-2}$	$Y_{n-3}$	$Y_{n-4}$	0	0	0
$n+1$	0		1	$Y_n$	$Y_{n-1}$	$Y_{n-2}$	$Y_{n-3}$	-1	0	0
$n+2$	0		1	0	$Y_n$	$Y_{n-1}$	$Y_{n-2}$	0	-1	0
$n+3$	0		1	0	0	$Y_n$	$Y_{n-1}$	0	0	-1

order autoregression. The estimated model is given in Example 8.4.1. To compute predictions, we use the method of indicator variables and a nonlinear regression program. To construct predictions for three periods, we write the regression model

$$Y_t = \theta_0 + \theta_1 X_{t1} + \theta_2 X_{t2} + \theta_3 X_{t3} + \theta_4 X_{t4} + \theta_5 X_{t5} \\ + (\theta_6 - \theta_1 \theta_5) X_{t6} + (\theta_7 - \theta_1 \theta_6 - \theta_2 \theta_5) X_{t7} + e_t,$$

where the regression variables are defined in Table 8.5.1. The estimator of  $(\theta_5, \theta_6, \theta_7)$  obtained by nonlinear least squares is the predictor of  $(Y_{n+1}, Y_{n+2}, Y_{n+3})$ . The estimated standard errors for  $(\hat{\theta}_5, \hat{\theta}_6, \hat{\theta}_7)$  output by a nonlinear regression program are standard errors for the predictions containing a component for estimation error. The estimated predictions and the estimated standard errors for the nonlinear regression are given in the last column of Table 8.5.2.

The next to last column of Table 8.5.2 contains the standard errors computed with the formula (8.5.15) using the ordinary least squares coefficients. The differences between the standard errors of the last two columns are due to the

Table 8.5.2. Alternative Predictors for Data of Example 8.4.1

Prediction Period	Method and model			
	ML, ARMA(2,2) (8.5.15)	ML, AR(4) (8.5.15)	OLS, AR(4) (8.5.15)	OLS, AR(4) (Nonlinear)
$n+1$	-7.96 (1.00)	-8.11 (1.00)	-8.17 (1.00)	-8.17 (1.01)
$n+2$	-5.17 (2.42)	-5.37 (2.40)	-5.51 (2.38)	-5.51 (2.45)
$n+3$	-1.14 (3.67)	-1.30 (3.58)	-1.48 (3.55)	-1.48 (3.68)

estimated effect of the estimation error. The estimation error makes a larger contribution to the estimated standard errors for the longer predictions.

Table 8.5.2 also contains predictions and standard errors for two models estimated by maximum likelihood. The maximum likelihood estimates of the autoregressive process differ slightly from the ordinary least squares estimates, and hence the predictions differ slightly. The autoregressive moving average predictions differ somewhat from the predictions based on the autoregressive model, but the differences are small relative to the standard errors. The differences among the predictions for the different models are larger than the differences among the estimated standard errors.  $\blacktriangle \blacktriangle$

## 8.6. NONLINEAR PROCESSES

In this section, we consider estimation for time series that fall outside the class of finite parameter autoregressive moving averages. As we have seen, estimation for autoregressive time series has many analogies to linear regression theory. Estimation for moving average time series is a particular nonlinear estimation problem, in that the model can be expressed as an autoregression in which the coefficients satisfy nonlinear restrictions. We now study extensions in which the conditional expectation of  $Y_t$  given  $(Y_{t-1}, \dots, Y_0)$  is a nonlinear function of  $(Y_{t-1}, \dots, Y_0)$ .

Assume

$$Y_t = f(\mathbf{Z}_t; \boldsymbol{\theta}) + e_t, \quad t = 1, 2, \dots, \quad (8.6.1)$$

where  $\mathbf{Z}_t = (Y_{t-1}, Y_{t-2}, \dots, Y_0)$ ,  $\boldsymbol{\theta} = (\theta_1, \theta_2, \dots, \theta_k)'$  is the parameter vector, and  $\{e_t\}$  is a sequence of iid  $(0, \sigma^2)$  random variables. Let  $\hat{\boldsymbol{\theta}}$  be the value of  $\boldsymbol{\theta}$  that minimizes

$$n^{-1} \sum_{t=1}^n [Y_t - f(\mathbf{Z}_t; \boldsymbol{\theta})]^2. \quad (8.6.2)$$

Assume  $f(\mathbf{Z}_t; \boldsymbol{\theta})$  has continuous first and second derivatives with respect to  $\boldsymbol{\theta}$  for all  $t$  on a convex compact set  $S$ ,  $S \subset \Theta$ , where the true value  $\boldsymbol{\theta}^0$  is an interior point of  $S$ , and  $\Theta$  is the parameter space. The model (8.6.1) is the same as the model (5.5.1) of Chapter 5. In Theorem 5.5.1, we gave conditions under which the nonlinear least squares estimator has a normal distribution in the limit. Basically, the derivatives must satisfy some convergence criteria.

In ordinary fitting problems, where  $Y_t$  is known to depend on a vector  $\mathbf{Z}_t$ , it is common practice to approximate an unknown functional relationship with a polynomial in  $\mathbf{Z}_t$ . The fitted function may be used directly as an approximation to the conditional expectation, or it may serve as an intermediate step in developing a nonlinear model with realistic properties. We demonstrate that similar approximations can be used in time series analysis.

**Theorem 8.6.1.** Let  $Y_t$  be a strictly stationary time series with finite  $2(r+1)$  moments for some positive integer  $r$ . Let  $\mathbf{X}_t$  be a vector of polynomials in lags of  $Y_t$ :

$$\mathbf{X}_t = (1, Y_{t-1}, Y_{t-2}, \dots, Y_{t-p}, Y_{t-1}^2, Y_{t-1}Y_{t-2}, \dots, Y_{t-p}').$$

Assume

$$Y_t = f(\mathbf{Z}_t; \boldsymbol{\theta}) + e_t,$$

where  $\{e_t\}$  is a sequence of iid( $0, \sigma^2$ ) random variables,  $\mathbf{Z}_t$  is the vector  $(Y_{t-1}, Y_{t-2}, \dots, Y_0)$ , and  $\boldsymbol{\theta}$  is a vector of parameters. Assume

$$n^{-1} \sum_{t=1}^n (\mathbf{Y}_t, \mathbf{X}_t)' (\mathbf{Y}_t, \mathbf{X}_t) = E\{(\mathbf{Y}_t, \mathbf{X}_t)' (\mathbf{Y}_t, \mathbf{X}_t)\} + O_p(n^{-1/2}), \quad (8.6.3)$$

where  $E\{(\mathbf{Y}_t, \mathbf{X}_t)' (\mathbf{Y}_t, \mathbf{X}_t)\}$  is positive definite. Let

$$\boldsymbol{\xi} = [E\{\mathbf{X}_t' \mathbf{X}_t\}]^{-1} E\{\mathbf{X}_t' f(\mathbf{Z}_t; \boldsymbol{\theta})\} \quad (8.6.4)$$

and

$$\hat{\boldsymbol{\xi}} = \left[ \sum_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \right]^{-1} \sum_{t=1}^n \mathbf{X}_t' \mathbf{Y}_t. \quad (8.6.5)$$

Then  $\hat{\boldsymbol{\xi}} - \boldsymbol{\xi} = O_p(n^{-1/2})$  and

$$\lim_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2 = E\{[f(\mathbf{Z}_t; \boldsymbol{\theta}) - \mathbf{X}_t \boldsymbol{\xi}]^2\} + \sigma^2,$$

where  $\hat{Y}_t = \mathbf{X}_t \hat{\boldsymbol{\xi}}$ .

**Proof.** We have

$$\hat{\boldsymbol{\xi}} - \boldsymbol{\xi} = \left[ \sum_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \right]^{-1} \sum_{t=1}^n \mathbf{X}_t' [f(\mathbf{Z}_t; \boldsymbol{\theta}) - \mathbf{X}_t \boldsymbol{\xi} + e_t].$$

By construction,  $E\{\mathbf{X}_t' [f(\mathbf{Z}_t; \boldsymbol{\theta}) - \mathbf{X}_t \boldsymbol{\xi}]\} = 0$ . Also,

$$n^{-1} \sum_{t=1}^n \mathbf{X}_t' [f(\mathbf{Z}_t; \boldsymbol{\theta}) - \mathbf{X}_t \boldsymbol{\xi}] = O_p(n^{-1/2})$$

by the assumption (8.6.3).

We demonstrate a stronger result in obtaining the order of  $\hat{\boldsymbol{\xi}} - \boldsymbol{\xi}$ . By assumption,  $e_t$  is independent of  $Y_{t-h}$  for  $h \geq 1$ . Therefore, letting  $\mathcal{A}_{t-1}$  denote the sigma-field generated by  $(Y_{t-1}, Y_{t-2}, \dots)$ ,  $\mathbf{X}_t' e_t$  satisfies the conditions of Theorem 5.3.4 because

$$\left[ E \left\{ \sum_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \right\} \right]^{-1} \sum_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \xrightarrow{P} \mathbf{I}.$$

Therefore,

$$\left[ \sum_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \right]^{-1/2} \sum_{t=1}^n \mathbf{X}_t' e_t \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}\sigma^2)$$

and  $\hat{\xi} - \xi = O_p(n^{-1/2})$ .

Because  $\hat{\xi} - \xi = O_p(n^{-1/2})$ ,

$$\begin{aligned} n^{-1} \sum_{t=1}^n (Y_t - \mathbf{X}_t \hat{\xi})^2 &= n^{-1} \sum_{t=1}^n [e_t + f(\mathbf{Z}_t; \theta) - \mathbf{X}_t \xi - \mathbf{X}_t (\hat{\xi} - \xi)]^2 \\ &= n^{-1} \sum_{t=1}^n [e_t + f(\mathbf{Z}_t; \theta) - \mathbf{X}_t \xi]^2 + o_p(n^{-1/2}) \\ &= \sigma^2 + E\{[f(\mathbf{Z}_t; \theta) - \mathbf{X}_t \xi]^2\} + O_p(n^{-1/2}), \end{aligned}$$

where we have used (8.6.3). ▲

It is a direct consequence of Theorem 8.6.1 that polynomial regressions can be used to test for nonlinearity in the conditional expectation.

**Corollary 8.6.1.** Let  $Y_t$  be the strictly stationary autoregressive process satisfying

$$Y_t = \alpha_0 + \sum_{i=1}^p \theta_i Y_{t-i} + e_t,$$

where  $\{e_t\}$  is a sequence of iid( $0, \sigma^2$ ) random variables. Assume  $Y_t$  satisfies the moment conditions of Theorem 8.6.1, including the assumption (8.6.3). Let  $\hat{\xi}$  be defined by (8.6.5), where  $\hat{\xi} = (\hat{\xi}_1, \hat{\xi}_2)$  and  $\hat{\xi}_1$  is the vector of coefficients of  $(1, Y_{t-1}, \dots, Y_{t-p})$ . Let

$$\hat{\sigma}^2 = (n - k)^{-1} \sum_{t=1}^n (Y_t - \hat{Y}_t)^2,$$

where  $k$  is the dimension of  $\mathbf{X}_t$ , and  $\hat{Y}_t = \mathbf{X}_t \hat{\xi}$ . Then

$$\hat{\sigma}^{-1} \hat{\mathbf{A}}_{22}^{-1/2} \hat{\xi}_2 \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}),$$

where

$$\left[ \sum_{t=1}^n \mathbf{X}_t' \mathbf{X}_t \right]^{-1} = \hat{\mathbf{A}} = \begin{bmatrix} \hat{\mathbf{A}}_{11} & \hat{\mathbf{A}}_{12} \\ \hat{\mathbf{A}}_{21} & \hat{\mathbf{A}}_{22} \end{bmatrix}$$

and the partition of  $\hat{\mathbf{A}}$  conforms to the partition of  $\hat{\xi}$ .

**Proof.** Under the assumptions,

$$\xi' = (\alpha_0, \theta_1, \dots, \theta_p, 0, 0, \dots, 0)$$

and  $f(\mathbf{Z}_t; \boldsymbol{\theta}) - \mathbf{X}_t \xi = 0$ . The limiting normal distribution follows from the proof of Theorem 8.6.1.  $\blacktriangle$

It follows from Corollary 8.6.1 that the null distribution of the test statistic

$$F = \hat{\sigma}^{-2} k_2^{-1} \hat{\xi}' \hat{\mathbf{A}}_{22}^{-1} \hat{\xi},$$

where  $k_2$  is the dimension of  $\hat{\xi}$  and  $k$  is the dimension of  $\xi$ , is approximately that of Snedecor's  $F$  with  $k_2$  and  $n - k$  degrees of freedom. The distribution when  $f(\mathbf{Z}_t; \boldsymbol{\theta})$  is a nonlinear function depends on the degree to which  $f(\mathbf{Z}_t; \boldsymbol{\theta})$  is well approximated by a polynomial. If the approximation is good, the test statistic will have good power. Other tests for nonlinearity are discussed by Tong (1990, p. 221), Hinich (1982), and Tsay (1986).

There are several nonlinear models that have received special attention in the time series literature. One is the *threshold autoregressive model* that has been studied extensively by Tong (1983, 1990). A simple first order threshold model is

$$Y_t = \theta_1 Y_{t-1} + \theta_2 \delta(Y_{t-1}, A) Y_{t-1} + e_t, \quad (8.6.6)$$

where

$$\delta(Y_{t-1}, A) = \begin{cases} 1 & \text{if } Y_{t-1} \geq A, \\ 0 & \text{otherwise.} \end{cases} \quad (8.6.7)$$

The indicator function,  $\delta(Y_{t-1}, A)$ , divides the behavior of the process into two regimes. If  $Y_{t-1} \geq A$ , the conditional expected value of  $Y_t$  given  $Y_{t-1}$  is  $(\theta_1 + \theta_2)Y_{t-1}$ . If  $Y_{t-1} < A$ , the conditional expected value of  $Y_t$  given  $Y_{t-1}$  is  $\theta_1 Y_{t-1}$ . Notice that the conditional expected value of  $Y_t$  is not a continuous function of  $Y_{t-1}$  for  $\theta_2 \neq 0$  and  $A \neq 0$ .

Threshold models with more than two regimes and models with more than one lag of  $Y$  entering the equation are easily constructed. Also, the indicator function  $\delta(\cdot)$  can be a function of different lags of  $Y$  and (or) a function of other variables. Practitioners are often interested in whether or not a coefficient has changed at some point in time. In such a case, the indicator function can be a function of time.

Most threshold models do not satisfy the assumptions of our theorems when the parameters specifying the regimes are unknown. The conditional expected value of  $Y_t$  given  $Y_{t-1}$  defined by threshold model (8.6.6) is not a continuous function of  $Y_{t-1}$ , and the conditional expected value is not continuous in  $A$ . Models that are continuous and differentiable in the parameters can be obtained by replacing the function  $\delta(\cdot)$  defining the regimes with a continuous differentiable function.

Candidate functions are continuous differentiable statistical cumulative distribution functions. An example of such a model is

$$Y_t = \theta_1 Y_{t-1} + \theta_2 \hat{\delta}(Y_{t-2}, \kappa) Y_{t-1} + e_t, \quad (8.6.8)$$

$$\hat{\delta}(x, \kappa) = [1 + \exp\{\kappa_1(x - \kappa_2)\}]^{-1}, \quad (8.6.9)$$

where  $\hat{\delta}(x, \kappa)$  is the logistic function. The parameter  $\kappa_1$  can be fixed or can be a parameter to be estimated. The conditional mean function of (8.6.8) is continuous and differentiable in  $(\theta_1, \theta_2, \kappa_1, \kappa_2)$  for  $\kappa_1$  in  $(0, \infty)$ . Tong (1990, p. 107) calls models with  $\delta(\cdot)$  a smooth function, such as (8.6.9), *smoothed threshold autoregressive models*. See Jones (1978) and Ozaki (1980).

**Example 8.6.1.** One of the most analyzed realizations of a time series is the series on the number of lynx trapped in the Mackenzie River district of Canada based on the records of the Hudson Bay Company as compiled by Elton and Nicholson (1942). The data are annual records for the period 1821–1934. The biological interest in the time series arises from the fact that lynx are predators heavily dependent on the snowshoe hare. The first statistical model for the data is that of Moran (1953). Tong (1990, p. 360) contains a description of other analyses and a detailed investigation of the series. We present a few computations heavily influenced by Tong (1990). The observation of analysis is  $\log_{10}$  of the original observations. The sample mean of the time series is 2.904, the smallest value of  $Y_t - \bar{y}$  is  $-1.313$ , and the largest value of  $Y_t - \bar{y}$  is  $0.941$ . The second order autoregressive model estimated by ordinary least squares is

$$\hat{Y}_t = 1.057 + 1.384 Y_{t-1} - 0.747 Y_{t-2}, \quad (8.6.10)$$

$$(0.122) \quad (0.064) \quad (0.064)$$

where  $\hat{\sigma}^2 = 0.053$ , and the numbers in parentheses are the ordinary least squares standard errors.

Let us assume that we are interested in a model based on  $Y_{t-1}$  and  $Y_{t-2}$  and are willing to consider a nonlinear model. For the moment, we ignore information from previously estimated nonlinear models. If we estimate a quadratic model as a first approximation, we obtain

$$\hat{y}_t = 0.079 + 1.366 y_{t-1} - 0.785 y_{t-2} + 0.063 y_{t-1}^2, \quad (8.6.11)$$

$$(0.034) \quad (0.073) \quad (0.068) \quad (0.159)$$

$$+ 0.172 y_{t-1} y_{t-2} - 0.450 y_{t-2}^2, \quad (0.284) \quad (0.172)$$

where  $y_t = Y_t - \bar{y} = Y_t - 2.904$  and  $\hat{\sigma}^2 = 0.0442$ . Also see Cox (1977). The ordinary regression  $F$ -test for the hypothesis that the three coefficients of the quadratic terms are zero is

$$F(3, 106) = (0.0442)^{-1} 0.3661 = 8.28,$$

and the hypothesis of zero values is strongly rejected. The coefficient on  $y_{t-1}^2$  is small, and one might consider the estimated model

$$\hat{y}_t = 0.089 + 1.350 y_{t-1} - 0.772 y_{t-2} + 0.272 y_{t-1}y_{t-2} - 0.497 y_{t-2}^2, \quad (8.6.11)$$

where  $\hat{\sigma}^2 = 0.0438$ .

The estimated conditional expectation of  $Y_t$ , given  $(Y_{t-1}, Y_{t-2})$ , is changed very little from that in (8.6.11) if we replace  $y_{t-2}$  in the last two terms of (8.6.11) by a bounded function of  $y_{t-2}$  that is nearly linear in the interval  $(-1.1, 1.1)$ . Such a function is

$$g(y_{t-2}; \kappa_1, \kappa_2) = [1 + \exp\{\kappa_1(y_{t-2} - \kappa_2)\}]^{-1}. \quad (8.6.12)$$

If  $\kappa_1$  is very small in absolute value, the function is approximately linear. As  $\kappa_1(y_{t-2} - \kappa_2)$  moves from  $-2$  to  $2$ ,  $g(y_{t-2}; \kappa_1, \kappa_2)$  moves from  $0.88$  to  $0.12$ , and as  $\kappa_1(y_{t-2} - \kappa_2)$  moves from  $-3$  to  $3$ ,  $g(y_{t-2}; \kappa_1, \kappa_2)$  moves from  $0.9526$  to  $0.0474$ . If  $\kappa_1$  is very large, the function is essentially a step function. Also, as  $\kappa_1$  increases, the derivative with respect to  $\kappa_1$  approaches the zero function except for values very close to  $\kappa_2$ . The range of  $y_t = Y_t - \bar{y}$  is from  $-1.313$  to  $0.941$ . Thus,  $g(y_{t-2}; -2.5, 0)$  is nearly linear for the range of the data.

The estimated equation obtained by replacing  $y_{t-2}$  with  $g(y_{t-2}; -2.5, 0)$  in the last two terms of (8.6.11) is

$$\begin{aligned} \hat{y}_t = & 0.107 + 1.092 y_{t-1} - 0.205 y_{t-2} + 0.512 g(y_{t-2}; -2.5, 0)y_{t-1} \\ & (0.029) \quad (0.144) \quad (0.140) \quad (0.251) \\ & - 1.110 g(y_{t-2}; -2.5, 0)y_{t-2} \end{aligned} \quad (8.6.13)$$

with  $\hat{\sigma}^2 = 0.0432$ . We note that  $g(y_{t-2}; \kappa_1, \kappa_2)$  converges to the indicator function with jump of height one at the point  $\kappa_2$  as  $\kappa_1 \rightarrow -\infty$ . Therefore, the threshold model is the limit as  $(\kappa_1, \kappa_2) \rightarrow (-\infty, \hat{\kappa}_2^*)$ , where  $\hat{\kappa}_2^*$  is the point dividing the space into two regimes.

We fit the nonlinear model obtained by letting  $\kappa_1$  and  $\kappa_2$  of (8.6.12) be parameters to be estimated. In the estimation we restricted  $\kappa_1$  to be the interval  $[-15, -0.5]$ , and the minimum sum of squares occurred on the boundary  $\kappa_1 = -15$ . The estimated function is

$$\begin{aligned} \hat{y}_t = & 0.119 + 1.241 y_{t-1} - 0.395 y_{t-2} + 0.371 g(y_{t-2}; -15, 0.357)y_{t-1} \\ & (0.033) \quad (0.073) \quad (0.092) \quad (0.143) \\ & - 0.870 g(y_{t-2}; -15, 0.357)y_{t-2}, \end{aligned} \quad (8.6.14)$$

where  $\hat{\sigma}^2 = 0.0420$  and the standard error of  $\hat{\kappa}_2$  is  $0.078$ . The standard errors are computed treating  $\kappa_1$  as known. If the standard errors are computed treating  $\kappa_1$  as

estimated, the standard error of  $\hat{\kappa}_1$  is larger than the estimated value. This reflects the fact that the derivative with respect to  $\kappa_1$  approaches zero as  $\kappa_1 \rightarrow -\infty$ .

The residual sum of squares for the model (8.6.14) is 4.448, while that for (8.6.13) is 4.620. The reduction due to fitting  $\kappa_1$  and  $\kappa_2$  is 0.172, and the F for a test against  $(-2.5, 0)$  is 2.05, which is less than the 5% tabular value of 3.08. If we assumed that we were only estimating  $\kappa_2$ , then the improvement in the fit would be significant. The  $\kappa_1$  of -15 means that about 76% of the shift occurs in an interval of length 0.27 centered at the estimated value of 0.357. The estimated variance of the original time series is 0.314. Therefore, the interval in which most of the estimated shift takes place is about one-half of one standard deviation of the original time series.

A threshold model fitted using  $\hat{\kappa}_2$  to define the regimes is

$$\hat{y}_t = \begin{cases} 0.102 + 1.278 y_{t-1} - 0.456 y_{t-2} & \text{if } y_{t-2} < 0.36, \\ (0.030) \quad (0.071) \quad (0.086) & \\ 0.166 + 1.527 y_{t-1} - 1.239 y_{t-2} & \text{if } y_{t-2} \geq 0.36, \\ (0.152) \quad (0.108) \quad (0.264) & \end{cases}$$

where  $\hat{\sigma}^2 = 0.0445$ . Tong (1990) has given threshold models with smaller residual mean square.

It is reasonable to conclude that the process generating the data is nonlinear, but a number of nonlinear models are consistent with the data. Selection of a final model would rest heavily on subject matter considerations.  $\blacktriangle \blacktriangle$

A second nonlinear model that has been studied extensively is the bilinear model. See, for example, Granger and Andersen (1978), Subba Rao and Gabr (1984), and Tong (1990). The first order model can be written

$$Y_t = \alpha_1 Y_{t-1} + \beta_1 Y_{t-1} e_{t-1} + e_t, \quad (8.6.15)$$

where  $e_t \sim NI(0, \sigma^2)$ . Sufficient conditions for (8.6.15) to be a stationary process are  $\alpha_1^2 + \beta_1^2 \sigma^2 < 1$  or  $E\{|\alpha_1 + \beta_1 e_t|\} < 1$ . Both conditions are intuitive in that they are analogous to a requirement that, on average, the coefficient of  $Y_{t-1}$  be less than one in absolute value. See Tong (1981) and Quinn (1982).

A third class of nonlinear models is the random coefficient autoregressive models. An example of such a model is

$$Y_t = \sum_{i=1}^p (\theta_i + a_{it}) Y_{t-i} + e_t,$$

where

$$\begin{bmatrix} a_t \\ e_t \end{bmatrix} \sim II \left[ \begin{pmatrix} \mathbf{0} \\ \mathbf{0} \end{pmatrix}, \begin{pmatrix} \Sigma_{aa} & \mathbf{0} \\ \mathbf{0} & \sigma_e^2 \end{pmatrix} \right],$$

$a'_t = (a_{1t}, a_{2t}, \dots, a_{pt})$ , and  $e_t$  is independent of  $a_t$ . Nichols and Quinn (1982) is a treatment of such models.

## 8.7. MISSING AND OUTLIER OBSERVATIONS

Estimation with missing observations requires a specification of the mechanism that is responsible for the observations being missing. The most common specification is that observations are *missing at random*. See Rubin (1976) and Little and Rubin (1987). Under the missing-at-random specification, the likelihood can be written as

$$L(\mathbf{Y}_B; \boldsymbol{\theta}) = \int L_C(\mathbf{Y}; \boldsymbol{\theta}) dF(\mathbf{Y}_m), \quad (8.7.1)$$

where  $L_C(\mathbf{Y}; \boldsymbol{\theta})$  is the likelihood for a complete data set, and  $F(\mathbf{Y}_m)$  is the distribution function of the missing observations,

$$\mathbf{Y} = (Y_{[1]}, Y_{[2]}, \dots, Y_{[n-m]}, Y_{[n-m+1]}, \dots, Y_{[n]})' = (\mathbf{Y}_B, \mathbf{Y}_m)',$$

$\mathbf{Y}_B$  is the vector of  $Y_t$  actually observed, and  $\mathbf{Y}_m$  is the vector of missing observations. The subscripts of  $\mathbf{Y}$  are placed in square brackets because they do not necessarily correspond to the index  $t$ .

A simple missing mechanism specifies some probability  $p_m = \eta$  that an observation is missing and that this probability is equal for all observations. A more general model would permit the probability to depend on the index  $t$ , but in order to estimate  $\boldsymbol{\theta}$  by maximizing only  $L(\mathbf{Y}_B; \boldsymbol{\theta})$ , the probability that the observation is missing must not be a function of  $\mathbf{Y}$ .

Assume that  $Y_t$  is a normal stationary autoregressive moving average of order  $(p, q)$  and that  $m$  observations in a sample of size  $n$  are missing at random. Then

$$L(\mathbf{Y}_B; \boldsymbol{\theta}) = (2\pi\sigma^2)^{-0.5(n-m)} |\Sigma_{BB}|^{-0.5} \exp\{-0.5\mathbf{Y}_B' \Sigma_{BB}^{-1} \mathbf{Y}_B\}, \quad (8.7.2)$$

where  $\mathbf{Y}_B$  is the  $(n-m)$ -dimensional vector of observations and  $\Sigma_{BB}$  is the covariance matrix of  $\mathbf{Y}_B$ . While it is relatively easy to write down the likelihood, there has been considerable research on computational methods.

Jones (1980) suggested using the Kalman filter representation of Section 4.6 to construct maximum likelihood estimators for autoregressive moving averages. Let

$$\mathbf{X}_{t|r} = (E\{Y_t | r\}, E\{Y_{t+1} | r\}, \dots, E\{Y_{t+s-1} | r\}),$$

where  $r \leq t$ ,  $s = \max(p, q + 1)$ , and  $E\{Y_{t+j} | r\} = E\{Y_{t+j} | (Y_1, Y_2, \dots, Y_r)\}$ . Let  $X_{t,j|r}$  denote the  $j$ th element of  $\mathbf{X}_{t|r}$ , and abbreviate  $\mathbf{X}_{t|r}$  to  $\mathbf{X}_t$ . The Kalman recursion is initiated at  $t = 1$  with  $\mathbf{X}_0$  of (4.6.43) equal to  $\mathbf{0}$ , and  $\Sigma_{vv00}$  equal to the covariance matrix of  $\mathbf{X}_t$ . Let the sample be complete. Then one can write the log-likelihood as

$$l(\mathbf{Y}; \boldsymbol{\theta}) = -0.5 \left[ n \log(2\pi\sigma^2) + \sum_{i=1}^n \log v_{ii} + \sum_{i=1}^n v_{ii}^{-1} (Y_i - \hat{X}_{i,1|i-1})^2 \right], \quad (8.7.3)$$

where  $\hat{X}_{t,1|t-1}$  is the best linear predictor of  $Y_t$  given  $(Y_{t-1}, \dots, Y_1)$ , and

$$v_{tt} = V\{Y_t - \hat{X}_{t,1|t-1}\}. \quad (8.7.4)$$

We have suppressed the dependence of  $v_{tt}$  and  $\hat{X}_{t,1|t-1}$  on the parameters to simplify the notation. In Example 4.6.4, we illustrated how the usual Kalman recursion formulas can be used in the presence of missing data. Thus, the maximum likelihood estimators can be obtained by maximizing (8.7.3) with respect to the unknown parameters, using the formulas of Section 4.6, where the sum in (8.7.3) is over the elements of  $\mathbf{Y}_B$ .

If there are a modest number of missing values, the method of indicator variables can be used in conjunction with an ordinary maximum likelihood program to construct the maximum likelihood estimators. We illustrate the procedure in Example 8.7.1.

**Example 8.7.1.** The data of Table 8.7.1 are computer generated observations from a second order autoregressive process. We assume that observations 40, 57, and 58 are missing at random. One method of computing the maximum likelihood estimates under this assumption is the method of indicator variables illustrated in the construction of predictions in Example 8.5.1.

Table 8.7.1. Data for Example 8.7.1

1.467	2.722	-1.198	-0.107	4.156
7.181	0.201	-1.803	-4.723	4.224
8.939	-3.239	0.096	-7.357	2.605
5.810	-5.287	-0.113	-4.938	-0.561
-0.267	-5.491	-0.023	1.647	-1.842
-4.053	-3.210	-0.746	6.374	-1.580
-5.427	0.682	-1.072	5.993	0.266
-3.624	4.971	-1.235	3.705	1.526
-0.704	6.265	-0.943	-0.784	3.401
1.256	3.347	1.218	-3.036	3.450
2.162	-0.860	3.173	-3.516	1.361
3.232	-3.697	2.158	-1.510	-2.423
1.076	-4.030	1.859	0.294	-5.100
-2.572	-1.710	2.676	1.637	-5.433
-4.135	-0.217	3.011	0.633	-3.377
-3.836	-1.254	1.540	-1.447	-0.070
-1.185	-0.618	—	-4.060	2.924
0.787	0.087	—	-5.559	3.432
1.446	0.553	5.081	-3.814	3.209
2.425	—	3.900	2.067	1.745

To implement the procedure, we define

$$\begin{aligned} X_{t1} &= \begin{cases} -1 & \text{if } t = 40, \\ 0 & \text{otherwise,} \end{cases} \\ X_{t2} &= \begin{cases} -1 & \text{if } t = 57, \\ 0 & \text{otherwise,} \end{cases} \\ X_{t3} &= \begin{cases} -1 & \text{if } t = 58, \\ 0 & \text{otherwise.} \end{cases} \end{aligned}$$

Then the model

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_{t1} + \beta_2 X_{t2} + \beta_3 X_{t3} + u_t, \\ u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} &= e_t, \end{aligned} \tag{8.7.5}$$

where  $e_t \sim NI(0, \sigma^2)$ , is estimated by maximum likelihood. Using the maximum likelihood option of the AUTOREG procedure of SAS®, we obtain

$$\begin{aligned} (\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\sigma}^2) \\ = (-0.001, -0.040, 1.442, 3.284, -1.377, -0.899, 1.183), \\ (0.207) \quad (0.554) \quad (0.787) \quad (0.786) \quad (0.045) \quad (0.042) \quad (0.173) \end{aligned}$$

where  $\hat{\sigma}^2$  is the estimator adjusted for degrees of freedom. Observe that  $(\hat{\beta}_1, \hat{\beta}_2, \hat{\beta}_3)$  is the best estimator of  $(Y_{40}, Y_{57}, Y_{58})$ , given the other observations and that the parameters are equal to  $(\hat{\alpha}_1, \hat{\alpha}_2)$ . The standard errors for  $Y_{57}$  and  $Y_{58}$  are larger than the standard error of the estimator of  $Y_{40}$  because  $Y_{57}$  and  $Y_{58}$  are adjacent missing observations. ▲▲

In any statistical analysis, it is wise practice to inspect the data at every stage of the analysis for extreme or unusual observations. Such observations are often called outlier observations. In time series analysis, the first step in the analysis is generally an inspection of plots of the original data. Such plots help one define the nature of the data. The plot of the data against time is often sufficient to identify very extreme observations, particularly extreme observations created by errors in recording data.

As with many important concepts, what defines an outlier is difficult to specify. This is because what is unusual requires a complete specification of the statistical model, a specification beyond that common in practice. Thus, in a sample of 100 normal  $(0, 1)$  random variables, a value of  $s_{(n-1)}^{-1}(X_{(n)} - \bar{x}_{(n-1)}) = 5.00$ , where  $X_{(n)}$  is the largest observation,  $s_{(n-1)}$  is the root mean square for the 99 smallest observations, and  $\bar{x}_{(n-1)}$  is the mean of the 99 smallest observations, is unusual. In a sample of 100 from a Cauchy distribution, the same value is less unusual.

Fox (1972) defined two types of unusual observations in time series. The first type is called an *additive outlier*. As an example model for additive outliers, assume that we observe the process

$$Y_t = X_t + \delta_{im} \zeta_m, \tag{8.7.6}$$

where  $X_t$  is a stationary autoregressive moving average,  $\delta_{tm}$  is zero for  $t \neq m$  and one for  $t = m$ , and  $\zeta_m$  is the value of the perturbation at  $t = m$ . Under such a model, one might attempt to estimate  $\zeta_m$  and to remove the effect of the  $\zeta_m$  on the estimates of the autoregressive parameters. One can expand the specification by permitting more than one outlier.

The second type of outlier is called an *innovation outlier*. To illustrate, assume that  $Y_t$  is a stationary autoregressive process satisfying

$$\sum_{i=0}^p \alpha_i Y_{t-i} = e_t, \quad (8.7.7)$$

where  $e_t = a_t + \delta_{tm} \zeta_m$ , and  $(\delta_{tm}, \zeta_m)$  has the properties described following (8.7.6). Then  $\zeta_m$  is called an innovation outlier.

If we assume that we know the point  $m$  at which an outlier may have occurred, treat the unknown  $\zeta_m$  as a parameter to be estimated, and assume  $Y_t$  is a normal process, then one can use likelihood methods to estimate  $\zeta_m$  and to test the hypothesis that  $\zeta_m = 0$ . If the point at which an outlier may have occurred is unknown, then a reasonable procedure is to search over the possible values. To construct the likelihood ratio for every value of  $t$ , or for every possible pair of values, is a large computational task. Therefore, it is common practice to fit a model to the original data and then inspect the residuals to see if any are unusual. The inspection is often done on the basis of plots. In an autoregressive process, an innovation outlier will generally appear as a single residual of large absolute value. This is because only the single observation fails to satisfy the autoregressive model. On the other hand, an additive outlier will affect the residual for the point at which it occurs and will also affect the  $p$  following residuals because they also fail to satisfy the basic autoregressive model. Statistics for testing for outliers were suggested by Fox (1972), and extensions have been considered by several authors, including Chang, Tiao, and Chen (1988) and Tsay (1988).

Assume we have an autoregressive process of order  $p$  with known parameters. The predicted value for  $Y_m$ , given that  $Y_m$  is not observed, is a function of  $\mathbf{y}_m = (Y_{m-p}, \dots, Y_{m-1}, Y_{m+1}, \dots, Y_{m+p})'$ . Therefore, one can construct the linear filter, say  $\mathbf{H}$ , which when applied to the  $\mathbf{y}_m$  gives an estimator of  $Y_m$ . The difference between  $Y_m$  and  $\mathbf{H}\mathbf{y}'_m$  divided by the estimated standard deviation of  $Y_m - \mathbf{H}\mathbf{y}'_m$  provides evidence on whether or not  $Y_m$  is an additive outlier. For a first order process with parameter  $\alpha_1$ ,

$$\hat{\zeta}_m = Y_m - \hat{Y}_m(Y_{m-1}, Y_{m+1}) = Y_m + \alpha_1(1 + \alpha_1^2)^{-1}(Y_{m-1} + Y_{m+1}) \quad (8.7.8)$$

and the variance of the difference is  $\sigma^2(1 + \alpha_1^2)^{-1}$ . In practice, it is necessary to replace the unknown parameters with estimates.

If the parameters of the autoregressive process have been estimated and the residuals calculated, then the residuals at the points  $m, m+1, \dots, m+p$  are affected by the presence of an additive outlier. Thus, an estimate of  $\zeta_m$ , calculated from the residuals, is

$$\xi_m = \left[ \sum_{i=0}^p \hat{\alpha}_i^2 \right]^{-1} \sum_{i=0}^p \hat{\alpha}_i \hat{e}_{m+i}, \quad (8.7.9)$$

where  $\hat{\alpha}_0 = 1$  and  $\hat{e}_{m+i} = \sum_{j=0}^p \hat{\alpha}_j Y_{m+i-j}$ ,  $i = 0, 1, \dots, p$ . A test statistic for the hypothesis that  $\xi_m = 0$  is

$$t_m = \hat{\sigma}^{-1} \left[ \sum_{i=0}^p \hat{\alpha}_i^2 \right]^{1/2} \xi_m, \quad (8.7.10)$$

where  $\hat{\sigma}^2$  is an estimator of  $\sigma^2$ , such as the regression residual mean square. If the process is normal,  $m$  is not determined by the data, and  $\xi_m = 0$ , then  $t_m$  is, approximately, a  $N(0, 1)$  random variable.

An innovation outlier at time  $m$  affects only the residual at time  $m$ . Thus, an estimate of  $\xi_m$  is the residuals at the point  $m$ . Fox (1972) and Chang, Tiao, and Chen (1988) have discussed choosing between the two types of outliers. A simple procedure is to choose the type of outlier that gives the largest absolute value of the estimate of  $\xi_m$ .

In Example 8.7.2, we use model fitting procedures similar to those used in Example 8.7.1 to identify unusual observations, to classify the unusual observations, and to construct estimates of the autoregressive parameters.

**Example 8.7.2.** Table 8.7.2 contains 100 computer generated observations from a second order autoregression. The basic model is

Table 8.7.2. Data for Example 8.7.2

25.267	23.722	19.802	20.893	26.266
27.081	21.201	19.197	16.277	27.653
26.819	17.761	21.096	13.643	25.962
25.110	15.713	20.887	16.062	21.455
20.733	15.509	20.977	25.712	22.665
16.947	17.790	20.254	31.666	17.355
15.573	21.682	19.928	30.120	19.120
17.376	25.971	19.765	25.048	21.462
20.296	27.265	20.057	17.757	24.929
22.256	24.347	22.218	14.199	26.190
23.162	20.140	24.173	14.525	24.300
24.232	17.303	23.158	18.885	19.655
22.076	16.970	22.859	23.230	15.589
18.428	19.290	23.676	25.915	14.117
16.865	20.783	24.011	24.403	15.886
17.164	19.746	22.540	20.349	19.861
19.815	20.382	22.818	15.452	24.060
21.787	21.087	24.914	12.609	25.628
22.446	21.553	26.081	14.620	25.755
23.425	22.063	24.900	22.137	23.786

$$Y_t - \mu + \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) = e_t, \quad (8.7.11)$$

where  $e_t \sim NI(0, \sigma^2)$ . The parameters of the second order autoregressive process estimated by maximum likelihood are

$$(\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\sigma}^2) = (21.150, -1.308, 0.840, 2.20). \quad (8.7.12)$$

Figure 8.7.1 is a plot of the standardized residuals from the second order autoregression, plotted against time. The standardized residuals are

$$\tilde{e}_1 = (Y_1 - \hat{\mu}) \hat{\gamma}_Y^{-1/2}(0),$$

$$\tilde{e}_2 = [(Y_2 - \hat{\mu}) - \hat{\rho}(1)(Y_1 - \hat{\mu})] \{ \hat{\gamma}_Y(0)[1 - \hat{\rho}^2(1)] \}^{-1/2},$$

$$\tilde{e}_t = [Y_t - \hat{\mu} + \hat{\alpha}_1(Y_{t-1} - \hat{\mu}) + \hat{\alpha}_2(Y_{t-2} - \hat{\mu})] \hat{\sigma}^{-1}, \quad t = 2, 3, \dots, n,$$

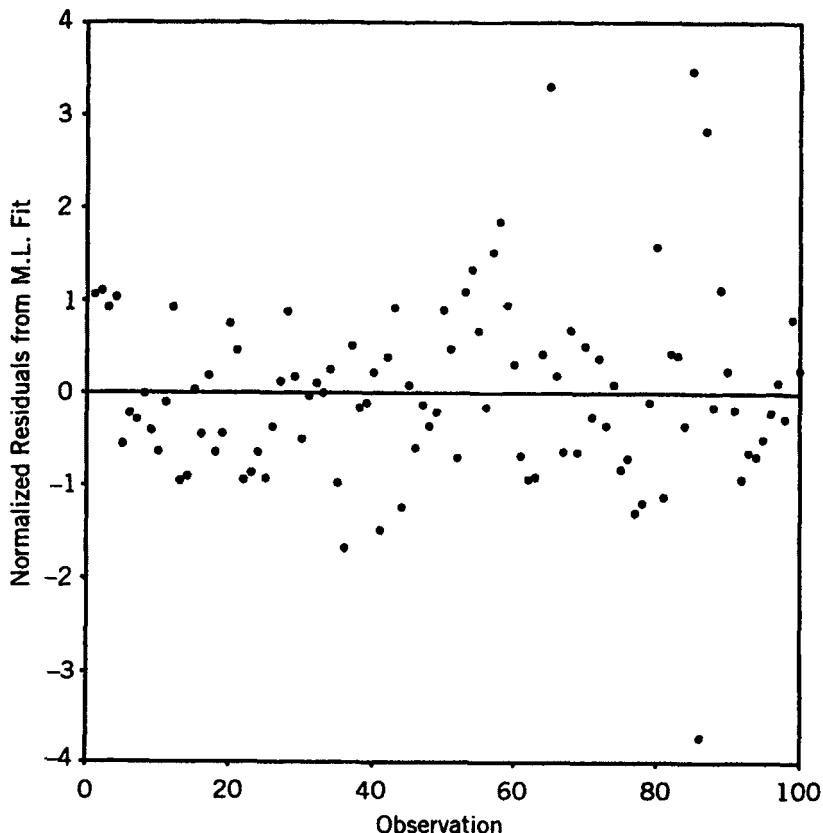


FIGURE 8.7.1. Residuals from second order autoregressive fit.

where  $\hat{\sigma}^2 = 2.20$  is the estimated variance of  $e_t$ . There is a large deviation at  $t = 65$  and three large deviations for  $t = 85, 86$ , and  $87$ . In practice, one would attempt to determine if an error has been made in data recording or if there is a subject matter basis for the unusual observations. We proceed under the assumption that no explanation for the large deviations is available.

The plot suggests that there is an innovation outlier at  $t = 65$ , because there is a large deviation only at that point, the deviations at  $t = 66$  and  $t = 67$  appearing rather ordinary. The deviations associated with  $t = 85, 86$ , and  $87$  indicate that there may be an additive outlier at  $t = 85$ . There is a large positive deviation at  $t = 85$ , followed by a negative deviation at  $t = 86$ , followed by a positive deviation at  $t = 87$ , and these signs match the signs of the autoregressive coefficients. The value of  $t_m$  of (8.7.10) for  $m = 85$  is 5.81, while the standardized residual is  $\hat{\sigma}^{-1}\hat{e}_{85} = 3.48$ . Thus, the data support the presence of an additive outlier over the presence of an innovation outlier at  $t = 85$ .

To estimate the parameters of the model treating  $Y_{65}$  as an innovation outlier and  $Y_{85}$  as an additive outlier, we write a regression model

$$\begin{aligned} Y_t &= \mathbf{X}_t \boldsymbol{\beta} + u_t, \\ u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} &= e_t, \end{aligned} \quad (8.7.13)$$

where  $u_t$  is the zero mean second order autoregressive process. In this representation, an additive outlier can be introduced by creating an indicator variable for the point. For our example, we let  $X_{t1} = 1$  for the parameter  $\mu$  and let

$$\delta_{t,85} = X_{t2} = \begin{cases} 1 & \text{if } t = 85, \\ 0 & \text{otherwise.} \end{cases} \quad (8.7.14)$$

The process of maximizing the likelihood for the model (8.7.13) with  $\mathbf{X}_t = (X_{t1}, X_{t2})$  can be visualized as transforming the data on the basis of the autoregressive model and then finding the  $\boldsymbol{\beta}$  that minimizes the regression residual mean square. Thus, after transformation,

$$\begin{aligned} Y_{85} &= \mu + \zeta_{85} - \alpha_1(Y_{84} - \mu) - \alpha_2(Y_{83} - \mu) + e_{85}, \\ Y_{86} &= \mu - \alpha_1(Y_{85} - \mu - \zeta_{85}) - \alpha_2(Y_{84} - \mu) + e_{86}, \\ Y_{87} &= \mu - \alpha_1(Y_{86} - \mu) - \alpha_2(Y_{85} - \mu - \zeta_{85}) + e_{87}, \end{aligned}$$

where  $(\beta_1, \beta_2) = (\mu, \zeta_{85})$ .

A different  $X$ -variable is required for an innovation outlier in regression model (8.7.13). We require a variable which, after the autoregressive transformation, is one for  $t = m$  and is zero for  $t \neq m$ . If we know  $(\alpha_1, \alpha_2)$ , the variable

$$X_{t3} = \begin{cases} 0 & \text{if } t < 65, \\ 1 & \text{if } t = 65, \\ \alpha_1 X_{t-1,3} + \alpha_2 X_{t-2,3} & \text{if } t > 65 \end{cases} \quad (8.7.15)$$

will satisfy the requirements for  $m = 65$ , because  $X_{t,3}$  satisfies the difference equation. We created the vector  $\mathbf{X}_t = (X_{t,1}, X_{t,2}, X_{t,3})$  using  $(\hat{\alpha}_1, \hat{\alpha}_2) = (-1.308, 0.840)$  from our initial fit of the autoregressive model. The model (8.7.13), estimated by maximum likelihood using the ARIMA procedure in SAS®, gives the estimates

$$(\tilde{\mu}, \tilde{\zeta}_{65}, \tilde{\zeta}_{85}, \tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\sigma}^2) = (20.991, 4.65, 5.27, -1.390, 0.877, 1.157). \\ (0.220) (0.55) (1.07) (0.048) (0.048) (0.170)$$

If we iterate, using  $(\tilde{\alpha}_1, \tilde{\alpha}_2) = (-1.390, 0.877)$  to create a new  $X_{t,3}$  variable, we obtain

$$(\hat{\mu}, \hat{\zeta}_{65}, \hat{\zeta}_{85}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\sigma}^2) = (20.982, 4.64, 5.16, -1.395, 0.880, 1.179). \\ (0.223) (0.55) (1.10) (0.048) (0.047) (0.173) \\ (8.7.16)$$

Note that iteration changed the estimates very little.

The estimates of  $\zeta_{65}$  and  $\zeta_{85}$  are estimated differences between the observed values and the values expected under the autoregressive model, where the deviation is assumed to be an additive outlier for  $t = 85$  and an innovation outlier for  $t = 65$ . Because we identified these points from the residual plots, we should not use ordinary critical values in making tests. Monte Carlo studies by Chang, Tiao, and Chen (1988) suggest that the largest  $t$  in a sample of 100 will exceed 4.00 about 1% of the time when the null model is true. Therefore, both observations are suspect.

Removing the outliers produces a large reduction in the estimated variance, where the estimated variance in (8.7.16) is only about 55% of that in (8.7.12). The changes in the autoregressive coefficients due to removing the outliers are between one and two standard errors, and the estimated mean is changed by about one-half the standard error.

If we use the estimates given in (8.7.12) to create predictions for the next three periods, we obtain

$$(\hat{Y}_{101}, \hat{Y}_{102}, \hat{Y}_{103}) = (20.73, 18.38, 17.89). \\ (1.48) (2.44) (2.76)$$

The corresponding predictions constructed with (8.7.16) are

$$(\hat{Y}_{101}, \hat{Y}_{102}, \hat{Y}_{103}) = (20.72, 18.14, 17.22). \\ (1.08) (1.85) (2.18)$$

The largest effect of removing the outliers is a reduction in the estimated standard error of the prediction errors. One should evaluate these effects when deciding to remove outliers. If the true model is one with long tailed errors, then removing

outliers and using the smaller estimate of  $\sigma^2$  will lead to confidence intervals for predictions that are too narrow.  $\blacktriangle \blacktriangle$

The estimation procedure of Example 8.7.2 assigns zero weight to observations that are identified as outliers. Alternatively, the presence of unusual observations might lead to a modification of the working model and the application of estimation procedures to the modified model. One might specify a long tailed distribution or a mixture distribution for the errors and use maximum likelihood estimation under the specified model. See Kitagawa (1987), Peña and Guttman (1989), and Durbin and Cordero (1993).

Estimators that perform well under a wide range of statistical models are called robust estimators. See Hampel et al. (1986) and Huber (1981) for general discussions of robust procedures. Robust procedures that automatically down-weight large deviation observations have also been considered for time series. See, for example, Martin (1980), Kreiss (1987), and Bruce and Martin (1989).

## 8.8. LONG MEMORY PROCESSES

We introduced long memory time series in Section 2.11. The long memory time series that is called fractionally differenced noise satisfies

$$(1 - \mathcal{B})^d Y_t = \sum_{j=0}^{\infty} \kappa_j(d) Y_{t-j} = e_t \quad (8.8.1)$$

and

$$Y_t = \sum_{j=0}^{\infty} u_j(d) e_{t-j},$$

where

$$\kappa_j(d) = [\Gamma(j+1)\Gamma(-d)]^{-1}\Gamma(j-d),$$

$$u_j(d) = [\Gamma(j+1)\Gamma(d)]^{-1}\Gamma(j+d),$$

and  $-0.5 < d < 0.5$ . As might be expected, the estimation theory associated with such processes differs from that of short memory processes. For example, consider the variance of the sample mean. We have

$$V\{\bar{y}_n\} = n^{-2} \left[ n\gamma_Y(0) + 2 \sum_{h=1}^{n-1} (n-h)\gamma_Y(h) \right].$$

If  $\gamma_Y(h)$  is proportional to  $h^{2d-1}$  and  $d < 0$ , then  $\gamma_Y(h)$  is absolutely summable and  $V\{\bar{y}_n\} = O(n^{-1})$ . If  $0 < d < 0.5$ ,  $\gamma_Y(h)$  is not summable, but Yajima (1988) has shown that

$$\lim_{n \rightarrow \infty} n^{1-2d} V\{\bar{y}_n\} = \sigma^2 \Gamma(1-2d) [\Gamma(d)\Gamma(1-d)d(1+2d)]^{-1}. \quad (8.8.2)$$

The distributions of the sample autocovariances are difficult to obtain, but it is relatively easy to show that the sample autocovariances are consistent.

**Lemma 8.8.1.** Let  $Y_t$  be the infinite moving average time series

$$Y_t = \sum_{r=0}^{\infty} v_r e_{t-r},$$

where  $e_t$  are independent  $(0, \sigma^2)$  random variables with bounded fourth moments and the  $v_r$  are square summable. Let

$$\hat{\gamma}_Y(h) = n^{-1} \sum_{t=1}^{n-h} Y_t Y_{t+h}.$$

Then  $\text{plim}_{n \rightarrow \infty} \hat{\gamma}_Y(h) = \gamma_Y(h)$  for  $h = 0, 1, 2, \dots$

**Proof.** Let

$$\hat{\gamma}_{X_j}(h) = n^{-1} \sum_{t=1}^n X_{tj} X_{t+h,j},$$

where  $X_{tj} = \sum_{r=0}^j v_r e_{t-r}$ . Because  $X_{tj}$  is a finite moving average,

$$\text{plim}_{n \rightarrow \infty} \hat{\gamma}_{X_j}(h) = \gamma_{X_j}(h) = E\{X_{tj} X_{t+h,j}\}$$

for every  $j$ . Also,

$$\text{plim}_{j \rightarrow \infty} \gamma_{X_j}(h) = E\{Y_t Y_{t+h}\}$$

because  $\lim_{j \rightarrow \infty} E\{(X_{tj} - Y_t)^2\} = 0$ . By Chebyshev's inequality,

$$P\{|\hat{\gamma}_Y(h) - \hat{\gamma}_{X_j}(h)| > \epsilon\} \leq (n\epsilon)^{-1} \sum_{t=1}^n E\{|X_{tj} X_{t+h,j} - Y_t Y_{t+h}|^2\}$$

and

$$\text{plim}_{j \rightarrow \infty} \{|\hat{\gamma}_Y(h) - \hat{\gamma}_{X_j}(h)|\} = 0$$

uniformly in  $n$ , because  $\lim_{j \rightarrow \infty} E\{(X_{tj} - Y_t)^2\} = 0$ , uniformly in  $t$ . Therefore, by Lemma 6.3.2,

$$\text{plim}_{n \rightarrow \infty} \hat{\gamma}_Y(h) = \gamma_Y(h). \quad \blacktriangle$$

To introduce estimation for the parameters of long memory processes, assume

we have  $n$  observations from the  $Y_t$  of (8.8.1) and we desire an estimator of  $d$ . Consider the estimator of  $d$  obtained by minimizing

$$Q_n(d) = n^{-1} \sum_{t=1}^n \left[ Y_t + \sum_{j=1}^{t-1} \kappa_j(d) Y_{t-j} \right]^2 \quad (8.8.3)$$

with respect to  $d$ . Note that the infinite autoregressive representation of  $Y_t$  is truncated at length  $t-1$  in (8.8.3). The function (8.8.3) is of the same type as (8.4.7).

**Lemma 8.8.2.** Let the model (8.8.1) hold, let  $d_0 \in \Theta$  be the true parameter, and let  $\Theta = [d_{10}, d_{20}] \subset (0.0, 0.5)$ . Assume the  $e_t$  are independent identically distributed  $(0, \sigma^2)$  random variables with finite fourth moment  $E\{e_t^4\} = \xi\sigma^4$ . Let  $\hat{d}$  be the value of  $d \in \Theta$  that minimizes (8.8.3). Then  $\text{plim}_{n \rightarrow \infty} \hat{d} = d_0$ .

**Proof.** We show that  $\hat{d}$  is a consistent estimator of  $d_0$  by showing that

$$\lim_{n \rightarrow \infty} P\left\{ \inf_{|d - d_0| > \eta} [Q_n(d) - Q_n(d_0)] > 0 \right\} = 1 \quad (8.8.4)$$

for all  $\eta > 0$ . Consider the difference

$$C_n(d) = n^{-1} \sum_{t=1}^n [S_t^2(d) - A_t^2(d)] = n^{-1} \sum_{t=1}^n \left\{ \left[ A_t(d) - \sum_{j=t}^{\infty} \kappa_j(d) Y_{t-j} \right]^2 - A_t^2(d) \right\},$$

where  $Q_n(d) = n^{-1} \sum_{t=1}^n S_t^2(d)$  and

$$[S_t(d), A_t(d)] = \left[ \sum_{j=0}^{t-1} \kappa_j(d) Y_{t-j}, \sum_{j=0}^{\infty} \kappa_j(d) Y_{t-j} \right].$$

Now,

$$\begin{aligned} E\{[A_t(d) - S_t(d)]^2\} &= \lim_{n \rightarrow \infty} E\left\{ \left[ \sum_{r=0}^{\infty} \kappa_{t+r}(d) \sum_{r=0}^n v_r(d_0) e_{t-r} \right]^2 \right\} \\ &\leq \lim_{n \rightarrow \infty} \left[ \sum_{r=0}^{\infty} |\kappa_{t+r}(d)| \right]^2 \sum_{r=0}^n v_r^2(d_0) \sigma^2 \\ &= O(t^{-2d}). \end{aligned}$$

Therefore

$$P\left\{ n^{-1} \sum_{t=1}^n [A_t(d) - S_t(d)]^2 > \epsilon \right\} = O(n^{-2d})$$

and

$$P\left\{\left|n^{-1} \sum_{t=1}^n A_t(d)[A_t(d) - S_t(d)]\right| > \epsilon\right\} = O(n^{-d}).$$

By the mean value theorem,

$$|C_n(d_1) - C_n(d_2)| \leq |d_1 - d_2| \sup_{d \in \Theta} \left| \frac{\partial C_n(d)}{\partial d} \right| \quad (8.8.5)$$

for all  $d_1$  and  $d_2$  in  $\Theta$ , where

$$\begin{aligned} \frac{\partial C_n(d)}{\partial d} &= n^{-1} \sum_{t=1}^n \left[ 2S_t(d) \frac{\partial S_t(d)}{\partial d} - 2A_t(d) \frac{\partial A_t(d)}{\partial d} \right], \\ \frac{\partial A_t(d)}{\partial d} &= \sum_{j=1}^{\infty} \kappa_j(d) \left[ d^{-1} - \sum_{i=2}^j (i-1-d)^{-1} \right] Y_{t-j} \\ &\stackrel{\text{say}}{=} \sum_{j=1}^{\infty} b_j(d) Y_{t-j} = U_t(d), \end{aligned} \quad (8.8.6)$$

and  $|b_j(d)| \leq M j^{-1-d} \log j$  as  $j \rightarrow \infty$ . Since the coefficients  $\kappa_j(d)$  and  $b_j(d)$  are absolutely summable and attain their supremum at some point in the set  $[d_{10}, d_{20}]$ , it follows that the supremum of the derivative on the right of (8.8.5) is  $O_p(1)$ . Thus, by Lemma 5.5.5,  $\text{plim}_{n \rightarrow \infty} C_n(d) = 0$  uniformly in  $d$ . Therefore, we consider the  $d$  associated with the minimum of  $\sum_{t=1}^n A_t^2(d)$ .

We have

$$\text{plim}_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n A_t^2(d) = E\{A_t^2(d)\}$$

because  $A_t(d) = \sum_{i=0}^{\infty} g_i(d) e_{t-i}$ , where  $g_i(d)$  is square summable by Theorem 2.2.3,

$$g_i(d) = \sum_{r=-\infty}^{\infty} \kappa_{i-r}(d) v_r(d_0),$$

and it is understood that  $\kappa_j(d) = 0$  and  $v_j(d_0) = 0$  for  $j < 0$ . Again, by the mean value theorem,

$$\left| n^{-1} \sum_{t=1}^n A_t^2(d_1) - n^{-1} \sum_{t=1}^n A_t^2(d_2) \right| \leq |d_1 - d_2| \sup_{d \in \Theta} \left| 2n^{-1} \sum_{t=1}^n A_t(d) U_t(d) \right|$$

for all  $d_1$  and  $d_2$  in  $\Theta$ , and, by the properties of  $\kappa_j(d)$  and  $b_j(d)$ , the supremum on the right is  $O_p(1)$ .

Now,

$$\begin{aligned} |E\{A_t^2(d_1)\} - E\{A_t^2(d_2)\}| &\leq |d_1 - d_2| \sup_{d \in \Theta} \left| \frac{\partial E\{A_t^2(d)\}}{\partial d} \right| \\ &\leq 2|d_1 - d_2| E\{\sup_{d \in \Theta} |A_t(d) U_t(d)|\} \\ &\leq M |d_1 - d_2| \end{aligned}$$

for some  $M < \infty$  because  $\sup_{d \in \Theta} |\kappa_j(d)|$  and  $\sup_{d \in \Theta} |b_j(d)|$  are absolutely summable; we have used the dominated convergence theorem. Hence, by Lemma 5.5.5,

$$\rho \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n A_i^2(d) = E\{A_i^2(d)\}$$

uniformly in  $d$ . Because  $E\{A_i^2(d)\}$  reaches its minimum at  $d_0$ , the condition (8.8.4) is established.  $\blacktriangle$

### Results on estimators of $d$ and of the parameters of the process

$$(1 - \mathcal{B})^d Y_t = Z_t, \quad (8.8.7)$$

where  $Z_t$  is a stationary autoregressive moving average, have been obtained by a number of authors. Maximum likelihood estimation for the normal distribution model has been studied by Fox and Taqqu (1986), Dahlhaus (1989), Haslett and Raftery (1989), and Beran (1992). Properties of Gaussian maximum likelihood estimators for linear processes have been investigated by Yajima (1985) and Giraitis and Surgailis (1990). Also see Robinson (1994a). The following theorem is due to Dahlhaus (1989). The result was extended to linear processes by Giraitis and Surgailis (1990).

**Theorem 8.8.1.** Let  $Y_t$  satisfy (8.8.7), where  $d \in [d_{10}, d_{20}] \subset (0.0, 0.5)$ , and  $Z_t$  is a stationary normal autoregressive moving average with  $(k-1)$ -dimensional parameter vector  $\theta_2$ , where  $\theta_2$  is in a compact parameter space  $\Theta_2$ . Let  $\theta = (d, \theta_2')$ . Let  $\hat{\theta}$  be the value of  $\theta$  that maximizes the likelihood. Then

$$n^{1/2}(\hat{\theta} - \theta) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{V}_{\theta\theta}),$$

where

$$\begin{aligned} \mathbf{V}_{\theta\theta}^{-1} &= (4\pi)^{-1} \int_{-\pi}^{\pi} \mathbf{g}'_{\theta}(\omega) \mathbf{g}_{\theta}(\omega) d\omega, \\ \mathbf{g}_{\theta}(\omega) &= \left[ \frac{\partial \log f_Y(\omega)}{\partial \theta_1}, \dots, \frac{\partial \log f_Y(\omega)}{\partial \theta_k} \right], \end{aligned}$$

$f_Y(\omega)$  is the spectral density of  $Y_t$ , and the derivatives are evaluated at the true  $\theta$ .

**Proof.** Omitted. See Dahlhaus (1989).  $\blacktriangle$

The inverse of the covariance matrix of Theorem 8.8.1 is also the limit of the expected value of  $n^{-1} \mathbf{h}'_{\theta}(\mathbf{Y}) \mathbf{h}_{\theta}(\mathbf{Y})$ , where

$$\mathbf{h}_{\theta}(\mathbf{Y}) = \left[ \frac{\partial \log L(\mathbf{Y}; \theta)}{\partial \theta_1}, \dots, \frac{\partial \log L(\mathbf{Y}; \theta)}{\partial \theta_k} \right],$$

$\log L(\mathbf{Y}; \boldsymbol{\theta})$  is the log-likelihood function, and the derivatives are evaluated at the true  $\boldsymbol{\theta}$ .

## REFERENCES

- Section 8.1.** Anderson (1971), Gonzalez-Farias (1992), Hasza (1980), Jobson (1972), Koopmans (1942).
- Section 8.2.** Anderson (1959, 1962, 1971), Box and Pierce (1970), Draper and Smith (1981), Hannan (1970), Kendall (1954), Mann and Wald (1943a), Marriott and Pope (1954), Reinsel (1993), Salem (1971), Shaman and Stine (1988).
- Sections 8.3, 8.4.** Anderson and Takemura (1986), Berk (1974), Box, Jenkins and Reinsel (1994), Brockwell and Davis (1991), Cryer and Ledolter (1981), Durbin (1959), Eltinge (1991), Hannan (1979), Kendall and Stuart (1966), Macpherson (1975), Nelson (1974), Pierce (1970a), Sarkar (1990), Walker (1961), Wold (1938).
- Section 8.5.** Davisson (1965), Fuller (1980), Fuller and Hasza (1980, 1981), Hasza (1977), Phillips (1979), Yamamoto (1976).
- Section 8.6.** Granger and Andersen (1978), Priestley (1988), Quinn (1982), Subba Rao and Gabr (1984), Tong (1983, 1990).
- Section 8.7.** Chang, Tiao, and Chen (1988), Fox (1972), Jones (1980), Ljung (1993), Tsay (1988).
- Section 8.8.** Beran (1992), Dahlhaus (1989), Deo (1995), Fox and Taqqu (1986), Yajima (1985).

## EXERCISES

- Using the first 25 observations of Table 7.2.1, estimate  $\rho$  and  $\lambda$  using the equations (8.1.7). Compute the estimated standard errors of  $\hat{\rho}$  and  $\hat{\lambda}$  using the standard regression formulas. Estimate  $\mu$  using the first equation of (8.1.9). Then compute the root of (8.1.10) using  $y_t = Y_t - \hat{\mu}$ . Iterate the computations.
  - Let  $Y_t$  be a stationary time series. Compare the limiting value of the coefficient obtained in the regression of  $Y_t - \hat{r}(1)Y_{t-1}$  on  $Y_{t-1} - \hat{r}(1)Y_{t-2}$  with the limiting value of the regression coefficient of  $Y_{t-2}$  in the multiple regression of  $Y_t$  on  $Y_{t-1}$  and  $Y_{t-2}$ .
  - Compare the variance of  $\bar{y}_n$  for a first order autoregressive process with  $\text{Var}\{\bar{\mu}\}$  and  $\text{Var}\{\hat{\mu}\}$ , where  $\bar{\mu} = \hat{\lambda}(1 - \rho)^{-1}$ , and  $\hat{\lambda}$  and  $\hat{\mu}$  are defined in (8.1.7) and (8.1.9), respectively. In computing  $\text{Var}\{\hat{\mu}\}$  and  $\text{Var}\{\bar{\mu}\}$ , assume that  $\rho$  is known without error. What are the numerical values for  $n = 10$  and  $\rho = 0.7$ ? For  $n = 10$  and  $\rho = 0.9$ ?
  - Assume that 100 observations on a time series gave the following estimates:
- $$\hat{r}(0) = 200, \quad \hat{r}(1) = 0.8, \quad \hat{r}(2) = 0.7, \quad \hat{r}(3) = 0.5.$$

Test the hypothesis that the time series is first order autoregressive against the alternative that it is second order autoregressive.

5. The estimated autocorrelations for a sample of 100 observations on the time series  $\{X_t\}$  are  $\hat{r}(1) = 0.8$ ,  $\hat{r}(2) = 0.5$ , and  $\hat{r}(3) = 0.4$ .

- (a) Assuming that the time series  $\{X_t\}$  is defined by

$$X_t = \beta_1 X_{t-1} + \beta_2 X_{t-2} + e_t,$$

where the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables, estimate  $\beta_1$  and  $\beta_2$ .

- (b) Test the hypothesis that the order of the autoregressive process is two against the alternative that the order is three.

6. Show that for a fixed  $\tilde{\beta}$ ,  $\tilde{\beta} \neq \beta^0$ , the derivative  $W_t(Y; \tilde{\beta})$  of (8.3.9) converges to an autoregressive moving average of order  $(2, 1)$ . Give the parameters.

7. Fit a first order moving average to the first fifty observations in Table 8.3.1. Predict the next observation in the realization. Establish an approximate 95% confidence interval for your prediction.

8. Assume that the  $e_t$  of Theorem 8.3.1 are independent with  $4 + \delta$  moments for some  $\delta > 0$ . Show that  $n^{1/2}[\hat{\sigma}^2 - (\sigma^0)^2]$  converges in distribution to a normal random variable.

9. Prove Theorem 8.3.1 for the estimator constructed with  $\bar{Y}_t - \bar{y}_n$  replacing  $\bar{Y}_t$  in all defining equations.

10. Fit an autoregressive moving average  $(1, 1)$  to the data of Table 8.3.1.

11. The sample variance of the Boone sediment time series discussed in Section 6.4 is 0.580, and the sample variance of the Saylorville sediment time series is 0.337. Let  $X_{1,t}$  and  $X_{2,t}$  represent the deviations from the sample mean of the Boone and Saylorville sediment, respectively. Using the correlations of Table 6.4.1, estimate the following models:

$$X_{2,t} = \theta_1 X_{1,t-1} + \theta_2 X_{1,t-2},$$

$$X_{2,t} = \theta_1 X_{1,t-1} + \theta_2 X_{1,t-2} + \theta_3 X_{2,t-1},$$

$$X_{2,t} = \theta_1 X_{1,t-1} + \theta_2 X_{1,t-2} + \theta_3 X_{1,t-3} + \theta_4 X_{2,t-1} + \theta_5 X_{2,t-2},$$

$$X_{2,t} = \sum_{i=1}^4 \theta_i X_{1,t-i} + \sum_{i=5}^7 \theta_i X_{2,t-i+4}.$$

On the basis of these regressions, suggest a model for predicting Saylorville sediment, given previous observations on Boone and Saylorville sediment.

12. Fit autoregressive equations of order 1 through 7 to the data of Exercise 10 of Chapter 6. Choose a model for these data.

13. A test statistic for partial autocorrelations is given in (8.2.19). An alternative statistic is  $t_i^* = n^{1/2} \hat{\beta}_{ii}$ . Show that

$$t_i^* \xrightarrow{d} N(0, 1)$$

under the assumptions of Theorem 8.2.1 and the assumption that  $\beta_{jj} = 0$  for  $j \geq i$ .

14. A simple symmetric estimator with mean adjustment can be defined as a function of the deviations  $Y_t - \bar{y}_n$ . Show that

$$\tilde{\rho}_{s0} - 1 = -2 \left\{ \sum_{t=2}^n [(Y_t - \bar{y}_n)^2 + (Y_{t-1} - \bar{y}_n)^2] \right\}^{-1} \sum_{t=2}^n (Y_t - Y_{t-1})^2,$$

where  $\tilde{\rho}_{s0} = \{\sum_{t=2}^n [(Y_t - \bar{y}_n)^2 + (Y_{t-1} - \bar{y}_n)^2]\}^{-1} \sum_{t=2}^n (Y_{t-1} - \bar{y}_n)(Y_t - \bar{y}_n)$ .

15. We can express the  $Q(\alpha)$  of (8.2.14) as

$$Q(\alpha) = \sum_{i=1}^n \sum_{j=0}^p \sum_{l=0}^p c_{ijl} Y_{t-i} Y_{t-j} \alpha_i \alpha_j,$$

where  $\alpha_0 = 1$  and  $c_{ijl} = 0$  if  $t-j < 1$  or  $t-l < 1$ . Show that  $c_{011} = 1$  for  $t = 2, 3, \dots, n$  for the weights as defined in (8.2.15). What are  $c_{001}$  and  $c_{111}$ , for  $w_t = 0.5$ ? What are  $c_{001}$  and  $c_{111}$ , for the weights of (8.2.15)?

16. Compute and plot the estimated autocorrelation functions for the AR(4) and ARMA(2, 2) models of Table 8.4.2 for  $h = 0, 1, \dots, 12$ . Does this help explain why it is difficult to choose between these models?

17. Consider a first order stationary autoregressive process  $Y_t = \rho Y_{t-1} + e_t$ , with  $|\rho| < 1$  and  $e_t$  satisfying the conditions of Theorem 8.2.1. Prove Theorem 8.2.2 for the first order autoregressive process by showing that the difference between any two of the estimators of  $\rho$  is  $o_p(n^{-1/2})$ .

18. Consider a first order invertible moving average process  $Y_t = \beta e_{t-1} + e_t$ , where  $|\beta| < 1$ , and the  $e_t$  are iid( $0, \sigma^2$ ) random variables with bounded fourth moments. Let  $\hat{c}_1, \hat{c}_2, \dots, \hat{c}_k$  be the regression coefficients in (8.3.15). Let  $\hat{\beta}_{D,k}$  denote the Durbin's estimator given by  $-[\sum_{i=1}^k \hat{c}_{i-1}^2]^{-1} [\sum_{i=1}^k \hat{c}_{i-1} \hat{c}_i]$ , where  $\hat{c}_0 = -1$ . Find the asymptotic distribution of  $\hat{\beta}_{D,k}$  for a fixed  $k$ . Find the limit as  $k \rightarrow \infty$  of the asymptotic mean and the variance of  $\hat{\beta}_{D,k}$ .

19. Let  $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\sigma}^2)$  be the estimated parameters for the second order auto-

regressive process of Example 8.7.1. Use  $(\hat{\alpha}_1, \hat{\alpha}_2, \hat{\sigma}^2)$  to construct estimates of the first five autocovariances of the process. Then use observations  $(Y_{38}, Y_{39}, Y_{41}, Y_{42})$  and the estimated autocovariances to estimate  $Y_{40}$ . Use  $(Y_{55}, Y_{56}, Y_{59}, Y_{60})$  to estimate  $(Y_{57}, Y_{58})$ .

20. Assume that  $Y_t$  is the first order normal stationary time series

$$Y_t = e_t + \beta e_{t-1}, e_t \sim \text{NI}(0, \sigma^2).$$

Show that the log-likelihood can be written

$$\begin{aligned} \log L(Y; \beta) &= -0.5n \log 2\pi - 0.5 \sum_{t=1}^n \log V_t \\ &\quad - 0.5n \log \sigma^2 - 0.5\sigma^{-2} \sum_{t=1}^n V_t^{-1} Z_t^2, \end{aligned}$$

where  $Z_1 = Y_1$ ,  $V_1 = (1 + \beta^2)$ ,  $Z_t = Y_t - V_{t-1}^{-1} \beta Z_{t-1}$  for  $t = 2, 3, \dots, n$ , and

$$V_t = 1 + \beta^2 - V_{t-1}^{-1} \beta^2, \quad t = 2, 3, \dots, n.$$

21. Prove the following.

**Lemma.** Let  $Y_t$  and  $e_t$  satisfy the assumptions of Theorem 8.2.1. Let  $c_{nt}$ ,  $t = 1, 2, \dots, n$  and  $n = 1, 2, \dots$ , be a triangular array of constants with

$$\sum_{t=1}^n c_{nt}^2 = 1 \quad \text{and} \quad \lim_{n \rightarrow \infty} \sup_{1 \leq t \leq n} c_{nt}^2 = 0.$$

Then

$$\left[ \sigma^{-1} \sum_{t=1}^n c_{nt} e_t, [n\gamma(0)]^{-1/2} \sum_{t=p+1}^n Y_{t-j} e_t \right]' \xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{I})$$

for  $j = 1, 2, \dots, p$ .

## CHAPTER 9

# Regression, Trend, and Seasonality

The majority of the theory presented to this point assumes that the time series under investigation is stationary. Many time series encountered in practice are not stationary. They may fail for any of several reasons:

1. The mean is a function of time, other than the constant function.
2. The variance is a function of time, other than the constant function.
3. The time series is generated by a nonstationary stochastic mechanism.

We consider processes of the third type in Chapter 10. Examples of hypothesized nonstationarity of the first kind occur most frequently in the applied literature, but there are also many examples of heterogeneous variances.

The traditional model for economic time series is

$$Y_t = T_t + S_t + Z_t, \quad (9.0.1)$$

where  $T_t$  is the “trend” component,  $S_t$  is the “seasonal” component, and  $Z_t$  is the “irregular” or “random” component. In our terminology,  $Z_t$  is a stationary time series. Often  $T_t$  is further decomposed into “cyclical” and “long-term” components. Casual inspection of many economic time series leads one to conclude that the mean is not constant through time, and that monthly or quarterly time series display a type of “periodic” behavior wherein peaks and troughs occur at “nearly the same” time each year. However, these two aspects of the time series typically do not exhaust the variability, and therefore the random component is included in the representation.

While the model (9.0.1) is an old one indeed, a precise definition of the components has not evolved. This is not necessarily to be viewed as a weakness of the representation. In fact, the terms acquire meaning only when a procedure is used to estimate them, and the meaning is determined by the procedure. The reader should not be disturbed by this. An example from another area might serve to clarify the issue. The “intelligence quotient” of a person is the person’s score on an I.Q. test, and I.Q. acquires meaning only in the context of the procedure used to

determine it. Although the test may be based on a theory of mental behavior, the I.Q. test score should not be taken to be the only estimator of that attribute of humans we commonly call intelligence.

For a particular economic time series and a particular objective, one model and estimation procedure for trend and seasonality may suffice; for a different time series or a different objective, an alternative specification may be preferred.

We shall now study some of the procedures used to estimate trend and seasonality and (or) to reduce a nonstationary time series to stationarity. Since the mean function of a time series may be a function of other time series or of fixed functions of time, we are led to consider the estimation of regression equations wherein the error is a time series.

## 9.1. GLOBAL LEAST SQUARES

In many situations we are able to specify the mean of a time series to be a simple function of time, often a low order polynomial in  $t$  or trigonometric polynomial in  $t$ . A sample of  $n$  observations can then be represented by

$$\mathbf{y} = \Phi\boldsymbol{\beta} + \mathbf{z}, \quad (9.1.1)$$

where  $\boldsymbol{\beta}' = (\beta_1, \beta_2, \dots, \beta_r)$  is a vector of unknown parameters,

$$\mathbf{y}' = (Y_1, Y_2, \dots, Y_n), \quad \mathbf{z}' = (Z_1, Z_2, \dots, Z_n),$$

$$\Phi = (\varphi_{11}, \varphi_{12}, \dots, \varphi_{1r}),$$

$\varphi_{ii}$ ,  $i = 1, 2, \dots, r$ , are  $n$ -dimensional column vectors, and  $Z_i$  is a zero mean time series. In many situations, we may be willing to assume that  $Z_i$  is stationary, but we also consider estimation under weaker assumptions.

The elements of  $\varphi_{ii}$ , say  $\varphi_{ii}$ , are functions of time. For example, we might have  $\varphi_{11} = 1$ ,  $\varphi_{12} = t$ ,  $\varphi_{13} = t^2$ . The elements of  $\varphi_{ii}$  may also be random functions of time, for example, a stationary time series. In the random case, we shall assume that  $\mathbf{z}$  is independent of  $\Phi$  and investigate the behavior of the estimators conditional on a particular realization of  $\varphi_{ii}$ . Thus, in this section, all  $\varphi_{ii}$  will be treated as fixed functions of time. Notice that  $\mathbf{y}$ ,  $\Phi$ ,  $\varphi_{ii}$ , and  $\mathbf{z}$  might properly be subscripted by  $n$ . To simplify the notation, we have omitted the subscript.

The simple least squares estimator of  $\boldsymbol{\beta}$  is

$$\hat{\boldsymbol{\beta}}_S = (\Phi'\Phi)^{-1}\Phi'\mathbf{y}. \quad (9.1.2)$$

In (9.1.2) and throughout this section, we assume  $\Phi'\Phi$  is nonsingular. Assume that the time series is such that the matrix  $\mathbf{V}_{zz} = E\{\mathbf{z}\mathbf{z}'\}$  is nonsingular. Then the generalized least squares (best linear unbiased) estimator of  $\boldsymbol{\beta}$  is

$$\hat{\boldsymbol{\beta}}_G = [\Phi'\mathbf{V}_{zz}^{-1}\Phi]^{-1}\Phi'\mathbf{V}_{zz}^{-1}\mathbf{y}. \quad (9.1.3)$$

The covariance matrix of  $\hat{\beta}_s$  is

$$E\{(\hat{\beta}_s - \beta)(\hat{\beta}_s - \beta)'\} = (\Phi'\Phi)^{-1}\Phi'V_{zz}\Phi(\Phi'\Phi)^{-1}, \quad (9.1.4)$$

while that of  $\hat{\beta}_G$  is

$$E\{(\hat{\beta}_G - \beta)(\hat{\beta}_G - \beta)'\} = (\Phi'V_{zz}^{-1}\Phi)^{-1}.$$

It is well known that  $\hat{\beta}_G$  is superior to  $\hat{\beta}_s$  in that the variance of any linear contrast  $\lambda'\hat{\beta}_G$  is no larger than the variance of the corresponding linear contrast  $\lambda'\hat{\beta}_s$ . However, the construction of  $\hat{\beta}_G$  requires knowledge of  $V_{zz}$ , and generally  $V_{zz}$  is not known. In fact, one may wish to estimate the mean function of  $Y_t$  prior to investigating the covariance structure of the error time series. Therefore, the properties of the simple least squares estimator are of interest.

To investigate the large sample behavior of least squares estimators, let

$$\begin{aligned} D_n &= \text{diag}\{d_{n11}, d_{n22}, \dots, d_{nnr}\} \\ &\stackrel{\text{def}}{=} \text{diag}\{(\varphi_1', \varphi_1)^{1/2}, (\varphi_2', \varphi_2)^{1/2}, \dots, (\varphi_r', \varphi_r)^{1/2}\}, \end{aligned}$$

where  $\varphi_i$  is the  $i$ th column of the matrix  $\Phi$ . The least squares estimator is consistent for  $\beta$  under mild assumptions.

**Proposition 9.1.1.** Let the model (9.1.1) hold where  $Z_t$  is a zero mean time series and  $\varphi_{ti}$ ,  $i = 1, 2, \dots, r$ ,  $t = 1, 2, \dots$ , are fixed functions of time. Assume

$$\lim_{n \rightarrow \infty} d_{nii} = \infty, \quad i = 1, 2, \dots, r, \quad (9.1.5)$$

$$\lim_{n \rightarrow \infty} D_n^{-1}\Phi'\Phi D_n^{-1} = A_0, \quad (9.1.6)$$

$$\lim_{n \rightarrow \infty} D_n^{-1}\Phi'V_{zz}\Phi D_n^{-1} = B, \quad (9.1.7)$$

where  $A_0$  is nonsingular. Then  $\hat{\beta}_{sj} - \beta_{sj}^0 = O_p(d_{njj}^{-1})$ , where  $\hat{\beta}_{sj}$  is the  $j$ th element of  $\hat{\beta}_s$  and  $\beta_{sj}^0$  is the true parameter value.

**Proof.** The covariance matrix of the least squares estimator is given in (9.1.4), and the variance of the normalized estimator is

$$V\{D_n(\hat{\beta}_s - \beta)\} = D_n(\Phi'\Phi)^{-1}D_n D_n^{-1}\Phi'V_{zz}\Phi D_n^{-1}D_n(\Phi'\Phi)^{-1}D_n.$$

It follows that  $D_n(\hat{\beta}_s - \beta^0) = O_p(1)$ , because

$$\lim_{n \rightarrow \infty} V\{D_n(\hat{\beta}_s - \beta^0)\} = A_0^{-1}BA_0^{-1}. \quad \blacktriangle$$

Under stronger assumptions, the limiting distribution of the least squares estimator can be obtained. Assume that the real valued functions  $\varphi_{ti}$  of  $t$  satisfy

$$\lim_{n \rightarrow \infty} \sup_{1 \leq i \leq n} \left( \sum_{j=1}^n \varphi_{ji}^2 \right)^{-1} \varphi_{ii}^2 = 0, \quad i = 1, 2, \dots, r, \quad (9.1.8)$$

and

$$\lim_{n \rightarrow \infty} d_{nii}^{-1} d_{njj}^{-1} \sum_{t=1}^{n-h} \varphi_{it} \varphi_{i+h,j} = a_{hij} = a_{-h,ij}, \quad (9.1.9)$$

$h = 0, 1, 2, \dots, i, j = 1, 2, \dots, r$ , where  $\mathbf{A}_0$  is nonsingular and  $\mathbf{A}_h$  is the matrix with typical element  $a_{hij}$ . The assumptions on the  $\varphi_{it}$  are quite modest and, for example, are satisfied by polynomial, trigonometric polynomial, and logarithmic functions of time. The following theorem demonstrates that the limiting distribution of the vector of standardized estimators is normal for a wide class of stationary time series.

**Theorem 9.1.1.** Let  $Z_t$  be a stationary time series defined by

$$Z_t = \sum_{j=0}^{\infty} u_j e_{t-j}, \quad t = 0, \pm 1, \pm 2, \dots,$$

where  $\{u_j\}$  is absolutely summable and the  $e_t$  are independent  $(0, \sigma^2)$  random variables with distribution functions  $F_t(e)$  such that

$$\lim_{\delta \rightarrow \infty} \sup_{i=1,2,\dots} \int_{|e| > \delta} e^2 dF_i(e) = 0.$$

Let the  $\varphi_{it}$ ,  $i = 1, 2, \dots, r$ ,  $t = 1, 2, \dots$ , be fixed functions of time satisfying the assumptions (9.1.8) and (9.1.9). Then

$$\mathbf{D}_n(\hat{\beta}_S - \beta) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{A}_0^{-1} \mathbf{B} \mathbf{A}_0^{-1}), \quad (9.1.10)$$

where  $\mathbf{A}_0$  is the matrix with elements  $a_{0ij}$  defined in (9.1.9),  $\mathbf{B}$  is defined in (9.1.7), and  $\hat{\beta}_S$  is defined in (9.1.2). The covariance matrix is nonsingular if  $\mathbf{B}$  of (9.1.7) is nonsingular.

**Proof.** Consider the linear combination  $\sum_{t=1}^r c_t Z_t$ , where

$$c_t = \sum_{i=1}^r \lambda_i d_{nii}^{-1} \varphi_{it}$$

and the  $\lambda_i$  are arbitrary real numbers. Now, by our assumptions,

$$\begin{aligned} \lim_{n \rightarrow \infty} \sum_{t=1}^{n-h} c_t c_{t+h} &= \lim_{n \rightarrow \infty} \sum_{i=1}^{n-h} \sum_{t=1}^r \sum_{j=1}^r \lambda_i \lambda_j d_{nii}^{-1} d_{njj}^{-1} \varphi_{it} \varphi_{i+h,j} \\ &= \sum_{i=1}^r \sum_{j=1}^r \lambda_i \lambda_j a_{hij} \stackrel{\text{(say)}}{=} g_\lambda(h), \end{aligned}$$

and  $c_i$  is completely analogous to  $(\sum_{j=1}^n C_j^2)^{-1/2} C_i$  of Theorem 6.3.4. By Theorem 6.3.4,  $\sum_{i=1}^n c_i Z_i$  converges to a normal random variable with variance

$$\sum_{h=-\infty}^{\infty} g_{\lambda}(h) \gamma_Z(h) = \sum_{i=1}^r \sum_{j=1}^r \lambda_i \lambda_j b_{ij},$$

where  $b_{ij} = \sum_{h=-\infty}^{\infty} a_{hij} \gamma_Z(h)$  is the  $ij$ th element of  $\mathbf{B}$  of (9.1.7). Since  $\lim_{n \rightarrow \infty} \mathbf{D}_n^{-1} \Phi' \Phi \mathbf{D}_n^{-1} = \mathbf{A}_0$  and since  $\lambda$  was arbitrary, the result follows from Theorem 5.3.3.  $\blacktriangle$

The matrix  $\mathbf{A}_h$  with  $ij$ th element equal to  $a_{hij}$  of (9.1.9) is analogous to the matrix  $\Gamma(h)$  of Section 4.4, and therefore the spectral representation

$$\mathbf{A}_h = \int_{-\pi}^{\pi} e^{i\omega h} d\mathbf{M}(\omega) \quad (9.1.11)$$

holds, where  $\mathbf{M}(\omega_2) - \mathbf{M}(\omega_1)$  is a positive semidefinite Hermitian matrix for all  $-\pi \leq \omega_1 < \omega_2 \leq \pi$ , and  $\mathbf{A}_0 = \mathbf{M}(\pi) - \mathbf{M}(-\pi)$ . We state without proof some of the results of Grenander and Rosenblatt (1957), which are based on this representation.

**Theorem 9.1.2.** Let the assumptions (9.1.1), (9.1.8), and (9.1.9) hold, and let the spectral density of the stationary time series  $Z_t$  be positive for all  $\omega$ . Then

$$\begin{aligned} \lim_{n \rightarrow \infty} \mathbf{D}_n^{-1} \Phi' \mathbf{V}_{zz} \Phi \mathbf{D}_n^{-1} &= 2\pi \int_{-\pi}^{\pi} f_z(\omega) d\mathbf{M}(\omega), \\ \lim_{n \rightarrow \infty} \mathbf{D}_n^{-1} \Phi' \mathbf{V}_{zz}^{-1} \Phi \mathbf{D}_n^{-1} &= \frac{1}{2\pi} \int_{-\pi}^{\pi} f_z^{-1}(\omega) d\mathbf{M}(\omega), \\ \lim_{n \rightarrow \infty} \mathbf{E}\{\mathbf{D}_n(\hat{\beta}_S - \boldsymbol{\beta})(\hat{\beta}_S - \boldsymbol{\beta})' \mathbf{D}_n\} &= 2\pi \mathbf{A}_0^{-1} \left[ \int_{-\pi}^{\pi} f_z(\omega) d\mathbf{M}(\omega) \right] \mathbf{A}_0^{-1}, \end{aligned} \quad (9.1.12)$$

$$\lim_{n \rightarrow \infty} \mathbf{E}\{\mathbf{D}_n(\hat{\beta}_G - \boldsymbol{\beta})(\hat{\beta}_G - \boldsymbol{\beta})' \mathbf{D}_n\} = \left[ \frac{1}{2\pi} \int_{-\pi}^{\pi} f_z^{-1}(\omega) d\mathbf{M}(\omega) \right]^{-1}, \quad (9.1.13)$$

where  $f_z(\omega)$  is the spectral density function of the process  $Z_t$ .

**Proof.** See Grenander and Rosenblatt (1957).  $\blacktriangle$

The asymptotic efficiency of the simple least squares and the generalized least squares is the same if the two covariance matrices (9.1.12) and (9.1.13) are equal. Following Grenander and Rosenblatt, we denote the set of points of increase of  $\mathbf{M}(\omega)$  by  $S$ . That is, the set  $S$  contains all  $\omega$  such that for any interval  $(\omega_1, \omega_2)$

where  $\omega_1 < \omega < \omega_2$ ,  $M(\omega_2) - M(\omega_1)$  is a positive semidefinite matrix and not the null matrix.

**Theorem 9.1.3.** Let the assumptions (9.1.1), (9.1.8), and (9.1.9) hold, and let the spectral density of the stationary time series  $Z_t$  be positive for all  $\omega$ . Then the simple least squares estimator and the generalized least squares estimator have the same asymptotic efficiency if and only if the set  $S$  is composed of  $q$  distinct points,  $\omega_1, \omega_2, \dots, \omega_q$ ,  $q \leq r$ .

**Proof.** See Grenander and Rosenblatt (1957). ▲

It can be shown that polynomials and trigonometric polynomials satisfy the conditions of Theorem 9.1.3. For example, we established the result for the special case of a constant mean function and autoregressive  $Y_t$  in Section 6.1. One may easily establish for the linear trend that

$$\lim_{n \rightarrow \infty} \frac{1}{2\pi n} \left( \frac{6}{n(n+1)(2n+1)} \right) \left( \sum_{s=1}^n \sum_{t=1}^n t e^{-\omega s t} s e^{-\omega s t} \right) = \begin{cases} 0, & \omega \neq 0, \\ \frac{3}{8\pi}, & \omega = 0, \end{cases}$$

from which it follows that the set  $S$  of points of increase is  $\omega_1 = 0$ .

The reader should not forget that these are asymptotic results. If the sample is of moderate size, it may be desirable to estimate  $V$  and transform the data to obtain final estimates of the trend function. Also, the simple least squares estimators may be asymptotically efficient, but it does not follow that the simple formulas for the estimated variances of the coefficients are consistent. In fact, the estimated variances may be badly biased. See Section 9.7.

## 9.2. GRAFTED POLYNOMIALS

In many applications the mean function is believed to be a "smooth" function of time but the functional form is not known. While it is difficult to define the term "smooth" in this context, several aspects of functional behavior can be identified. For a function defined on the real line, the function would be judged to be continuous and, in most situations, to have a continuous first derivative. This specification is incomplete in that one also often expects few changes in the sign of the first derivative.

Obviously low order polynomials satisfy the stated requirements. Also, by the Weierstrass approximation theorem, we know that any continuous function defined on a compact interval of the real line can be uniformly approximated by a polynomial. Consequently, polynomials have been heavily used to approximate the mean function. However, if the mean function is such that higher order polynomials are required, the approximating function may be judged unsatisfactory in that it contains a large number of changes in sign of the derivative. An alternative

approximation that generally overcomes this problem is to approximate segments of the function by low order polynomials and then join the segments to form a continuous function. The segments may be joined together so that the derivatives of a desired order are continuous.

Our presentation is based on the fact that, on the real line, the functions

$$\varphi_{ii} = \begin{cases} (t - A_i)^k, & t > A_i, \\ 0 & \text{otherwise,} \end{cases}$$

where  $i = 1, 2, \dots, M$ ,  $k$  and  $M$  are positive integers, and  $A_i > A_{i-1}$ , are continuous with continuous  $(k-1)$ st derivatives. It follows that

$$g(t) = b_0 + b_1 t + \dots + b_k t^k + \sum_{i=1}^M b_{k+i} \varphi_{ii}$$

is also continuous with continuous  $(k-1)$ st derivative. We call the function  $g(t)$  a *grafted polynomial of degree k*.

To illustrate the use of grafted polynomials in the estimation of the mean function of a time series let us make the following assumption: "The time series may be divided into periods of length  $A$  such that the mean function in each period is adequately approximated by a quadratic in time. Furthermore the mean function possesses a continuous first derivative." Let  $n$  observations indexed by  $t = 1, 2, \dots, n$  be available. We construct the functions  $\varphi_{ii}$ ,  $i = 1, 2, \dots, M$ , for  $k = 2$  and  $A_i = Ai$ , where  $M$  is an integer such that  $|A(M+1) - n| < A$ . In this formulation we assume that, if  $n \neq A(M+1)$ , the last interval is the only one whose length is not  $A$ . At the formal level all we need do is regress our time series upon  $t, t^2, \varphi_{i1}, \varphi_{i2}, \dots, \varphi_{iM}$  to obtain estimates of  $b_j$ ,  $j = 0, 1, \dots, M+2$ . However, if  $M$  is at all large, we can expect to encounter numerical problems in obtaining the inverse and regression coefficients. To reduce the possibility of numerical problems, we suggest that the  $\varphi$ 's be replaced by the linear combinations

$$w_{ii} = \varphi_{ii} - 3\varphi_{i,i+1} + 3\varphi_{i,i+2} - \varphi_{i,i+3}, \quad i = 1, 2, \dots, M,$$

where, for convenience, we define

$$\varphi_{i,M+1} = \varphi_{i,M+2} = \varphi_{i,M+3} = 0$$

for all  $t$ . Note that

$$w_{ii} = \begin{cases} (t - Ai)^2, & Ai < t < (i+1)A, \\ (t - Ai)^2 - 3[t - (i+1)A]^2, & (i+1)A \leq t < (i+2)A, \\ [t - (i+3)A]^2, & (i+2)A \leq t < (i+3)A, \\ 0 & \text{otherwise.} \end{cases}$$

Since the function  $w_{ii}$  is symmetric about  $(i+1.5)A$ , the  $w$ -variables can be

written down immediately. The  $w_{ii}$  remain correlated, but there should be little trouble in obtaining the inverse.

This procedure is felt to have merit when one is called on to extrapolate a time series. Since polynomials tend to plus or minus infinity as  $t$  increases, practitioners typically hesitate to use high order polynomials in extrapolation. Although a time series may display a nonlinear trend, we might wish to extrapolate on the basis of a linear trend. To accomplish this, we approximate the trend of the last  $K$  periods of a time series of  $n$  observations by a straight line tangent to a higher degree trend for the earlier portion of the time series. If the mean function for the first part of the sample is to be approximated by a grafted quadratic one could construct the regression variables

$$\begin{aligned}\varphi_{i1} &= t, \\ \varphi_{i,1+i} &= \begin{cases} [n - K - (i - 1)A - t]^2, & t < n - K - (i - 1)A, \\ 0 & \text{otherwise} \end{cases}\end{aligned}$$

for  $i = 1, 2, \dots, M$ , where  $|K + AM - n| < A$ . Using these variables, the estimated regression equation can be written as

$$\hat{b}_0 + \sum_{j=1}^{M+1} \hat{b}_j \varphi_{ij},$$

where the  $\hat{b}$ 's are the estimated regression coefficients. The forecast equation for the mean function is  $\hat{b}_0 + \hat{b}_1 t$ ,  $t = n + 1, n + 2, \dots$ .

We can also estimate a mean function that is linear for both the first and last portions of the observational period. Using a grafted quadratic in periods of length  $A$  for the remainder of the function, the required regression variables would be, for example,

$$\begin{aligned}\varphi_{i1} &= t, \\ \varphi_{i,1+i} &= \begin{cases} [t - K - A(i - 1)]^2, & K + A(i - 1) < t \leq K + Ai, \\ A^2 + 2A(t - K - Ai), & t > K + Ai, \\ 0 & \text{otherwise} \end{cases}\end{aligned}$$

for  $i = 1, 2, \dots, M$ , where  $n > K + MA$  and the mean function is linear for  $t < K$  and  $t > K + MA$ . If  $M$  is large, a transformation similar to that discussed above can be used to reduce computational problems.

**Example 9.2.1.** Yields of wheat in the United States for the period 1908 through 1991 are given in Table 9.2.1. In this example, we use only the data from 1908 through 1971. Yields for the first 20 to 30 years of the period are relatively constant, but there is a definite increase in yields from the late 1930s through 1971. As an approximation to the trend line, we fit a function that is constant for the first 25 years, increases at a quadratic rate until 1961, and is linear from 1961

Table 9.2.1. United States Wheat Yields (1908 to 1991)

Year	Yield	Year	Yield	Year	Yield
1908	14.3	1936	12.8	1964	25.8
1909	15.5	1937	13.6	1965	26.5
1910	13.7	1938	13.3	1966	26.3
1911	12.4	1939	14.1	1967	25.9
1912	15.1	1940	15.3	1968	28.4
1913	14.4	1941	16.8	1969	30.6
1914	16.1	1942	19.5	1970	31.0
1915	16.7	1943	16.4	1971	33.9
1916	11.9	1944	17.7	1972	32.7
1917	13.2	1945	17.0	1973	31.7
1918	14.8	1946	17.2	1974	27.3
1919	12.9	1947	18.2	1975	30.6
1920	13.5	1948	17.9	1976	30.3
1921	12.7	1949	14.5	1977	30.7
1922	13.8	1950	16.5	1978	31.4
1923	13.3	1951	16.0	1979	34.2
1924	16.0	1952	18.4	1980	33.5
1925	12.8	1953	17.3	1981	34.5
1926	14.7	1954	18.1	1982	35.5
1927	14.7	1955	19.8	1983	39.4
1928	15.4	1956	20.2	1984	38.8
1929	13.0	1957	21.8	1985	37.5
1930	14.2	1958	27.5	1986	34.4
1931	16.3	1959	21.6	1987	37.7
1932	13.1	1960	26.1	1988	34.1
1933	11.2	1961	23.9	1989	32.7
1934	12.1	1962	25.0	1990	39.5
1935	12.2	1963	25.2	1991	34.3

SOURCE: U.S. Department of Agriculture, *Agricultural Statistics*, various issues.

to 1971. A continuous function with continuous derivative that satisfies these requirements is

$$\varphi_{t1} = \begin{cases} 0, & 1 \leq t \leq 25, \\ (t - 25)^2, & 25 \leq t \leq 54, \\ 841 + 58(t - 54), & 54 \leq t, \end{cases} \quad (9.2.1)$$

where  $t = 1$  for 1908.

The trend line obtained by regressing the 64 yield observations on the vector  $(1, \varphi_{t1})$  is displayed in Figure 9.2.1. The equation for the estimated trend line is

$$\hat{Y}_t = 13.97 + 0.0123\varphi_{t1}$$

and the residual mean square is 2.80.

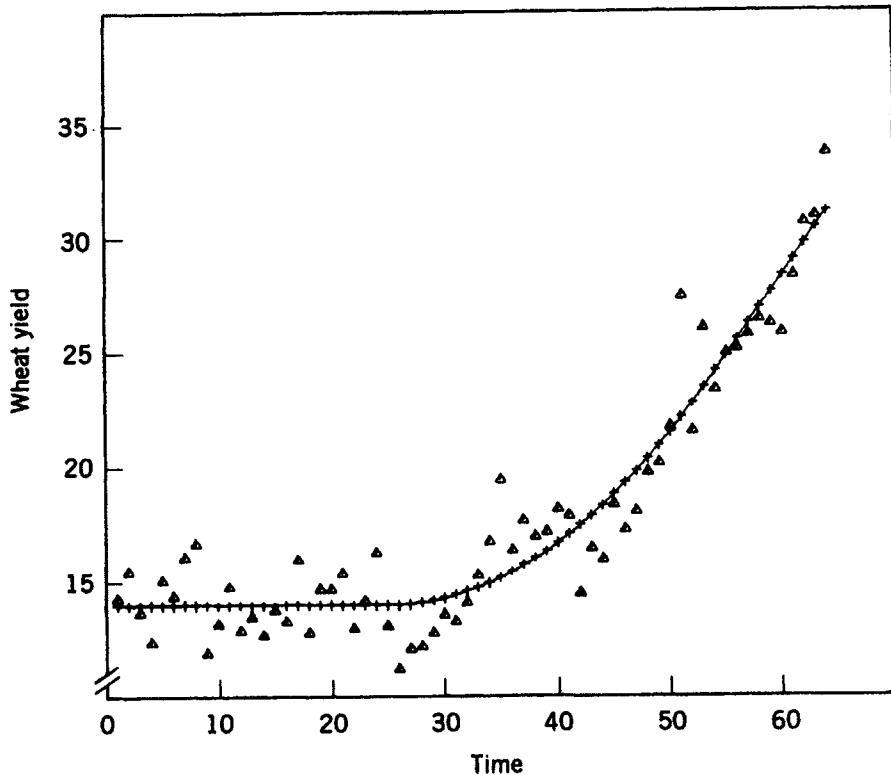


Figure 9.2.1. U.S. wheat yields 1908–1971 and grafted quadratic trend.

The trend in the data and the properties of the time series will be investigated further in Example 9.5.1 and Example 9.7.2. ▲▲

In the example and in our presentation we assumed that one was willing to specify the points at which the different polynomials are joined. The problem of using the data to estimate the join points has been considered by Gallant and Fuller (1973).

Grafted polynomials passing through a set of given points and joined in such a way that they satisfy certain restrictions on the derivatives are called *spline functions* in approximation theory. A discussion of such functions and their properties is contained in Ahlberg, Nilson, and Walsh (1967), Greville (1969), and Wahba (1990).

### 9.3. ESTIMATION BASED ON LEAST SQUARES RESIDUALS

#### 9.3.1. Estimated Autocorrelations

One of the reasons for estimating the trend function is to enable one to investigate the properties of the time series  $Z_t$ . Let the regression model be that defined in (9.1.1):

$$\mathbf{y} = \Phi\boldsymbol{\beta} + \mathbf{z}. \quad (9.3.1)$$

The calculated residuals using the simple least squares estimator are

$$\hat{Z}_t = Y_t - \varphi_t \hat{\beta}_S = Z_t - \varphi_t (\hat{\beta}_S - \boldsymbol{\beta}), \quad (9.3.2)$$

where

$$\hat{\beta}_S = (\Phi' \Phi)^{-1} \Phi' \mathbf{y}$$

and  $\varphi_t$  is the  $t$ th row of  $\Phi$ . The estimated autocovariance of  $Z_t$ , computed using the estimated  $Z_t$ , is

$$\hat{\gamma}_Z(h) = \frac{1}{n} \sum_{t=1}^{n-h} \hat{Z}_t \hat{Z}_{t+h}, \quad h = 0, 1, 2, \dots, n-1. \quad (9.3.3)$$

**Theorem 9.3.1.** Let  $Y_t$  be defined by the model (9.3.1) where  $Z_t = \sum_{j=0}^{\infty} \alpha_j e_{t-j}$ ,  $\{\alpha_j\}$  is absolutely summable,  $\{e_t\}$  is a sequence of independent  $(0, \sigma^2)$  random variables with bounded  $2 + \delta$  ( $\delta > 0$ ) moments, and the  $\varphi_{ti}$ ,  $i = 1, 2, \dots, r$ ,  $t = 1, 2, \dots$ , are fixed functions of time satisfying the assumptions (9.1.5) and (9.1.6). Then

$$\hat{\gamma}_Z(h) - \gamma_Z(h) = O_p(n^{-1}),$$

where  $\hat{\gamma}_Z(h)$  is defined in (9.3.3) and

$$\hat{\gamma}_Z(h) = \frac{1}{n} \sum_{t=1}^{n-h} Z_t Z_{t+h}, \quad h = 0, 1, 2, \dots, n-1.$$

**Proof.** We have, for  $h = 0, 1, \dots, n-1$ ,

$$\begin{aligned} \sum_{t=1}^{n-h} \hat{Z}_t \hat{Z}_{t+h} &= \sum_{t=1}^{n-h} Z_t Z_{t+h} - \sum_{t=1}^{n-h} Z_t \varphi_{t+h} \mathbf{D}_n^{-1} \mathbf{D}_n (\hat{\beta}_S - \boldsymbol{\beta}) \\ &\quad - \sum_{t=1}^{n-h} Z_{t+h} \varphi_t \mathbf{D}_n^{-1} \mathbf{D}_n (\hat{\beta}_S - \boldsymbol{\beta}) \\ &\quad + \sum_{t=1}^{n-h} (\hat{\beta}_S - \boldsymbol{\beta})' \mathbf{D}_n \mathbf{D}_n^{-1} \varphi_t' \varphi_{t+h} \mathbf{D}_n^{-1} \mathbf{D}_n (\hat{\beta}_S - \boldsymbol{\beta}), \end{aligned}$$

where  $\mathbf{D}_n$  is the diagonal matrix whose elements are the square roots of the elements of the principal diagonal of  $\Phi' \Phi$ . By the absolute summability of  $\gamma_Z(h)$ , the assumption (9.1.5), and the assumption (9.1.6)

$$\sum_{t=1}^{n-h} Z_t \varphi_{t+h} \mathbf{D}_n^{-1} = O_p(1),$$

$$\mathbf{D}_n (\hat{\beta}_S - \boldsymbol{\beta}) = O_p(1).$$

Since each element of the  $r \times r$  matrix

$$\sum_{t=1}^{n-h} \mathbf{D}_n^{-1} \boldsymbol{\varphi}'_t \boldsymbol{\varphi}_{t+h} \mathbf{D}_n^{-1}$$

is less than one in absolute value, the result follows.  $\blacktriangle$

It is an immediate consequence of Theorem 9.3.1 that the limiting behavior of the estimated autocovariances and autocorrelations defined by Theorem 6.3.5 and Corollary 6.3.5 also holds for autocovariances and autocorrelations computed using  $\hat{Z}_t$  in place of  $Z_t$ . The bias in the estimated autocorrelations is  $O(n^{-1})$  and therefore can be ignored in large samples. However, if the sample is small and if several  $\varphi_{ii}$  are included in the regression, the bias in  $\hat{\gamma}_2(h)$  may be sizable.

To briefly investigate the nature of the bias in  $\hat{\gamma}_2(h)$  we use the statistic studied by Durbin and Watson (1950, 1951). They suggested the von Neumann ratio (see Section 6.2) computed from the calculated residuals as a test of the hypothesis of independent errors. The statistic is

$$d = \frac{\sum_{t=2}^n (\hat{Z}_t - \hat{Z}_{t-1})^2}{\sum_{t=1}^n \hat{Z}_t^2} = \frac{\hat{\mathbf{z}}' \mathbf{H} \hat{\mathbf{z}}}{\hat{\mathbf{z}}' \hat{\mathbf{z}}}, \quad (9.3.4)$$

where

$$\mathbf{H} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ -1 & 2 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & -1 & 1 \end{bmatrix},$$

$\hat{\mathbf{z}} = [\mathbf{I} - \Phi(\Phi'\Phi)^{-1}\Phi']\mathbf{z}$  is the vector of calculated residuals, and  $\mathbf{z}$  is the vector of observations on the original time series. Recall that

$$d = 2 - 2\hat{r}_2(1) - \left( \sum_{t=1}^n \hat{Z}_t^2 \right)^{-1} (\hat{Z}_1^2 + \hat{Z}_n^2),$$

where

$$\hat{r}_2(1) = \left( \sum_{t=1}^n \hat{Z}_t^2 \right)^{-1} \sum_{t=2}^n \hat{Z}_t \hat{Z}_{t-1}.$$

The expected values of the numerator and denominator of  $d$  are given by

$$\begin{aligned} E\{\hat{\mathbf{z}}' \mathbf{H} \hat{\mathbf{z}}\} &= E\{\mathbf{z}' [\mathbf{I} - \Phi(\Phi'\Phi)^{-1}\Phi'] \mathbf{H} [\mathbf{I} - \Phi(\Phi'\Phi)^{-1}\Phi'] \mathbf{z}\} \\ &= \text{tr}[\mathbf{V}_{zz} \mathbf{H}] - \text{tr}[\mathbf{V}_{zz} \Phi(\Phi'\Phi)^{-1}\Phi' \mathbf{H}] - \text{tr}[\mathbf{V}_{zz} \mathbf{H} \Phi(\Phi'\Phi)^{-1}\Phi'] \\ &\quad + \text{tr}[\mathbf{V}_{zz} \Phi(\Phi'\Phi)^{-1}\Phi' \mathbf{H} \Phi(\Phi'\Phi)^{-1}\Phi'] \end{aligned} \quad (9.3.5)$$

and

$$E\{\hat{z}'\hat{z}\} = \text{tr}[\mathbf{V}_{zz} - \mathbf{V}_{zz}\Phi(\Phi'\Phi)^{-1}\Phi'], \quad (9.3.6)$$

where  $\mathbf{V}_{zz} = E\{\mathbf{z}\mathbf{z}'\}$ , and  $\text{tr } \mathbf{B}$  is the trace of the matrix  $\mathbf{B}$ .

It is easy to see that the expectation (9.3.5) may differ considerably from  $2(n-1)[\gamma_z(0) - \gamma_z(1)]$ . The expectation of the ratio is not necessarily equal to the ratio of the expectations, and this may further increase the bias in some cases (e.g., see Exercise 9.3).

We now turn to a consideration of  $d$  as a test of independence. For normal independent errors Durbin and Watson have shown that the ratio can be reduced to a canonical form by a simultaneous diagonalization of the numerator and denominator quadratic forms. Therefore, in this case, we can write

$$d = \frac{\sum_{j=1}^n \lambda_j \epsilon_j^2}{\sum_{j=1}^n \delta_j \epsilon_j^2} = \frac{\sum_{j=1}^n \lambda_j \epsilon_j^2}{\sum_{j=1}^{n-k'-1} \epsilon_j^2},$$

where it is assumed that the model contains an intercept and  $k'$  other independent variables, the  $\lambda_j$  are the roots of  $[\mathbf{I} - \Phi(\Phi'\Phi)^{-1}\Phi']\mathbf{H}[\mathbf{I} - \Phi(\Phi'\Phi)^{-1}\Phi]$ , the  $\delta_j$  are the roots of  $\mathbf{I} - \Phi(\Phi'\Phi)^{-1}\Phi'$ , and the  $\epsilon_j$  are normal independent random variables with mean zero and variance  $\sigma_\epsilon^2 = \sigma_z^2$ . Furthermore, the distribution of the ratio is independent of the distribution of the denominator. The exact distribution of  $d$ , depending on the roots  $\lambda$  and  $\delta$ , can be obtained. For normal independent errors  $Z_i$ , Durbin and Watson were able to obtain upper and lower bounds for the distribution, and they tabled the percentage points of these bounding distributions.

For small sample sizes and (or) a large number of independent variables the bounding distributions may differ considerably. Durbin and Watson suggested that the distribution of  $\frac{1}{4}d$  for a particular  $\Phi$ -matrix could be approximated by a beta distribution with the same first two moments as  $\frac{1}{4}d$ .

As pointed out in Section 6.2, in the null case the  $t$ -statistic associated with the first order autocorrelation computed from the von Neumann ratio has percentage points approximately equal to those of Student's  $t$  with  $n+3$  degrees of freedom. For the  $d$ -statistic computed from the regression residuals it is possible to use the moments of the  $d$ -statistic to develop a similar approximation. The expected value of  $d$  under the assumption of independent errors is

$$\begin{aligned} E\{d\} &= \{\text{tr } \mathbf{H} - \text{tr}[\Phi'\mathbf{H}\Phi(\Phi'\Phi)^{-1}]\}(n-k'-1)^{-1} \\ &= \{2(n-1) - \text{tr}[(\Delta\Phi)'(\Delta\Phi)(\Phi'\Phi)^{-1}]\}(n-k'-1)^{-1}, \end{aligned} \quad (9.3.7)$$

where  $(\Delta\Phi)'(\Delta\Phi)$  is the matrix of sums of squares and products of the first differences. Consider the statistics

$$\hat{\rho} = r_d + \frac{1}{2}(E\{d\} - 2) \frac{n-k'+1}{n-k'} (1-r_d^2) \quad (9.3.8)$$

and

$$t_d = (n - k' + 1)^{1/2} \hat{\rho} (1 - \hat{\rho}^2)^{-1/2}, \quad (9.3.9)$$

where  $r_d = \frac{1}{2}(2 - d)$ . The factor  $1 - r_d^2$  keeps  $\hat{\rho}$  in the interval  $[-1, 1]$  for most samples. For normal independent  $(0, \sigma^2)$  errors the expected value of  $\hat{\rho}$  is  $O(n^{-2})$  and the variance of  $\hat{\rho}$  differs from that of an estimated autocorrelation computed from a sample of size  $n - k'$  by terms of order  $n^{-2}$ . This suggests that the calculated  $t_d$  be compared with the tabular value of Student's  $t$  for  $n - k' + 3$  degrees of freedom. By using the second moment associated with the particular  $\Phi$ -matrix, the  $t$ -statistic could be further modified to improve the approximation, though there is little reason to believe that the resulting approximation would be as accurate as the beta approximation suggested by Durbin and Watson.

### 9.3.2. Estimated Variance Functions

The residuals can be used to investigate the nature of the variance of  $Z_t$  as well as to study the autocorrelation structure. By Proposition 9.1.1, the least squares estimator of  $\beta$  is consistent in the presence of variance heterogeneity. A simple form of variance heterogeneity is that in which the variance of the errors depends on fixed functions of time.

In Proposition 9.3.1, we show that the errors in the least squares estimated autoregressive coefficients are  $O_p(n^{-1/2})$  in the presence of variance heterogeneity.

**Proposition 9.3.1.** Let  $Z_t$  be a time series satisfying

$$\sum_{i=0}^p \alpha_i Z_{t-i} = a_t, \quad (9.3.10)$$

$$a_t = \sigma_{a_t} e_t$$

for  $t = 0, \pm 1, \pm 2, \dots$ , where  $e_t$  are independent  $(0, 1)$  random variables with bounded  $2 + \delta$  moments,  $\delta > 0$ ,  $\{\sigma_{a_t}\}$  is a fixed sequence bounded above by  $M_a$  and bounded away from zero, and the roots of the characteristic equation associated with (9.3.10) are less than one in absolute value. Define the least squares estimator of  $\alpha$  by

$$\hat{\alpha} = - \left( \sum_{t=p+1}^n \mathbf{X}_t' \mathbf{X}_t \right)^{-1} \sum_{t=p+1}^n \mathbf{X}_t' Z_t,$$

where  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)'$  and  $\mathbf{X}_t = (Z_{t-1}, Z_{t-2}, \dots, Z_{t-p})$ . Then

$$\mathbf{G}_n^{-1/2} (\hat{\alpha} - \alpha) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}),$$

where

$$\mathbf{G}_n = \left( \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \right)^{-1} \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \sigma_{at}^2 \left( \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \right)^{-1}. \quad (9.3.11)$$

**Proof.** The  $Z_t$  can be written as  $\sum_{j=0}^{\infty} w_j a_{t-j}$ , where the  $w_j$  are defined in (2.6.4). Now,

$$(n-h)^{-1} \sum_{t=1}^{n-h} Z_t Z_{t+h} = (n-h)^{-1} \sum_{t=1}^{n-h} \sum_{j=0}^{\infty} w_j w_{j+h} a_{t-j}^2 + (n-h)^{-1} \sum_{t=0}^{n-h} \sum_{j=0}^{\infty} \sum_{\substack{i=0 \\ i \neq j}}^{\infty} w_j w_{j+h} a_{t-j} a_{t-i}. \quad (9.3.12)$$

The  $a_t$  are independent and have bounded  $2 + \delta$  moments. Therefore,

$$p\lim n^{-1} \sum_{t=1}^n (a_t^2 - \sigma_{at}^2) = 0$$

and

$$p\lim (n-h)^{-1} \sum_{t=1}^{n-h} \sum_{j=0}^{\infty} w_j w_{j+h} (a_t^2 - \sigma_{at}^2) = 0$$

by Theorem 6.3.2.

Using the arguments of the proof of Theorem 6.3.5, it can be shown that the second term of (9.3.12) converges to zero in probability. Therefore,

$$p\lim \left[ n^{-1} \sum_{t=1}^{n-h} Z_t Z_{t+h} - n^{-1} \sum_{t=1}^{n-h} \sum_{j=0}^{\infty} w_j w_{j+h} \sigma_{a,t-j}^2 \right] = 0.$$

The error in the least squares estimator is

$$\hat{\alpha} - \alpha = - \left( \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \right)^{-1} \sum_{t=p+1}^n \mathbf{X}'_t a_t \quad (9.3.13)$$

and by the assumption (9.3.10),

$$E\{(\mathbf{X}'_t a_t, \mathbf{X}'_t \mathbf{X}_t a_t^2) | \mathcal{A}_{t-1}\} = (\mathbf{0}, \mathbf{X}'_t \mathbf{X}_t \sigma_{at}^2) \quad (9.3.14)$$

for  $t = 0, \pm 1, \pm 2, \dots$ , where  $\mathcal{A}_{t-1}$  is the sigma-field generated by  $Z_{t-1}, Z_{t-2}, \dots$ . Because  $a_{t-j}^2 \sigma_{at}^2$  has absolute moment greater than one, the arguments associated with obtaining the probability limit of (9.3.12) can be used to show that

$$p\lim_{n \rightarrow \infty} (n-p)^{-1} \left[ \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \sigma_{at}^2 - \sum_{t=p+1}^n E\{\mathbf{X}'_t \mathbf{X}_t\} \sigma_{at}^2 \right] = 0.$$

It follows that

$$\left( \sum_{t=p+1}^n \mathbf{X}'_t \mathbf{X}_t \sigma_{\alpha t}^2 \right)^{-1/2} \sum_{t=p+1}^n \mathbf{X}'_t \alpha_t \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}) \quad (9.3.15)$$

by Theorem 5.3.4.  $\blacktriangle$

The result of Proposition 9.3.1 can be extended to the estimated autoregressive parameters computed from least squares residuals.

**Proposition 9.3.2.** Let the model (9.1.1) hold, where  $Z_t$  is the autoregressive time series of (9.3.10) and  $\varphi_{it}$ ,  $i = 1, 2, \dots, r$ ,  $t = 1, 2, \dots$ , are fixed functions of time. Let the assumptions (9.1.5) and (9.1.6) of Proposition 9.1.1 and the assumptions of Proposition 9.3.1 hold. Let  $\hat{Z}_t$  be the least squares residuals of (9.3.2), and let

$$\tilde{\alpha} = - \left( \sum_{t=p+1}^n \hat{\mathbf{W}}'_t \hat{\mathbf{W}}_t \right)^{-1} \sum_{t=p+1}^n \hat{\mathbf{W}}'_t \hat{Z}_t,$$

where  $\hat{\mathbf{W}}_t = (\hat{Z}_{t-1}, \hat{Z}_{t-2}, \dots, \hat{Z}_{t-p})$ . Then

$$\mathbf{G}_n^{-1/2} (\tilde{\alpha} - \alpha) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}),$$

where  $\mathbf{G}_n$  is defined in (9.3.11).

**Proof.** Because

$$|E\{Z_t Z_{t+h}\}| < (1 - \lambda^2) \lambda^{|h|} M_a^2$$

for some  $0 < \lambda < 1$ , one can use the arguments of the proof of Theorem 9.3.1 to show that

$$\sum_{t=1}^{n-h} \hat{Z}_t \hat{Z}_{t+h} = \sum_{t=1}^{n-h} Z_t Z_{t+h} + O_p(1). \quad \blacktriangle$$

**Example 9.3.1.** The 150 observations in Table 9.3.1 were computer generated. A plot of the data suggests that the mean function is approximately a linear function of time. If we fit a linear trend by ordinary least squares, we obtain

$$\hat{Y}_t = 15.472 + 0.3991t. \quad (9.3.16)$$

If a second order autoregression is fitted to the residuals by ordinary least squares, we obtain

$$\hat{Z}_t = \frac{1.524}{(0.067)} \hat{Z}_{t-1} - \frac{0.616}{(0.067)} \hat{Z}_{t-2}, \quad (9.3.17)$$

Table 9.3.1. Data for Example 9.3.1

16.45	25.13	46.49	49.57	73.07
16.39	26.21	47.78	51.03	72.83
16.25	26.98	48.05	50.80	70.05
16.23	28.60	46.54	51.49	66.30
17.16	30.80	43.15	52.91	61.47
17.55	32.65	40.34	56.95	60.47
17.48	32.92	41.30	59.32	61.30
18.32	34.41	42.49	59.76	62.38
18.31	35.65	43.00	59.29	63.46
19.09	37.15	43.75	58.74	62.38
19.53	37.49	45.86	57.80	61.61
18.08	37.64	45.95	56.27	61.82
19.21	37.03	44.87	52.98	61.03
19.96	36.78	44.22	52.27	62.18
20.69	37.26	43.95	56.17	62.60
21.15	36.83	43.52	61.33	61.83
19.69	36.38	41.51	64.69	64.23
19.36	36.79	39.41	68.49	65.62
18.83	35.87	36.07	70.19	67.17
19.84	35.89	37.23	71.47	68.61
22.20	36.21	40.14	70.98	69.36
22.42	36.95	40.41	68.63	69.91
22.18	37.93	41.06	67.47	70.28
22.38	36.55	39.56	67.68	71.92
22.45	33.94	40.33	68.47	72.48
23.47	34.28	43.90	69.52	71.59
25.03	37.51	47.28	71.47	70.19
25.41	40.42	48.71	73.33	69.69
24.91	41.70	49.46	72.17	73.55
25.49	43.39	50.61	72.72	77.55

where  $s^2 = 1.646$  and the standard errors are those of the ordinary regression program.

If the absolute values of the residuals from the autoregression are plotted against time, there is an increase in the mean of the absolute values. Several models can be considered for the variance of the errors. We postulate a model in which the variance is related to the mean. Our model for the variance of  $a_t$  is

$$\sigma_{at}^2 = E\{a_t^2\} = e^{\kappa_1}[E\{Y_t\}]^{\kappa_2}, \quad (9.3.18)$$

where  $a_t$  is the error in the autoregression estimated by (9.3.17). We estimated  $(\kappa_1, \kappa_2)$ , replacing  $a_t^2$  with  $\hat{a}_t^2$  from regression (9.3.17) and replacing  $E\{Y_t\}$  with  $\hat{Y}_t$  from (9.3.16), by maximum likelihood under the assumption that  $a_t$  are NI(0,  $\sigma_{at}^2$ ). The estimates are

$$(\hat{\kappa}_1, \hat{\kappa}_2) = (-3.533, 1.051), \\ (0.558) (0.148)$$

where the standard errors are from the inverse of the estimated information matrix.

We have estimated the variance function for the  $a_t$ , but if the variance function is a smooth function, the variance of the autoregressive process  $Z_t$  will be nearly a multiple of the variance of the  $a_t$ . Therefore, we define new variables

$$(\hat{\sigma}_{at}^{-1} Y_t, \hat{\sigma}_{at}^{-1}, \hat{\sigma}_{at}^{-1} t) = (L_t, \varphi_{t1}, \varphi_{t2}),$$

where  $\hat{\sigma}_{at}$  is the function (9.3.18) evaluated at  $E\{Y_t\} = \hat{Y}_t$  and  $(\hat{\kappa}_1, \hat{\kappa}_2) = (-3.533, 1.051)$ . The parameters of the model

$$\begin{aligned} L_t &= \varphi_{t1}\beta_1 + \varphi_{t2}\beta_2 + u_t, \\ u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} &= e_t, \end{aligned} \tag{9.3.19}$$

estimated by Gaussian maximum likelihood, are

$$(\beta_1, \beta_2, \alpha_1, \alpha_2, \sigma_e^2) = (15.113, 0.407, -1.473, 0.570, 0.995). \\ (1.465) (0.021) (0.068) (0.068) (0.090)$$

Estimation for a model such as (9.3.19) is discussed in Section 9.7. The estimate of  $(\beta_1, \beta_2, \alpha_1, \alpha_2)$  is not greatly different from that obtained in (9.3.16) and (9.3.17) by ordinary least squares. However, the confidence interval for a prediction is different under the model in which the variance is a function of the mean. The standard errors for predictions one, two, and three periods ahead are (1.283, 2.328, 3.178) for the model computed under the assumption of homogeneous variances. The corresponding prediction standard errors are (1.673, 2.985, 4.023) for the model with the variance function of (9.3.18). ▲

In Example 9.3.1, the variance was assumed to be a fixed function of time. Models in which the variance is a random function have also been considered in the literature. Most often the variance at the current time is expressed as a function of the past behavior of the time series. Under mild assumptions, it can be shown that the least squares estimator of the parameter vector of the autoregressive process has a normal distribution in the limit.

**Proposition 9.3.3.** Let  $(Z_t, \sigma_{at})$ ,  $t = 0, \pm 1, \pm 2, \dots$ , be a covariance stationary time series satisfying

$$Z_t = \sum_{j=0}^{\infty} w_j a_{t-j}, \\ a_t = \sigma_{at} e_t,$$

where the  $w_j$  are absolutely summable,  $e_t$  are iid(0, 1) random variables with  $2 + \delta_2$  ( $\delta_2 > 0$ ) moments,  $e_{t+j}$  is independent of  $\sigma_{at}$  for  $j \geq 0$ , and

$$\operatorname{plim}_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \sigma_{ai}^2 = \sigma_a^2,$$

where  $\sigma_a^2 = E\{\sigma_{ai}^2\}$ . Then

$$\operatorname{plim}_{n \rightarrow \infty} n^{-1} \sum_{i=1}^{n-h} (Z_i - \bar{z}_n)(Z_{i+h} - \bar{z}_n) = \gamma_Z(h),$$

where  $\bar{z}_n = n^{-1} \sum_{t=1}^n Z_t$  and

$$\gamma_Z(h) = \sum_{j=0}^{\infty} w_j w_{j+h} \sigma_a^2.$$

If, in addition,  $Z_t$  is an autoregressive process satisfying

$$Z_t + \sum_{i=1}^p \alpha_i Z_{t-i} = a_t,$$

where the roots of the characteristic equation are less than one in absolute value, and if

$$\operatorname{plim}_{n \rightarrow \infty} n^{-1} \sum_{i=p+1}^n \mathbf{X}'_i \mathbf{X}_i \sigma_{ai}^2 = E\{\mathbf{X}'_i \mathbf{X}_i \sigma_{ai}^2\}, \quad (9.3.20)$$

then

$$\mathbf{G}_n^{-1/2}(\hat{\alpha} - \alpha) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}),$$

where  $\alpha$ ,  $\hat{\alpha}$ ,  $\mathbf{G}_n$ , and  $\mathbf{X}_i$  are defined in Proposition 9.3.1.

**Proof.** The sequence  $\{a_t\}$  is a martingale difference sequence and

$$n^{1/2} \bar{a}_n \xrightarrow{\mathcal{L}} N(0, \sigma_a^2)$$

by Theorem 5.3.4. Hence,  $n^{1/2} \bar{z}_n$  also converges in distribution to a normal random variable by Theorem 6.3.3. Thus,

$$(n-h)^{-1} \sum_{i=1}^{n-h} (Z_i - \bar{z}_n)(Z_{i+h} - \bar{z}_n) = (n-h)^{-1} \sum_{i=1}^n Z_i Z_{i+h} + O_p(n^{-1}).$$

We expand  $n^{-1} \sum_{i=1}^{n-h} Z_i Z_{i+h}$  as in (9.3.12) and observe that

$$\operatorname{plim}_{n \rightarrow \infty} (n-h)^{-1} \sum_{i=1}^{n-h} \sum_{j=0}^h w_j w_{j+h} a_{i-j}^2 = \gamma_Z(h)$$

by Theorem 6.3.2. It can be shown that

$$\operatorname{plim}_{n \rightarrow \infty} \sum_{t=1}^{n-h} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} w_j w_{i+h} a_{t-i} a_{t-j} = 0$$

by applying the arguments of the proof of Theorem 6.3.5. Therefore, the sample covariance is consistent for the population covariance.

The error in the least squares estimator is given in (9.3.13). Under the present assumptions, (9.3.14) holds and, by the assumption (9.3.20) and Theorem 5.3.4, we obtain the limiting normal distribution result of (9.3.15). The limiting result for  $\mathbf{G}_n^{-1/2}(\hat{\alpha} - \alpha)$  then follows. ▲

Engle (1982) introduced a variance model called the autoregressive conditional heteroscedastic model (ARCH). An example is

$$\begin{aligned} Z_t &= \theta Z_{t-1} + a_t, \\ a_t &= (\kappa_1 + \kappa_2 a_{t-1}^2)^{1/2} e_t, \end{aligned} \quad (9.3.21)$$

where the  $e_t$  are  $N(0, 1)$  random variables and  $0 \leq \kappa_2 < 3^{-1/2}$ . Some properties of this model are developed in Exercise 2.43 of Chapter 2. Clearly, many extensions of the simple model are possible, and a number have been considered in the literature. See Engle and Bollerslev (1986), Weiss (1984, 1986), Pantula (1986), Bollerslev, Chou, and Kroner (1992), Higgins and Bera (1992), and Granger and Teräsvirta (1993).

A model closely related to the ARCH model postulates the variances to be an unobserved stochastic process. See Melino and Turnbull (1990), Jacquier, Polson, and Rossi (1994), and Harvey, Ruiz, and Shephard (1994). An example model is

$$Z_t = \theta Z_{t-1} + a_t, \quad (9.3.22)$$

$$a_t = \sigma_a e_t, \quad (9.3.23)$$

$$h_t - \mu_h = \psi(h_{t-1} - \mu_h) + b_t, \quad (9.3.24)$$

where  $h_t = \log \sigma_a$  and  $(e_t, b_t) \sim \Pi[(0, 0), \operatorname{diag}(1, \sigma_b^2)]$ . In (9.3.24), the logarithm of the unobserved variance process is a first order autoregressive process. Thus,  $\sigma_a$  is always positive.

If the  $a_t$  are such that  $\log a_t^2$  has second moments, it is natural to take logarithms of (9.3.23) to create a linear model. Then,

$$\log a_t^2 = 2h_t + \log e_t^2. \quad (9.3.25)$$

If  $e_t$  is normally distributed, then  $\log e_t^2$  is distributed as log chi square which has mean of about  $-1.27$  and variance of about  $4.93$ . The distribution of  $\log e_t^2$  is heavily skewed to the left. If (9.3.24) is a stationary autoregressive process and  $e_t$  is normally distributed, then  $\log a_t^2$  is a stationary autoregressive moving average process. Nonlinear estimation methods can be used to estimate the parameters of

(9.3.24) treating the variance of  $\log e^2$ , as known or unknown. In Example 9.3.2, we consider a modification of the log transformation in the estimation.

**Example 9.3.2.** As an example of the fitting of a stochastic volatility model, we consider the prices of June futures contracts for the Standard and Poor's 500 index, collected as the average price per minute from 8:30 a.m. to 3:15 p.m. on April 13, 1987. Our basic observations are the first differences of the prices. The data are considered by Breidt and Carriquiry (1994). If a first order autoregression is fitted to the data, the estimated coefficient is 0.340 with a standard error of 0.047. The residuals from the first order autoregression, denoted by  $\hat{a}_t$ , are plotted against time in Figure 9.3.1.

The observations in Figure 9.3.1 show evidence of a serially correlated variance. At the end of the series, large deviations tend to be followed by large deviations, while in the middle, small deviations tend to be followed by small deviations. This dependence is reflected in the sample autocorrelations for the  $\hat{a}_t^2$ . The first five autocorrelations of the  $\hat{a}_t^2$  are 0.204, 0.501, 0.205, 0.346, and 0.263. These correlations are significantly different from zero because the standard error under zero correlation is about 0.05.

We wish to fit a variance model of the form (9.3.23), (9.3.24) to the residuals from the autoregressive fit. In practice, there are some disadvantages to a logarithmic transformation of the  $a_t^2$ , as defined in (9.3.25). There may be some zero values for  $a_t^2$ , and the original distribution may not be normal. Therefore, we

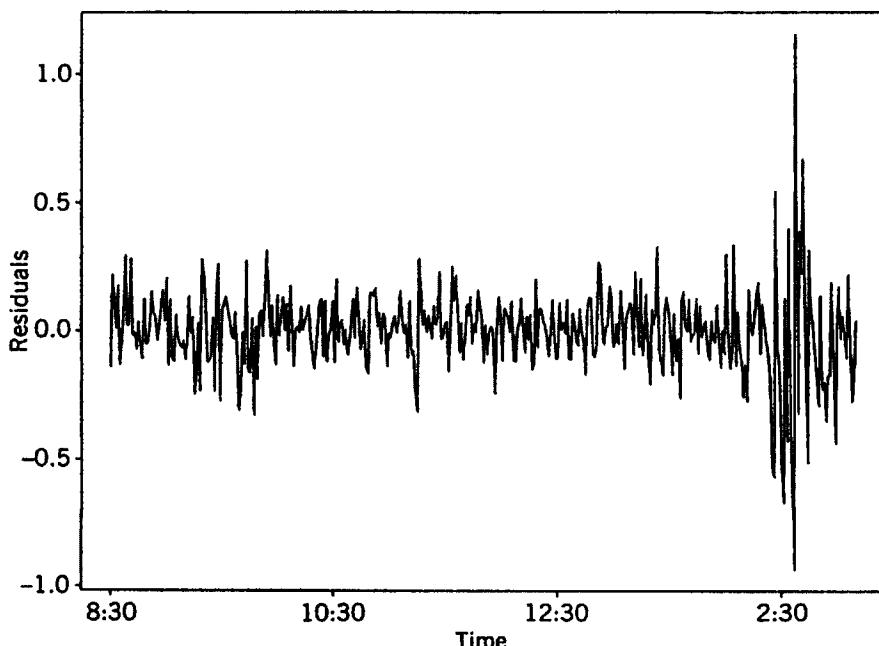


Figure 9.3.1. Residuals from autoregression of price differences.

suggest the following estimation scheme that is more robust against nonnormality. Let  $\xi$  be a small fraction, say 0.02, of the average of the  $a_i^2$ . Let

$$\begin{aligned} U_i &= \log(a_i^2 + \xi) - (a_i^2 + \xi)^{-1}\xi \\ &= 2h_i + w_i, \end{aligned} \quad (9.3.26)$$

where

$$w_i = \log(e_i^2 + \sigma_{a_i}^{-2}\xi) - (e_i^2 + \sigma_{a_i}^{-2}\xi)^{-1}\sigma_{a_i}^{-2}\xi. \quad (9.3.27)$$

If  $e_i^2$  is large,  $w_i$  differs little from  $\log e_i^2$ . For example, the difference is less than 0.016 if  $\sigma_{a_i}^{-2}\xi = 0.02$  and  $e_i^2 > 0.10$ . For normal (0, 1) random variables and  $\sigma_{a_i}^{-2}\xi = \xi = 0.02$ , the mean of  $w_i$  is about -1.09 and the variance is about 3.19. Thus, the  $\xi$ -modification of the log transformation produces variables with smaller variance than  $\log e_i^2$ , because the function (9.3.26) is bounded below by  $\log \xi - 1$ . Also, the distribution of  $w_i$  is much less skewed than that of  $\log e_i^2$ .

The  $U_i$  of (9.3.26) can be used to estimate the parameters of the  $h_i$ -process up to an additive constant. For example, if our specification for the  $h_i$ -process is a  $p$ th order autoregressive process, we write

$$\begin{aligned} U_i &= \mu_U + X_i + g_i, \\ X_i &= \sum_{i=1}^p \psi_i X_{i-i} + b_i, \end{aligned} \quad (9.3.28)$$

where  $\mu_U = E\{2h_i\} + E\{w_i\}$ ,  $g_i = w_i - E\{w_i\}$ ,  $X_i = 2h_i - E\{2h_i\}$ ,  $w_i$  is defined in (9.3.27),  $\{b_i\}$  is independent of  $\{g_i\}$ , and  $b_i \sim II(0, \sigma_b^2)$ . Because of the nature of our transformation, the conditional distributions of the  $g_i$  given  $h_i$  are only approximately equal.

We use a first order autoregressive model for the  $h_i$  of stock prices. Then the  $U_i$  of (9.3.26) is an autoregressive moving average, which we write as

$$U_i - \mu_U = \psi(U_{i-1} - \mu_U) + \eta_i + \beta\eta_{i-1}. \quad (9.3.29)$$

We replace  $a_i^2$  with  $\hat{a}_i^2$  in the definition of  $U_i$  of (9.3.26) and set  $\xi = 0.00058$ , where the average of the  $\hat{a}_i^2$  is 0.0290. The estimates of the parameters are

$$(\hat{\psi}, \hat{\beta}, \hat{\sigma}_{\eta}^2) = (0.939, -0.816, 2.961). \\ (0.033) \quad (0.055)$$

From (9.3.29) and (9.3.28),  $\beta\sigma_{\eta}^2 = -\psi\sigma_g^2$  and

$$(1 + \beta^2)\sigma_{\eta}^2 = \sigma_b^2 + (1 + \psi^2)\sigma_g^2.$$

It follows that

$$\hat{\sigma}_g^2 = -\hat{\psi}^{-1}\hat{\beta}\hat{\sigma}_{\eta}^2 = 2.5742,$$

$$\hat{\sigma}_b^2 = \hat{\sigma}_{\eta}^2[1 + \hat{\beta}^2 + \hat{\psi}^{-1}\hat{\beta}(1 + \hat{\psi}^2)] = 0.0903.$$

The fact that  $\hat{\sigma}_g^2$  differs considerably from 3.19 suggests that the original  $e_t$  are not normally distributed. Given an estimate of the parameter  $\psi$  of the  $X_t$ -process, we can construct smoothed estimates of  $X_t$ ,  $t = 1, 2, \dots, n$  and, hence, estimates of  $2h_t$ , up to an additive constant. Then the constant can be estimated as a multiplicative constant in the original (nonlogarithmic) scale. Let  $\zeta$  be defined by

$$\sigma_{at}^2 = \zeta \exp\{2(h_t - \mu_h)\}, \quad (9.3.30)$$

where  $\mu_h = E\{h_t\}$ . Then an estimator of  $\zeta$  is

$$\hat{\zeta} = n^{-1} \sum_{t=1}^n [\exp(\hat{X}_t)]^{-1} \hat{a}_t^2, \quad (9.3.31)$$

where  $\hat{X}_t$  is the smoothed estimator of  $X_t$  constructed from the fitted model (9.3.28). The quantities

$$\hat{\sigma}_{at}^2 = \hat{\zeta} \exp(\hat{X}_t) \quad (9.3.32)$$

are smoothed estimates of the unknown  $\sigma_{at}^2$ .

If the  $\hat{\sigma}_{at}^2$  are quite variable over the sample, a second round of calculations can be carried out in which the initial transformation is

$$U_t = \log(\hat{a}_t^2 + \hat{\sigma}_{at}^2 k) - (\hat{a}_t^2 + \hat{\sigma}_{at}^2 k)^{-1} \hat{\sigma}_{at}^2 k,$$

where  $k$  is a constant, such as 0.02, and  $\hat{\sigma}_{at}^2 = \hat{\zeta} \exp(\hat{X}_t)$ . In this example, the estimates are changed very little by the second round of computations. However, in some situations there is a considerable change, and in general a second round of computation is recommended. See Breidt and Carriquiry (1994).

Given the estimates of  $\sigma_g^2$ ,  $\sigma_b^2$ , and  $\psi$  for the price data, we constructed smoothed estimates  $\hat{X}_t$  using the Kalman filter recursions for fixed interval smoothing described, for example, in Harvey (1989, p. 154) and Anderson and Moore (1979, Chapter 7). The estimate of  $\zeta$  of (9.3.30) is  $\hat{\zeta} = 0.01550$ . The smoothed estimates of the standard deviations  $\sigma_{at}$  from (9.3.32) are plotted against time in Figure 9.3.2. They reflect the pattern observed in the data. That is, there is a period of high volatility near the end of the series and low volatility in the middle of the observation period. ▲▲

## 9.4. MOVING AVERAGES—LINEAR FILTERING

### 9.4.1. Moving Averages for the Mean

In the previous sections we considered methods of estimating the mean function for the entire period of observation. One may be interested in a simple approximation to the mean as a part of a preliminary investigation, where one is not willing to specify the mean function for the entire period, or one may desire a relatively simple method of removing the mean function that will permit simple

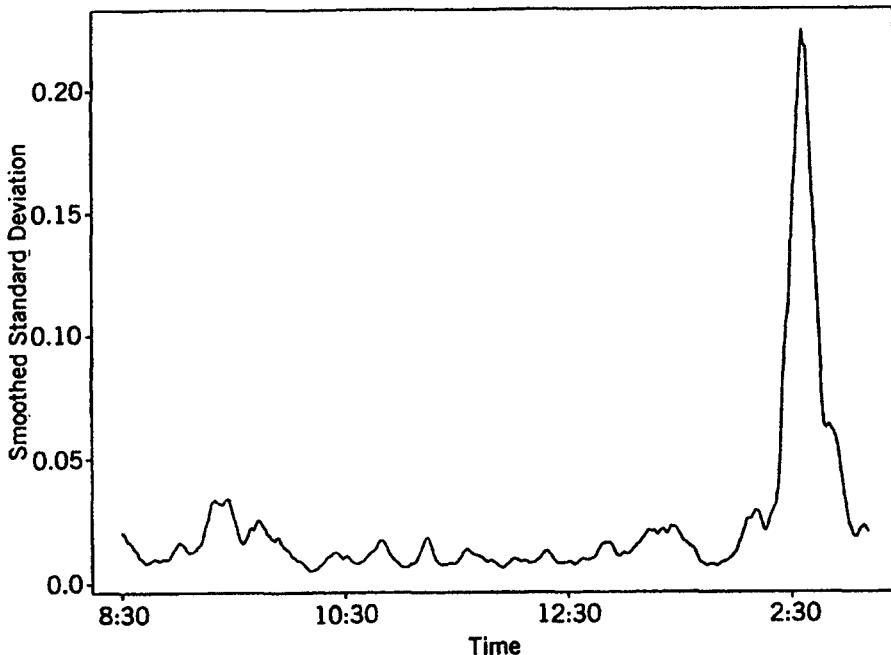


Figure 9.3.2. Smoothed estimates of  $\sigma_u$  for price differences.

extrapolation of the time series. In such cases the method of moving averages may be appropriate.

One basis for the method of moving averages is the presumption that for a period of  $M$  observations, the mean is adequately approximated by a specified function. The function is typically linear in the parameters and most commonly is a polynomial in  $t$ . Thus, the specification is

$$Y_{t+j} = g(j; \beta_t) + Z_{t+j}, \quad j = -M_1, -M_1 + 1, \dots, M_2 - 1, M_2, \quad (9.4.1)$$

where  $Z_t$  is a stationary time series with zero expectation,  $\beta_t$  is a vector of parameters, and  $M = M_1 + M_2 + 1$ . The form of the approximating function  $g$  is assumed to hold for all  $t$ , but the parameters are permitted to be a function of  $t$ . Both the "local" and "approximate" nature of the specification should now be clear. If the functional form for the expectation of  $Y_t$  held exactly in the interval, then it would hold for the entire realization, and the constants  $M_1$  and  $M_2$  would become  $t - 1$  and  $n - t$ , respectively.

Given specification (9.4.1), a set of weights,  $w_j(s)$ , are constructed that, when applied to the  $Y_{t+j}$ ,  $j = -M_1, -M_1 + 1, \dots, M_2 - 1, M_2$ , furnish an estimator of  $g(s; \beta_t)$  for the specified  $s$ . It follows that an estimator of  $Z_{t+s}$  is given by

$$Y_{t+s} - \hat{g}(s; \beta_t),$$

where

$$\hat{g}(s; \beta_i) = \sum_{j=-M_1}^{M_2} Y_{t+j} w_j(s).$$

In the terminology of Section 4.3 the set of weights  $\{w_j(s)\}$  is a linear filter.

Let us consider an example. Assume that for a period of five observations the time series is adequately represented by

$$Y_{t+j} = \beta_{0i} + \beta_{1i} j + \beta_{2i} j^2 + Z_{t+j}, \quad j = -2, -1, 0, 1, 2, \quad (9.4.2)$$

where  $Z_t$  is a stationary time series with zero expectation. Using this specification, we calculate the least squares estimator for the trend value of the center observation,  $s = 0$ , as a linear function of the five observations. The model may be written in matrix form as

$$\mathbf{y}_t = \Phi \boldsymbol{\beta}_t + \mathbf{z}_t,$$

where

$$\boldsymbol{\beta}'_t = (\beta_{0i}, \beta_{1i}, \beta_{2i}),$$

$\mathbf{y}_t = (Y_{t-2}, Y_{t-1}, Y_t, Y_{t+1}, Y_{t+2})'$  is the vector in the first column of Table 9.4.1,  $\Phi$  is the matrix defined by the second, third, and fourth columns of Table 9.4.1 (those columns headed by  $\beta_{0i}$ ,  $\beta_{1i}$ , and  $\beta_{2i}$ ), and  $\mathbf{z}_t = (Z_{t-2}, Z_{t-1}, Z_t, Z_{t+1}, Z_{t+2})'$  is the vector of (unobservable) elements of the stationary time series. The least squares estimator of the mean at  $j = 0$ ,  $g(0; \boldsymbol{\beta}_t)$ , is given by

$$\begin{aligned} \hat{\beta}_{0i} &= (1, 0, 0)(\Phi' \Phi)^{-1} \Phi' \mathbf{y}_t, \\ &= \sum_{j=-2}^2 w_j(0) Y_{t+j} = \mathbf{w}' \mathbf{y}_t, \end{aligned}$$

where the vector  $(1, 0, 0)$  is the value of the vector of independent variables associated with the third observation. That is,  $(1, 0, 0)$  is the third row of the matrix  $\Phi$  and is associated with  $j = 0$ . The vector of weights

Table 9.4.1. Calculation of Weights for a Five Period Quadratic Moving Average

$y_t$	Weights for Trend					Weights for Trend Adjusted Series ( $s = 0$ )	
	$\beta_{0i}$	$\beta_{1i}$	$\beta_{2i}$	$g(0; \boldsymbol{\beta}_t)$	$g(1; \boldsymbol{\beta}_t)$	$g(2; \boldsymbol{\beta}_t)$	
$Y_{t-2}$	1	-2	4	-6/70	-10/70	6/70	6/70
$Y_{t-1}$	1	-1	1	24/70	12/70	-10/70	-24/70
$Y_t$	1	0	0	34/70	24/70	-6/70	36/70
$Y_{t+1}$	1	1	1	24/70	26/70	18/70	-24/70
$Y_{t+2}$	1	2	4	-6/70	18/70	62/70	6/70

$$\mathbf{w}' = (1, 0, 0)(\Phi' \Phi)^{-1} \Phi',$$

once computed, can be applied to any vector  $\mathbf{y}_t$  of five observations. These weights are given in the fifth column of Table 9.4.1 under the heading “ $g(0; \beta_t)$ .” The least squares estimator of  $Z_t$  for the third observation ( $s = 0$ ) is given by

$$Y_t - \hat{\beta}_{0t} = Y_t - (1, 0, 0)(\Phi' \Phi)^{-1} \Phi' \mathbf{y}_t.$$

The vector of weights giving the least squares estimator of  $Z_t$  is presented in the last column of Table 9.4.1. One may readily verify that the inner product of this vector of weights with each of the columns of the original matrix  $\Phi$  is zero. Thus, if the original time series satisfies the specification (9.4.2), the time series of estimated residuals created by applying the weights in the last column will be a stationary time series with zero mean. The created time series  $X_t$  is given by

$$\begin{aligned} X_t &= \sum_{j=-2}^2 r_j Y_{t+j} = \sum_{j=-2}^2 r_j (\beta_{0t} + \beta_{1t} j + \beta_{2t} j^2 + Z_{t+j}) \\ &= \frac{1}{70} (6Z_{t-2} - 24Z_{t-1} + 36Z_t - 24Z_{t+1} + 6Z_{t+2}), \end{aligned}$$

where  $r_j$  denotes the weights in the last column of Table 9.4.1. Although  $X_t$  can be thought of as an estimator of  $Z_t$ ,  $X_t$  is a linear combination of the five original  $Z_t$  included in the moving average.

A regression program that computes the estimated value and deviation from fit for each observation on the dependent variable is a convenient method of computation. Form a regression problem with the independent variables given by the trend specification (e.g., constant, linear, and quadratic) and the dependent variable defined by

$$u_{t+j} = \begin{cases} 1, & j = s, \\ 0 & \text{otherwise,} \end{cases}$$

where the estimated mean is to be computed for  $j = s$ . The vector of weights for the estimated mean is then given by

$$\mathbf{w} = \hat{\mathbf{v}} = \Phi(\Phi' \Phi)^{-1} \Phi' \mathbf{v} = \Phi(\Phi' \Phi)^{-1} \varphi_s', \quad (9.4.3)$$

where  $\varphi_s$  is the vector of observations on the independent variables associated with  $j = s$ . Similarly, the vector of weights for computing the trend adjusted time series at  $j = s$  is given by

$$\mathbf{v} - \hat{\mathbf{v}} = \mathbf{v} - \Phi(\Phi' \Phi)^{-1} \Phi' \mathbf{v} = \mathbf{v} - \Phi(\Phi' \Phi)^{-1} \varphi_s'. \quad (9.4.4)$$

The computation of (9.4.3) is recognized as the computation of  $\hat{\mathbf{y}}$  associated with the regression of the vector  $\mathbf{v}$  on  $\Phi$ . The weights in (9.4.4) are then the deviations from fit of the same regression.

The moving average estimator of the trend and of the trend adjusted time series are frequently calculated for  $j = 0$  and  $M_1 = M_2$ . This configuration is called a *centered moving average*. The popularity of the centered moving average is perhaps due to the following theorem.

**Theorem 9.4.1.** The least squares weights constructed for the estimator of trend for a centered moving average of  $2M + 1$  observations,  $M$  a positive integer, under the assumption of a  $p$ th degree polynomial,  $p$  nonnegative and even, are the same as those computed for the centered moving average of  $2M + 1$  observations under the assumption of a  $(p + 1)$ st degree polynomial.

**Proof.** We construct our matrix  $\Phi$  in a manner analogous to that of Table 9.4.1, that is, with columns given by  $j^k$ ,  $k = 0, 1, 2, \dots, p$ ,  $j = 0, \pm 1, \pm 2, \dots, \pm M$ .

Since we are predicting trend for  $j = 0$ , the regression coefficients of all odd powers are multiplied by zero. That is, the elements of  $\varphi_s'$  in (9.4.3) associated with the odd powers are all zero. It remains only to show that the regression coefficients of the even powers remain unchanged by the presence or absence of the  $(p + 1)$ st polynomial in the regression. But

$$\sum_{j=-M}^{M} j^{2k+1} = 0$$

for  $k$  a positive integer. Therefore, the presence or absence of odd powers of  $j$  in the matrix  $\Phi$  leaves the coefficients of the even powers unchanged, and the result follows. ▲

The disadvantage of a centered moving average is the loss of observations at the beginning and at the end of the realization. The loss of observations at the end of the observed time series is particularly critical if the objective of the study is to forecast future observations. If the trend and trend adjusted time series are computed for the end observations using the same model, the variance-covariance structure of these estimates differs from those computed for the center of the time series. We illustrate with our example.

Assume that the moving average associated with equation (9.4.2) and Table 9.4.1 is applied to a sequence of independent identically distributed random variables. That is, the trend value for the first observation is given by applying weights  $g(-2, \beta_1)$  to the first five observations, and the trend adjusted value for the first observation is obtained by subtracting the trend value from  $Y_1$ . The weights  $g(-2, \beta_1)$  are the weights  $g(2, \beta_1)$  arranged in reverse order. For the second observation the weights  $g(-1, \beta_1)$  are used. For  $t = 3, 4, \dots, n - 2$ , the weights  $g(0, \beta_1)$  are used. Denote the trend adjusted time series by  $X_t$ ,  $t = 1, 2, \dots, n$ . If  $3 \leq t \leq n - 2$  and  $3 \leq t + h \leq n - 2$ , then

$$\text{Cov}\{X_t, X_{t+h}\} = \begin{cases} 2520/4900, & h = 0, \\ -2016/4900, & h = \pm 1, \\ 1008/4900, & h = \pm 2, \\ -288/4900, & h = \pm 3, \\ 36/4900, & h = \pm 4, \\ 0 & \text{otherwise.} \end{cases}$$

For the first observation,

$$\text{Cov}\{X_1, X_t\} = \begin{cases} 560/4900, & t = 1, \\ -1260/4900, & t = 2, \\ 420/4900, & t = 3, \\ 252/4900, & t = 4, \\ -420/4900, & t = 5, \\ 204/4900, & t = 6, \\ -36/4900, & t = 7, \\ 0 & \text{otherwise.} \end{cases}$$

#### 9.4.2. Moving Averages of Integrated Time Series

In Chapter 10 we shall study nonstationary time series that can be represented as an autoregressive process with root of unit absolute value. We have been constructing moving average weights to remove a mean that is polynomial in time. We shall see that weights constructed to remove such a mean will also eliminate the nonstationarity arising from an autoregressive component with unit root. The time series  $\{W_t, t \in (0, 1, 2, \dots)\}$  is called an *integrated time series of order s* if it is defined by

$$W_t = \sum_{j_1=0}^t \cdots \sum_{j_2=0}^{j_3} \sum_{j_1=0}^{j_2} Z_{j_1},$$

where  $\{Z_t, t \in (0, 1, 2, \dots)\}$  is a stationary time series with zero mean and positive spectral density at  $\omega = 0$ .

**Theorem 9.4.2.** A moving average constructed to remove the mean will reduce a first order integrated time series to stationarity, and a moving average constructed to remove a linear trend will reduce a second order integrated time series to stationarity.

**Proof.** We first prove that a moving average constructed to remove the mean (zero degree polynomial) will reduce a first order integrated time series  $W_t = \sum_{s=0}^t Z_s$  to stationarity. Define the time series created by applying a moving average to remove the mean by

$$X_t = \sum_{j=1}^M c_j W_{t+j}, \quad t = 0, 1, 2, \dots,$$

where the weights satisfy

$$\sum_{j=1}^M c_j = 0.$$

Then

$$\begin{aligned} X_t &= \sum_{j=1}^M c_j \sum_{s=0}^{t+j} Z_s \\ &= \sum_{j=1}^M c_j \left( W_t + \sum_{r=1}^j Z_{t+r} \right) \\ &= \sum_{j=1}^M c_j \sum_{r=1}^j Z_{t+r}, \end{aligned}$$

which is a finite moving average of a stationary time series and therefore is stationary.

Consider next the second order integrated time series

$$U_t = \sum_{r=0}^t \sum_{s=0}^r Z_s = \sum_{r=0}^t W_r = \sum_{j=1}^{t+1} j Z_{t-j+1},$$

where  $W_r$  is a first order integrated time series. The weights  $d_j$  constructed to remove a linear trend satisfy  $\sum_{j=1}^M d_j = 0$  and  $\sum_{j=1}^M j d_j = 0$ . Therefore, the time series created by applying such weights is given by

$$\begin{aligned} X_t &= \sum_{j=1}^M d_j U_{t+j} = \sum_{j=1}^M d_j \left( U_t + \sum_{r=1}^j W_{t+r} \right) \\ &= \sum_{j=1}^M d_j \sum_{r=1}^j \left( W_t + \sum_{s=1}^r Z_{t+s} \right) \\ &= \sum_{j=1}^M d_j j W_t + \sum_{j=1}^M d_j \sum_{s=1}^j s Z_{t+j-s+1} \\ &= \sum_{j=1}^M \sum_{s=1}^j d_j s Z_{t+j-s+1}, \end{aligned}$$

which, once again, is a finite moving average of a stationary time series. ▲

The reader may extend this theorem to higher orders.

Moving averages are sometimes repeatedly applied to the same time series.

Theorem 9.4.3 can be used to show that the repeated application of a moving average constructed to remove a low order polynomial trend will remove a high order polynomial trend.

**Theorem 9.4.3.** Let  $p$  and  $q$  be integers,  $p \geq 0, q \geq 1$ . A moving average constructed to remove a  $p$ th degree polynomial trend will reduce a  $(p+q)$ th degree polynomial trend to degree  $q-1$ .

**Proof.** We write the trend function as

$$T_t = \sum_{r=0}^{p+q} b_r t^r.$$

If a moving average with weights  $\{w_j: j = -M_1, -M_1 + 1, \dots, M_2\}$  is applied to this function, we obtain

$$\begin{aligned} \sum_{j=-M_1}^{M_2} w_j T_{t+j} &= \sum_{j=-M_1}^{M_2} w_j \sum_{r=0}^{p+q} b_r (t+j)^r \\ &= \sum_{j=-M_1}^{M_2} w_j \sum_{r=0}^{p+q} b_r \sum_{k=0}^r \binom{r}{k} j^k t^{r-k}. \end{aligned}$$

Interchanging the order of summation and using the fact that a filter constructed to remove a  $p$ th degree polynomial trend satisfies

$$\sum_{j=-M_1}^{M_2} w_j j^r = 0, \quad r = 0, 1, 2, \dots, p,$$

we obtain the conclusion. ▲

### 9.4.3. Seasonal Adjustment

Moving averages have been heavily used in the analysis of time series displaying seasonal variation. Assume that  $Y_t$  is a monthly time series that can be represented locally by

$$Y_{t+j} = \alpha_t + \beta_t j + \sum_{m=1}^{12} \delta_{tm} D_{t+j,m} + Z_{t+j}, \quad (9.4.5)$$

where  $\alpha_t, \beta_t, \delta_{tm}$  are parameters,

$$D_{tm} = \begin{cases} 1 & \text{if } Y_t \text{ is observed in month } m, \\ 0 & \text{otherwise,} \end{cases}$$

$$\sum_{m=1}^{12} \delta_{im} = 0,$$

and  $Z_t$  is a stationary time series with zero mean. In this representation  $\alpha_i + \beta_j, j$  is the trend component and  $\sum_{m=1}^{12} \delta_{im} D_{t+j,m}$  is the seasonal component. Since the sum of the seasonal effects is zero over a period of 12 observations, it follows that a moving average such as

$$\hat{\tau}_t = \frac{1}{12} (\frac{1}{2} Y_{t-6} + Y_{t-5} + Y_{t-4} + \cdots + Y_{t+4} + Y_{t+5} + \frac{1}{2} Y_{t+6}) \quad (9.4.6)$$

or

$$\frac{1}{36} (\frac{1}{2} Y_{t-18} + Y_{t-17} + Y_{t-16} + \cdots + Y_{t+16} + Y_{t+17} + \frac{1}{2} Y_{t+18})$$

will not contain the seasonal component. Letting

$$\begin{aligned} d_0 &= \frac{11}{12}, \\ d_j &= -\frac{1}{12}, \quad j = \pm 1, \pm 2, \dots, \pm 5, \\ d_{-6} &= d_6 = -\frac{1}{24}, \end{aligned}$$

the difference

$$Y_t - \hat{\tau}_t = \sum_{m=1}^{12} \delta_{im} D_{im} + \sum_{j=-6}^6 d_j Z_{t+j} \quad (9.4.7)$$

constructed for the time series (9.4.5) contains the original seasonal component but no trend. A moving average of the time series  $Y_t - \hat{\tau}_t$  can then be used to estimate the seasonal component of the time series. For example,

$$\frac{1}{5} \sum_{i=-2}^2 (Y_{t+12i} - \hat{\tau}_{t+12i}) = \hat{S}_t \quad (9.4.8)$$

furnishes an estimator of the seasonal component for time  $t$  based on 5 years of data. The 12 values of  $\hat{S}_t$  computed for a year by this formula do not necessarily sum to zero. Therefore, one may modify the estimators to achieve a zero sum for the year by defining

$$\tilde{S}_{t(k)} = \hat{S}_t - \frac{1}{12} \sum_{j=-k+1}^{12-k} \hat{S}_{t+j}, \quad (9.4.9)$$

where  $\tilde{S}_{t(k)}$  is the seasonal component at time  $t$  that is associated with the  $k$ th month. The seasonally adjusted time series is then given by the difference  $Y_t - \tilde{S}_{t(k)}$ .

The seasonal adjustment procedure described above was developed from an additive model and the adjustment was accomplished with a difference. It is quite common to specify a multiplicative model and use ratios in the construction.

It is also possible to construct directly a set of weights using regression procedures and a model such as (9.4.5). To illustrate the procedure of seasonal adjustment based directly on a regression model, we assume that a quarterly time series can be represented for a period of 21 quarters by

$$Y_{t+j} = \sum_{r=0}^3 \beta_r j^r + \sum_{k=1}^4 \alpha_k D_{t+j,k} + Z_{t+j}, \quad j = 0, \pm 1, \dots, \pm 10,$$

where

$$D_{tk} = \begin{cases} 1 & \text{if } Y_t \text{ is observed in quarter } k, \\ 0 & \text{otherwise} \end{cases}$$

and

$$\sum_{k=1}^4 \alpha_k = 0.$$

We call  $\sum_{k=1}^4 \alpha_k D_{tk}$  the quarter (or seasonal) effect and  $\sum_{r=0}^3 \beta_r j^r$  the trend effect. The first seven columns of Table 9.4.2 give a  $\Phi$ -matrix for this problem. We have incorporated the restriction  $\sum_{k=1}^4 \alpha_k = 0$  by setting  $\alpha_4 = -\alpha_1 - \alpha_2 - \alpha_3$ .

With our coding of the variables, the trend value for  $t$ , the center observation, is given by  $\beta_0$ , and the seasonal value for  $t$  is given by  $\alpha_1$ . To compute the weights needed for the trend value at  $t$ , we regress the  $\beta_0$ -column on the remaining six columns, compute the deviations, and divide each deviation by the sum of squares of the deviations. It is readily verified that this is equivalent to computing the vector of weights by

$$\mathbf{w}'_{\beta_0} = \mathbf{J}_{\beta_0} (\Phi' \Phi)^{-1} \Phi',$$

where

$$\mathbf{J}_{\beta_0} = (1, 0, 0, 0, 0, 0, 0).$$

The weights for the seasonal component at time  $t$  can be calculated by regressing the column associated with  $\alpha_1$  on the remaining columns, computing the deviations, and dividing each deviation by the sum of squares. These operations are equivalent to computing

$$\mathbf{w}'_{\alpha_1} = \mathbf{J}_{\alpha_1} (\Phi' \Phi)^{-1} \Phi',$$

where

$$\mathbf{J}_{\alpha_1} = (0, 0, 0, 0, 1, 0, 0).$$

Let  $\mathbf{v}$  be a vector with 1 in the 11th ( $t$ th) position and zeros elsewhere. Then the weights for the trend and seasonally adjusted time series are given by

**Table 9.4.2. Calculation of Weights for the Trend and Seasonal Components of a Quarterly Time Series**

Index j	$\Phi$							Weights		
	$\beta_0$	$\beta_1$	$\beta_2$	$\beta_3$	$\alpha_1$	$\alpha_2$	$\alpha_3$	$w_{\beta_0}$	$w_{\alpha_1}$	$w_A$
-10	1	-10	100	-1000	0	0	1	-0.0477	-0.0314	0.0791
-9	1	-9	81	-729	-1	-1	-1	-0.0304	-0.0407	0.0711
-8	1	-8	64	-512	1	0	0	-0.0036	0.1562	-0.1526
-7	1	-7	49	-343	0	1	0	0.0232	-0.0469	0.0237
-6	1	-6	36	-216	0	0	1	0.0595	-0.0437	-0.0158
-5	1	-5	25	-125	-1	-1	-1	0.0634	-0.0515	-0.0119
-4	1	-4	16	-64	1	0	0	0.0768	0.1469	-0.2237
-3	1	-3	9	-27	0	1	0	0.0902	-0.0546	-0.0356
-2	1	-2	4	-8	0	0	1	0.1131	-0.0499	-0.0632
-1	1	-1	1	-1	-1	-1	-1	0.1036	-0.0562	-0.0474
0	1	0	0	0	1	0	0	0.1036	0.1438	0.7526
1	1	1	1	1	0	1	0	0.1036	-0.0562	-0.0474
2	1	2	4	8	0	0	1	0.1131	-0.0499	-0.0632
3	1	3	9	27	-1	-1	-1	0.0902	-0.0546	-0.0356
4	1	4	16	64	1	0	0	0.0768	0.1469	-0.2237
5	1	5	25	125	0	1	0	0.0634	-0.0515	-0.0119
6	1	6	36	216	0	0	1	0.0595	-0.0437	-0.0158
7	1	7	49	343	-1	-1	-1	0.0232	-0.0469	0.0237
8	1	8	64	512	1	0	0	-0.0036	0.1562	-0.1526
9	1	9	81	729	0	1	0	-0.0304	-0.0407	0.0711
10	1	10	100	1000	0	0	1	-0.0477	-0.0314	0.0791

$$w_A = v - w_{\beta_0} - w_{\alpha_1}.$$

The weights  $w_A$  can also be obtained by regressing  $v$  on  $\Phi$  and computing the deviations from regression.

#### 9.4.4. Differences

Differences have been used heavily when the objective is to reduce a time series to stationarity and there is little interest in estimating the mean function of the time series. Differences of the appropriate order will remove nonstationarity associated with locally polynomial trends in the mean and will reduce to stationarity integrated time series. For example, if  $Y_t$  is defined by

$$Y_t = \alpha + \beta t + W_t, \quad (9.4.10)$$

where

$$W_t = \sum_{j=0}^t Z_{t-j}$$

and  $Z_t$  is a stationary time series with zero mean, then

$$\Delta Y_t = Y_t - Y_{t-1} = \beta + Z_t$$

is a stationary time series.

The first difference operator can be viewed as a multiple of the moving average constructed to remove a zero degree polynomial trend from the second of two observations. The second difference is a multiple of the moving average constructed to remove a linear trend from the third of three observations. The second difference can also be viewed as a multiple of the moving average constructed to remove a linear trend from the second of three observations, and so forth. Therefore, the results of the previous subsections and of Section 2.4 may be combined in the following lemma.

**Lemma 9.4.1.** Let  $Y_t$  be a time series that is the sum of a polynomial trend of degree  $r$  and an integrated time series of order  $p$ . Then the  $q$ th difference of  $Y_t$ , where  $r \leq q$  and  $p \leq q$ , is a stationary time series. The mean of  $\Delta^q Y_t$  is zero for  $r \leq q - 1$ .

**Proof.** See Corollary 2.4.1 and Theorems 9.4.2 and 9.4.3. ▲

Differences of lag other than one are important in transforming time series. For the function  $f(t)$  defined on the integers, we define the *difference of lag  $H$*  by

$$\Delta_{(H)} f(t) = f(t) - f(t - H), \quad (9.4.11)$$

where  $H$  is a positive integer.

In Section 1.6 we defined a periodic function of period  $H$  with domain  $T$  to be a function satisfying

$$f(t + H) = f(t) \quad \forall t, \quad t + H \in T.$$

For  $T$  the set of integers and  $H$  a positive integer, the difference of lag  $H$  of a periodic function of period  $H$  is identically zero. Therefore, differences of lag  $H$  have been used to remove seasonal and other periodic components from time series. For example, if the expected value of a monthly time series is written as

$$E\{Y_t\} = \mu + \beta t + \sum_{i=1}^{12} \alpha_i D_{it}, \quad (9.4.12)$$

where

$$D_{it} = \begin{cases} 1 & \text{if } Y_t \text{ is observed in month } i, \\ 0 & \text{otherwise,} \end{cases}$$

$$\sum_{i=1}^{12} \alpha_i = 0,$$

then the expected value of  $\Delta_{(12)}Y_t$  is  $E\{Y_t - Y_{t-12}\} = 12\beta$ . The difference of lag 12 removed the periodic component and reduced the linear trend to a constant as well. Also, a difference of any finite lag will reduce a first order integrated time series to stationarity.

Mixtures of differences of different lags can be used. For example, if we take the first difference of the difference of lag 12 (or the difference of lag 12 of the first difference) of the time series  $Y_t$  of (9.4.12), the expected value of the resultant time series is zero; that is,  $E\{\Delta_{(12)}\Delta Y_t\} = 0$ .

The effect of repeated application of differences of different lags is summarized in Lemma 9.4.2.

**Lemma 9.4.2.** Let the time series  $Y_t$  be the sum of (1) a polynomial trend of degree  $r$ , (2) an integrated time series of order  $p$ , and (3) a sum of periodic functions of order  $H_1, H_2, \dots, H_q$ , where  $r \leq q$  and  $p \leq q$ . Then the difference  $X_t = \Delta_{(H_1)}\Delta_{(H_2)} \cdots \Delta_{(H_q)}Y_t$ , where  $1 \leq H_i < \infty$  for all  $i$ , is a stationary time series. If  $r \leq q-1$ , the mean of  $X_t$  is zero.

**Proof.** Reserved for the reader. ▲

## 9.5. STRUCTURAL MODELS

In Sections 9.1 and 9.2, we demonstrated the use of specified functions of time to approximate a mean function. In Section 9.4, we studied moving averages as local approximations to the mean function. Stochastic functions are also used as models for the means of time series. Models with stochastic means are sometimes called structural models in econometrics. See, for example, Harvey (1989). To introduce the approach, consider the simple model

$$\begin{aligned} Y_t &= \mu_t + e_t, \quad t = 1, 2, \dots, \\ \mu_t &= \mu_{t-1} + a_t, \end{aligned} \tag{9.5.1}$$

where  $(e_t, a_t)' \sim NI(\mathbf{0}, \text{diag}\{\sigma_e^2, \sigma_a^2\})$ . Models such as (9.5.1) are also called *unobserved components* models and can be considered a special case of (9.0.1). It follows from (9.5.1) that

$$X_t = \Delta Y_t = a_t + e_t - e_{t-1} \tag{9.5.2}$$

and hence  $X_t$  is a normally distributed stationary process with mean zero and autocovariances  $\gamma_X(0) = \sigma_a^2 + 2\sigma_e^2$ ,  $\gamma_X(1) = -\sigma_e^2$ , and  $\gamma_X(h) = 0$  for  $|h| > 1$ . That is,  $X_t$  is a first order moving average time series with representation

$$X_t = u_t + \beta u_{t-1}, \tag{9.5.3}$$

where  $\beta(1 + \beta^2)^{-1} = -(\sigma_a^2 + 2\sigma_e^2)^{-1}\sigma_e^2$ ,  $\beta \in [-1, 0)$ , and  $u_t$  are uncorrelated

with mean zero and variance  $\sigma_u^2 = (1 + \beta^2)^{-1}(2\sigma_e^2 + \sigma_a^2)$ . If we let  $\kappa = \sigma_e^{-2}\sigma_a^2$ , then  $\kappa = 2 - \beta^{-1}(1 + \beta)^2$  and  $\sigma_e^2 = (\kappa + 2)^{-1}(1 + \beta^2)\sigma_u^2$ . When  $\sigma_a^2 = 0$ , the model (9.5.1) reduces to the constant mean model and  $X_t = e_t - e_{t-1}$  is a noninvertible moving average.

Given a sample segment from a realization of  $Y_t$ , we can estimate the unknown parameters  $(\sigma_e^2, \sigma_a^2)$  by fitting a first order moving average to  $X_t = \Delta Y_t$ , with the parameter  $\beta$  restricted to  $[-1, 0)$ . The resulting estimates of  $\beta$  and  $\sigma_e^2$  are used to construct estimates of  $(\sigma_e^2, \sigma_a^2)$ . Given an estimate of  $(\sigma_e^2, \sigma_a^2)$ , estimates of the  $\mu_t$  can be obtained with filtering methods. Because of the form (9.5.1), it is natural to use the Kalman filter procedures of Section 4.6 to estimate the  $\mu_t$ .

It is also clear that a prediction constructed with estimates of  $\sigma_a^2$  and  $\sigma_e^2$  is identical to the prediction obtained by using the moving average representation for  $X_t$ .

A more general trend model is obtained by including a random change component. The expanded model is

$$\begin{aligned} Y_t &= \mu_t + e_{t1}, \\ \mu_t &= \zeta_{t-1} + \mu_{t-1} + e_{t2}, \\ \zeta_t &= \zeta_{t-1} + e_{t3}, \end{aligned} \tag{9.5.4}$$

where  $(e_{t1}, e_{t2}, e_{t3}) \sim NI(0, \text{diag}\{\sigma_1^2, \sigma_2^2, \sigma_3^2\})$ . The  $\zeta$ -component is a local linear trend.

If we take second differences of  $Y_t$ , we obtain

$$X_t = \Delta^2 Y_t = e_{t-1,3} + \Delta e_{t2} + \Delta^2 e_{t1}. \tag{9.5.5}$$

By the assumptions on  $e_t$ , we have

$$[\gamma_X(0), \gamma_X(1), \gamma_X(2)] = [\sigma_3^2 + 2\sigma_2^2 + 6\sigma_1^2, -\sigma_2^2 - 4\sigma_1^2, \sigma_1^2],$$

and  $\gamma_X(h) = 0$  for  $h > 2$ . Thus,  $X_t = \Delta^2 Y_t$  can be represented as a second order moving average

$$X_t = u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2}. \tag{9.5.6}$$

If we let  $\kappa_2 = \sigma_1^{-2}\sigma_2^2$  and  $\kappa_3 = \sigma_1^{-2}\sigma_3^2$ , then

$$\begin{aligned} \kappa_2 &= -4 - \rho_X^{-1}(2)\rho_X(1), \\ \kappa_3 &= \rho_X^{-1}(2) - 2\kappa_2 - 6, \\ \sigma_1^2 &= (\kappa_3 + 2\kappa_2 + 6)^{-1}\sigma_u^2(1 + \beta_1^2 + \beta_2^2). \end{aligned}$$

As with the simple model, the fact that variances are nonnegative restricts the possible values for  $(\beta_1, \beta_2)$ .

**Example 9.5.1.** To illustrate the use of the structural model to estimate the

trend, we use the data on wheat yields of Example 9.2.1. We consider the simple model

$$\begin{aligned} Y_t &= \mu_t + e_t, \\ \mu_t &= \mu_{t-1} + a_t, \end{aligned} \quad (9.5.7)$$

where  $(e_t, a_t)' \sim \text{II}(0, \text{diag}\{\sigma_e^2, \sigma_a^2\})$ . We let  $X_t = \Delta Y_t$  and fit the first order moving average of (9.5.3) to the first differences of the data to obtain

$$X_t = -0.4109 u_{t-1} + u_t, \quad (0.1010)$$

with  $\hat{\sigma}_u^2 = 4.4019$ . Eighty-three differences were used, and the zero mean model is estimated. Using

$$\begin{aligned} (1 + \beta^2)^{-1} \beta &= -(2\sigma_e^2 + \sigma_a^2)^{-1} \sigma_e^2, \\ (1 + \beta^2) \sigma_u^2 &= 2\sigma_e^2 + \sigma_a^2, \end{aligned}$$

we obtain  $(\hat{\sigma}_e^2, \hat{\sigma}_a^2) = (1.8086, 1.5278)$ . Using  $\hat{V}\{\hat{\sigma}_u^2\} = 0.4726$  and Taylor approximation methods, the estimated standard errors are 0.5267 and 0.5756 for  $\hat{\sigma}_e^2$  and  $\hat{\sigma}_a^2$ , respectively.

Given the estimates and the model, one can use filtering procedures to estimate the individual  $\mu_t$ . We use the covariance structure to construct a linear filter. If the process begins at time one with a fixed unknown initial value, denoted by  $\mu_1$ , the covariance matrix of the vector  $(Y_1, Y_2, \dots, Y_n)' = \mathbf{Y}$  is

$$\mathbf{V}_{nn} = \mathbf{I}\sigma_e^2 + \mathbf{L}\mathbf{L}'\sigma_a^2,$$

where  $\mathbf{L}$  is an  $n \times n$  lower triangular matrix with  $L_{ij} = 1$  for  $j < i$  and  $L_{ij} = 0$  for  $j \geq i$ . The best linear unbiased estimator of the unknown initial value is

$$\hat{\mu}_1 = (\mathbf{J}'\mathbf{V}_{nn}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{V}_{nn}^{-1}\mathbf{Y}, \quad (9.5.8)$$

where  $\mathbf{J}'$  is an  $n$ -dimensional row vector composed of all ones. Now, from (9.5.7),

$$\boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_n)' = \mu_1 \mathbf{J} + \mathbf{L}\mathbf{a},$$

where  $\mathbf{a} = (a_1, a_2, \dots, a_n)'$ . Therefore, the best unbiased estimator of  $\boldsymbol{\mu}$  is

$$\begin{aligned} \hat{\boldsymbol{\mu}} &= \mathbf{J}\hat{\mu}_1 + \sigma_a^2 \mathbf{L}\mathbf{L}'\mathbf{V}_{nn}^{-1}(\mathbf{Y} - \mathbf{J}\hat{\mu}_1) \\ &= \{\mathbf{J}(\mathbf{J}'\mathbf{V}_{nn}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{V}_{nn}^{-1} + \sigma_a^2 \mathbf{L}\mathbf{L}'\mathbf{V}_{nn}^{-1}[\mathbf{I} - \mathbf{J}(\mathbf{J}'\mathbf{V}_{nn}^{-1}\mathbf{J})^{-1}\mathbf{J}'\mathbf{V}_{nn}^{-1}]\}\mathbf{Y}. \end{aligned} \quad (9.5.9)$$

For fixed  $\sigma_a^2 \sigma_e^{-2}$ , the estimator is a linear function of  $\mathbf{Y}$ , say  $\mathbf{K}\mathbf{Y}$ , and the weights to be applied to the elements of  $\mathbf{Y}$  to estimate any particular  $\mu_t$  decline rapidly as

the distance between  $j$  and  $t$  increases. In equation (9.5.9),  $\hat{\mu}$  is an  $n$ -dimensional vector, but any subset of the original vector of observations can be used to construct a smaller vector of estimates. Using our estimates of  $\sigma_a^2$  and  $\sigma_e^2$ , the vector of optimal weights for estimating  $\mu_m$  using  $(Y_{m-5}, Y_{m-4}, \dots, Y_{m+4}, Y_{m+5})$  is

$$\mathbf{H}_m = (0.007, 0.013, 0.029, 0.071, 0.171, 0.418, 0.171, 0.071, 0.029, 0.013, 0.007).$$

Notice that the sum of the weights is one and that the weights are symmetric about the center weight. If  $\mathbf{K}$  denotes the matrix multiplying  $\mathbf{Y}$  in (9.5.9), the vector  $\mathbf{H}_m$  is the sixth row of the eleven by eleven matrix  $\mathbf{K}$ .

Figure 9.5.1 contains a plot of the estimated  $\mu_t$  values. Values for 1913 through 1986 were computed using  $\mathbf{H}_m$ . That is, the estimates are centered moving averages with weights  $\mathbf{H}_m$ . The end values were computed using the optimal estimator based on the first eleven and last eleven observations. The vector of estimated variances of the errors  $\hat{\mu}_t - \mu_t$  for the last six observations is

$$(0.755, 0.755, 0.757, 0.764, 0.808, 1.066).$$

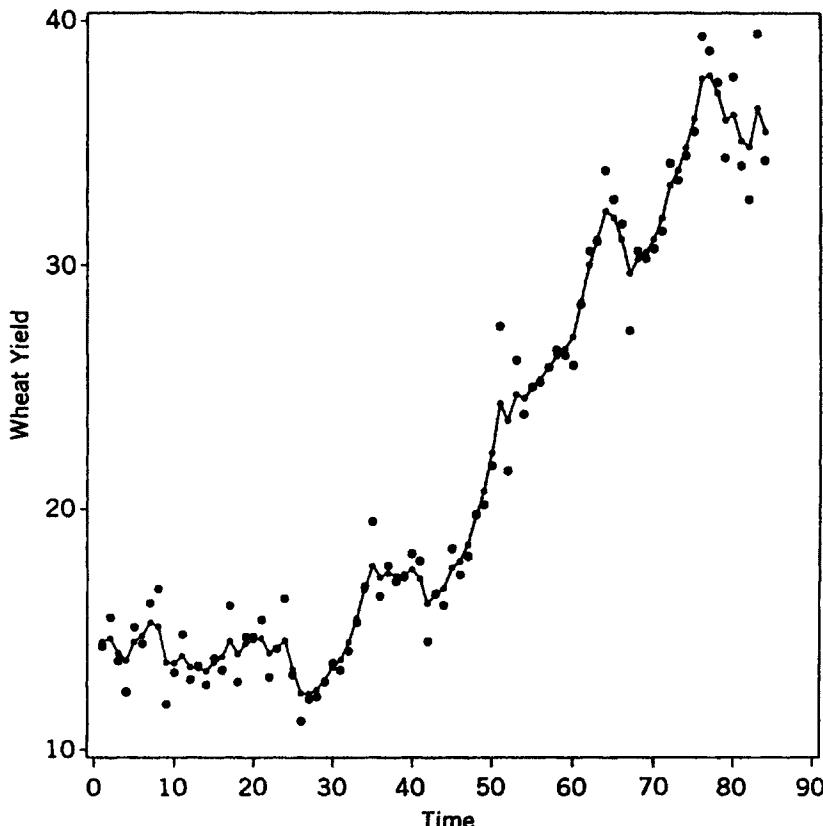


Figure 9.5.1. U.S. wheat yields 1908–1991 and structural trend estimates from model (9.5.7).

The variance for  $Y_{n-5}$  is the variance for all estimators in the center of the observation vector, that is, more than 5 years from the end points. The estimated variances are the last six diagonal elements of the matrix

$$\mathbf{L}\mathbf{L}'\hat{\sigma}_a^2 - \mathbf{K}\mathbf{L}\mathbf{L}'\hat{\sigma}_a^2 - \mathbf{L}\mathbf{L}'\mathbf{K}'\hat{\sigma}_a^2 + \mathbf{K}\hat{\mathbf{V}}_{nn}\mathbf{K}'.$$

The random walk model for  $\mu_t$  provides a rather flexible model for the mean function. As a result, the  $\hat{\mu}_t$  of Figure 9.5.1 form a much more variable estimated mean function than the grafted polynomial function of Example 9.2.1. The random walk model might be questioned for these data because of the rather extended period of increasing yields. Although one may not be satisfied with the final estimates of  $\mu_t$  obtained from the simple structural model, the estimates furnish useful information about the general movement of the time series.

Under the model (9.5.7), the predictor for  $\mu_{n+s}$ ,  $s > 0$ , is  $\hat{\mu}_n$ . The variance of the prediction error is  $V\{\hat{\mu}_{n+s} - \mu_{n+s}\} = V\{\hat{\mu}_n - \mu_n\} + s\sigma_a^2$ . Thus, the predictions of the mean for 1992, 1993, and 1994 are (35.46, 35.46, 35.46), respectively, and the estimated standard errors of the estimated means are (1.61, 2.03, 2.38). The predictions of the individual yields for 1992, 1993, and 1994 are also (35.46, 35.46, 35.46). The variances of the prediction errors are the variances of the estimated means increased by  $\sigma_a^2$ . The estimated standard errors of the prediction errors are (2.03, 2.38, 2.68).  $\blacktriangle\blacktriangle$

Example 9.5.1 illustrates the fact that the structural model can be used for preliminary investigation of a time series. We discussed the construction of moving averages of arbitrary length based on polynomial models in Section 9.4. By specifying a structural model for the trend, we also obtain a moving average estimator of the trend. The "length" of the moving average is based on estimates constructed with the data.

## 9.6. SOME EFFECTS OF MOVING AVERAGE OPERATORS

Linear moving averages have well-defined effects on the correlation and spectral properties of stationary time series. These were discussed earlier (see Theorems 2.1.1 and 4.3.1), but the effects of trend-removal filters are of sufficient importance to merit special investigation.

**Proposition 9.6.1.** Let  $X_t$  be a stationary time series with absolutely summable covariance function, and let  $Y_t = \sum_{j=-L}^M a_j X_{t-j}$ , where  $L$  and  $M$  are nonnegative integers and the weights  $a_j$  satisfy  $\sum_{j=-L}^M a_j = 0$ . Then the spectral density of  $Y_t$  evaluated at zero is zero; that is,  $f_Y(0) = 0$ .

**Proof.** By Theorem 4.3.1, the spectral density of  $Y_t$  is given by

$$f_Y(\omega) = (2\pi)^2 f_a(\omega) f_a^*(\omega) f_X(\omega),$$

where

$$2\pi f_a(\omega) = \sum_{j=-L}^M a_j e^{-j\omega j}.$$

Since  $\sum_{j=-L}^M a_j = 0$ , it follows that  $f_a(0) = 0$ .  $\blacktriangle$

Since  $f_y(0) = (2\pi)^{-1} \sum_{h=-\infty}^{\infty} \gamma_y(h)$ , it follows that the sum of the auto-covariances is zero for a stationary time series that has been filtered to remove the mean. For example, one may check that the covariances of  $X_t$  and  $X_{t+h}$ ,  $h = -4, -3, \dots, 4$ , for the example of Table 9.4.1 sum to zero.

The weights in the last column of Table 9.4.1 were constructed to remove a quadratic trend. The transfer function of that filter is

$$2\pi f_a(\omega) = (70)^{-1} (36 - 48 \cos \omega + 12 \cos 2\omega),$$

and the squared gain is

$$|2\pi f_a(\omega)|^2 = (70)^{-2} (36 - 48 \cos \omega + 12 \cos 2\omega)^2.$$

The squared gain is plotted in Figure 9.6.1. The squared gain is zero at zero and rises very slowly. Since the weights remove a quadratic trend, the filter removes much of the power from the spectral density at low frequencies. Since the spectral density of white noise is constant, the squared gain is a multiple of the spectral density of a moving average of uncorrelated random variables where the coefficients are given in Table 9.4.1.

The squared gain of the first difference operator is displayed in Figure 9.6.2. While it is of the same general appearance as the function of Figure 9.6.1, the squared gain of the first difference operator rises from zero more rapidly.

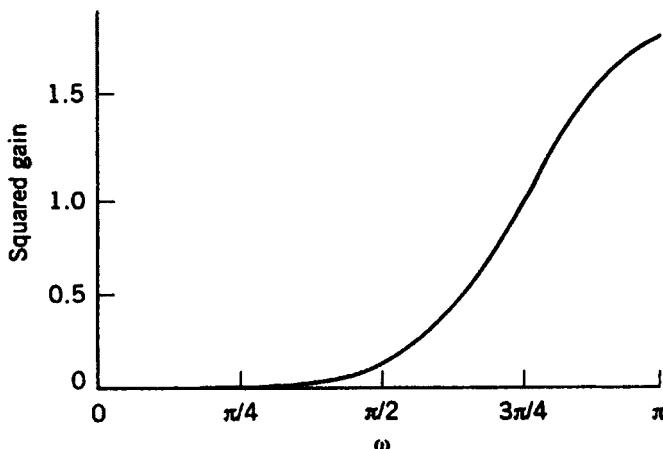


Figure 9.6.1. Squared gain of filter in Table 9.4.1.

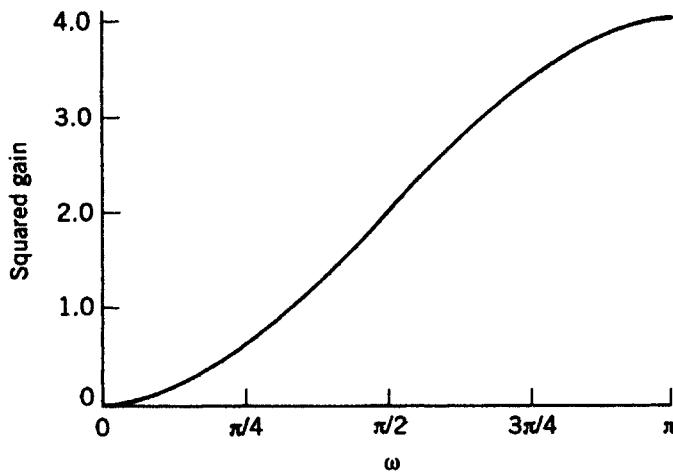


Figure 9.6.2. Squared gain of first difference operator.

If a time series contains a perfect sine component, the difference of the time series will also contain a perfect sine component of the same period, but of different amplitude and phase. That is, if  $Y_t = \sin \omega t$ , the difference is

$$\begin{aligned}\Delta Y_t &= \sin \omega t - \sin \omega(t-1) \\ &= 2 \sin \frac{1}{2} \omega \cos \omega(t - \frac{1}{2}).\end{aligned}$$

We note that the amplitude of the sine wave is changed. If  $\omega$  is such that  $|\sin \frac{1}{2} \omega| < \frac{1}{2}$ , the amplitude will be reduced. This will be true for long period waves. On the other hand, for short period waves,  $\pi/3 < \omega < \pi$ , the amplitude will be increased. Note that this agrees with Figure 9.6.2, which shows the transfer function to be greater than one for  $\omega > \pi/3$ .

Filtering a stationary time series with least squares weights to remove the seasonal effects will reduce the power of the spectral density to zero at the seasonal frequencies. For a time series with  $p$  observations per period of interest, the seasonal frequencies are defined to be  $2mp^{-1}\pi$ ,  $m = 1, 2, \dots, L[p]$ , where  $L[p]$  is the largest integer less than or equal  $p/2$ . For example, with a monthly time series, the seasonal frequencies are  $\pi/6, \pi/3, \pi/2, 2\pi/3, 5\pi/6$ , and  $\pi$ . We have not included the zero frequency because most seasonal adjustment schemes are not constructed to remove the mean.

**Proposition 9.6.2.** Let  $a_j$  be a least squares linear filter of length  $R$  constructed to remove seasonal variation from a time series with  $p$  ( $R \geq p$ ) observations per period of interest. Let the time series have an absolutely summable covariance function. Then the spectral density of the filtered time series is zero at the seasonal frequencies.

**Proof.** A linear least squares filter constructed to remove the seasonal effects

satisfies

$$\sum_{j=1}^R a_j \sin \frac{2\pi m}{p} j = 0, \\ \sum_{j=1}^R a_j \cos \frac{2\pi m}{p} j = 0, \quad m = 1, 2, \dots, L[p].$$

This is so because any periodic function of period  $p$  defined on the integers can be represented as a sum of  $p$  sines and cosines. We have

$$2\pi f_a(\omega) = \sum_{j=1}^R a_j e^{-i\omega j}$$

and, setting  $\omega = 2\pi m/p$ ,

$$2\pi f_a\left(\frac{2\pi m}{p}\right) = \sum_{j=1}^R a_j \left( \cos \frac{2\pi m}{p} j - i \sin \frac{2\pi m}{p} j \right) = 0$$

for  $m = 1, 2, \dots, L[p]$ . ▲

The first difference is a filter satisfying the conditions of Proposition 9.6.1, and the difference of lag  $p$  satisfies the conditions of Proposition 9.6.2. We now consider the effects of difference operators on autoregressive moving average time series.

**Proposition 9.6.3.** Let  $X_t$  be an autoregressive moving average time series of order  $(p, q)$  expressible as

$$\sum_{j=0}^p \alpha_j X_{t-j} = \sum_{r=0}^q \beta_r e_{t-r},$$

where  $\{e_t\}$  is a sequence of uncorrelated  $(0, \sigma^2)$  random variables. Let the roots of  $m^p + \alpha_1 m^{p-1} + \dots + \alpha_p = 0$  be less than one in absolute value, and let the roots of  $s^q + \beta_1 s^{q-1} + \dots + \beta_q = 0$  be  $s_1, s_2, \dots, s_q$ . Then the first difference  $Y_t = X_t - X_{t-1}$  is an autoregressive moving average  $(p, q+1)$  with the autoregressive portion unchanged and the roots of the moving average portion given by  $s_1, s_2, \dots, s_q, 1$ .

**Proof.** The spectral density of  $X_t$  is given by

$$f_X(\omega) = \frac{\sigma^2}{2\pi} \frac{\left( \sum_{j=0}^q \beta_j e^{ij\omega} \right) \left( \sum_{j=0}^q \beta_j e^{-ij\omega} \right)}{\left( \sum_{j=0}^p \alpha_j e^{ij\omega} \right) \left( \sum_{j=0}^p \alpha_j e^{-ij\omega} \right)}$$

$$= \frac{\sigma^2}{2\pi} \frac{\prod_{j=1}^q (e^{\omega} - s_j) \prod_{j=1}^q (e^{-\omega} - s_j)}{\prod_{j=1}^p (e^{\omega} - m_j) \prod_{j=1}^p (e^{-\omega} - m_j)}.$$

The spectral density of  $Y_t$  is given by

$$\begin{aligned} f_Y(\omega) &= (1 - e^{\omega})(1 - e^{-\omega})f_X(\omega) \\ &= (e^{\omega} - 1)(e^{-\omega} - 1)f_X(\omega) \\ &= \frac{\sigma^2}{2\pi} \frac{\prod_{j=1}^{q+1} (e^{\omega} - s_j) \prod_{j=1}^{q+1} (e^{-\omega} - s_j)}{\prod_{j=1}^p (e^{\omega} - m_j) \prod_{j=1}^p (e^{-\omega} - m_j)}, \end{aligned}$$

where  $s_{q+1} = 1$ . ▲

It follows immediately from Proposition 9.6.3 that the  $k$ th difference of the stationary autoregressive moving average time series of order  $(p, q)$  is an autoregressive moving average time series of order  $(p, q+k)$ , where at least  $k$  of the roots of the auxiliary equation associated with the moving average are one. Differences of lag  $r$  have similar effects on a time series.

**Proposition 9.6.4.** Given a time series  $X_t$  satisfying the assumptions of Proposition 9.6.3, the difference of lag  $r$  of  $X_t$  is an autoregressive moving average time series of order  $(p, q+r)$  where the autoregressive portion is unchanged and the roots of the moving average portion are  $s_1, s_2, \dots, s_q$  plus the  $r$  roots of the equation  $s^r = 1$ .

**Proof.** The spectral density of  $Y_t = X_t - X_{t-r}$  is given by

$$f_Y(\omega) = (1 - e^{-\omega r})(1 - e^{\omega r})f_X(\omega)$$

and, for example, the factor  $e^{-\omega r} - 1$  can be written as

$$\prod_{j=q+1}^{q+r} (e^{-\omega} - s_j),$$

where the  $s_j$ ,  $j = q+1, q+2, \dots, q+r$ , are the  $r$  roots of  $s^r = 1$ . ▲

These results have important practical implications. If one is attempting to identify an autoregressive moving average model for a time series that has been differenced, it is wise to consider a moving average of order at least as large as the order of differences applied to the original time series. If the time series was stationary before differencing, the characteristic polynomial of the moving average

portion of the differenced time series contains at least one unit root and the estimation theory of Section 8.3 is not applicable.

## 9.7. REGRESSION WITH TIME SERIES ERRORS

In this section we treat the problem of obtaining estimates of the parameters of regression equations with time series errors. We investigated the properties of the ordinary least squares estimators in Section 9.1 and found that for some special types of independent variables the ordinary least squares estimators are asymptotically fully efficient. For other independent variables, including most stationary time series, the ordinary least squares estimators are not efficient. Moreover, in most situations, the variance estimators associated with ordinary least squares are biased.

To illustrate some of these ideas, consider the simple model

$$\begin{aligned} Y_t &= \beta X_t + Z_t, \\ (Z_t, X_t) &= (\rho Z_{t-1}, \lambda X_{t-1}) + (e_t, u_t), \end{aligned} \quad (9.7.1)$$

where the  $e_t$  are normal independent  $(0, \sigma_e^2)$  random variables, the  $u_t$  are normal independent  $(0, \sigma_u^2)$  random variables,  $e_t$  is independent of  $u_j$  for all  $t, j$ ,  $|\rho| < 1$ , and  $|\lambda| < 1$ . Under the model (9.7.1), the expected value of

$$\hat{\beta}_S = \left( \sum_{t=1}^n X_t^2 \right)^{-1} \sum_{t=1}^n X_t Y_t$$

conditional on the observed  $X_t$ ,  $t = 1, 2, \dots, n$ , is  $\beta$ , and hence  $\hat{\beta}_S$  is unbiased.

The variance of the least squares estimator of  $\beta$  conditional on  $\mathbf{X} = (X_1, X_2, \dots, X_n)$  is given by equation (9.1.4),

$$V\{\hat{\beta}_S | \mathbf{X}\} = \left( \sum_{t=1}^n X_t^2 \right)^{-2} \sum_{t=1}^n \sum_{j=1}^n X_t X_j \rho^{|t-j|} \sigma_z^2,$$

where  $\sigma_z^2 = \gamma_Z(0)$ . Now the sample autocovariance of  $X_t$  converges to the population autocovariance, and

$$\begin{aligned} \operatorname{plim}_{n \rightarrow \infty} n^{-1} \sum_{t=1}^n \sum_{j=1}^n X_t X_j \rho^{|t-j|} \sigma_z^2 &= \lim_{n \rightarrow \infty} \sum_{h=-n+1}^{n-1} n^{-1} (n-|h|) \lambda^{|h|} \rho^{|h|} \sigma_x^2 \sigma_z^2 \\ &= (1 - \rho \lambda)^{-1} (1 + \rho \lambda) \sigma_x^2 \sigma_z^2, \end{aligned}$$

where  $\sigma_x^2 = \gamma_X(0)$ . Hence,

$$\operatorname{plim}_{n \rightarrow \infty} n V\{\hat{\beta}_S | \mathbf{X}\} = [(1 - \rho \lambda) \sigma_x^2]^{-1} (1 + \rho \lambda) \sigma_z^2.$$

If  $\rho$  is known, the variance of the generalized least squares estimator computed with known  $\rho$ , conditional on  $\mathbf{X}$ , is

$$V\{\hat{\beta}_G | \mathbf{X}\} = \left[ (1 - \rho^2) X_1^2 + \sum_{i=2}^n (X_i - \rho X_{i-1})^2 \right]^{-1} \sigma_e^2.$$

It follows that the large sample relative efficiency of generalized least squares to ordinary least squares is

$$p\lim_{n \rightarrow \infty} \frac{V\{\hat{\beta}_S | X_1, X_2, \dots, X_n\}}{V\{\hat{\beta}_G | X_1, X_2, \dots, X_n\}} = \frac{(1 + \rho\lambda)(1 - 2\rho\lambda + \rho^2)}{(1 - \rho\lambda)(1 - \rho^2)}.$$

This expression is greater than one for all nonzero  $(\rho, \lambda)$  less than one in absolute value. For  $\rho = \lambda = (0.5)^{1/2}$ , the relative efficiency is 300%!

The estimated variance of the ordinary least squares estimator obtained by the usual formulas is

$$\hat{V}_S\{\hat{\beta}_S\} = \left( \sum_{i=1}^n X_i^2 \right)^{-1} s^2,$$

where  $s^2 = (n - 1)^{-1} \sum_{i=1}^n (Y_i - \hat{\beta}_S X_i)^2$ .

Since  $V\{\hat{\beta}_S | \mathbf{X}\} = O_p(n^{-1})$ , we have  $\hat{\beta}_S - \beta = O_p(n^{-1/2})$  and  $s^2$  converges to the variance of  $Z_i$ . Therefore,

$$p\lim n \hat{V}_S\{\hat{\beta}_S\} = p\lim n \left( \sum_{i=1}^n X_i^2 \right)^{-1} s^2 = \sigma_x^{-2} \sigma_z^2$$

and

$$p\lim \left\{ \frac{\hat{V}_S\{\hat{\beta}_S\} - V\{\hat{\beta}_S | \mathbf{X}\}}{V\{\hat{\beta}_S | \mathbf{X}\}} \right\} = \frac{-2\rho\lambda}{1 + \rho\lambda}.$$

The common least squares estimator of variance will be an over- or underestimate, depending on the signs of  $\rho$  and  $\lambda$ . If  $\rho = \lambda = (0.5)^{1/2}$ , the ordinary least squares estimator will be estimating a quantity approximately one-third of the true variance. Given these results on the efficiency of ordinary least squares and on the bias in the ordinary least squares estimator of variance, it is natural to consider an estimator such as (9.1.3), where the estimated autocorrelations are used in place of the unknown parameters.

In Section 9.3 we established that the autocorrelations of the error time series could be estimated from the calculated regression residuals. We apply those results to our current model.

**Proposition 9.7.1.** Let  $Y_t$  satisfy

$$Y_t = \sum_{i=1}^r \varphi_{it} \beta_i + Z_t, \quad (9.7.2)$$

where  $Z_t$  is a stationary autoregressive process,

$$Z_t + \sum_{j=1}^p \alpha_j Z_{t-j} = e_t,$$

the roots of the characteristic equation are less than one in absolute value, the  $\varphi_{it}$  satisfy (9.1.7), (9.1.8), and (9.1.9), and  $\{e_t\}$  is a sequence of independent  $(0, \sigma^2)$  random variables with  $E[e_t^4] = \eta \sigma^4$ . Denote the simple least squares estimator of  $\beta$  by  $\hat{\beta} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_k)'$  and the simple least squares residuals by  $\hat{Z}_t$ .

Let  $\tilde{\alpha}$  be the estimator of  $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_p)'$  obtained by regressing  $\hat{Z}_t$  on  $(\hat{Z}_{t-1}, \hat{Z}_{t-2}, \dots, \hat{Z}_{t-p})$ ,  $t = p+1, p+2, \dots, n$ , and let  $\hat{\alpha}$  be the estimator obtained by regressing  $Z_t$  on  $(Z_{t-1}, Z_{t-2}, \dots, Z_{t-p})$ . Then

$$\tilde{\alpha} - \hat{\alpha} = O_p(n^{-1}).$$

**Proof.** The result is an immediate consequence of Theorem 9.3.1.  $\blacktriangle$

Given an estimator of the correlational structure of the error time series, we wish to construct the estimated generalized least squares estimator. Often the most expeditious way to obtain this estimator is to transform the data. In the present case, the Gram-Schmidt orthogonalization leads one to the transformed variables

$$\begin{aligned} \epsilon_1 &= \delta_{11} Z_1, \\ \epsilon_2 &= \delta_{22} Z_2 - \delta_{21} Z_1, \\ &\vdots \\ \epsilon_p &= \delta_{pp} Z_p - \sum_{j=1}^{p-1} \delta_{p,p-j} Z_{p-j}, \\ \epsilon_t &= e_t = Z_t + \sum_{j=1}^p \alpha_j Z_{t-j}, \quad t = p+1, p+2, \dots, n, \end{aligned} \quad (9.7.3)$$

where  $\delta_{11} = \gamma_z^{-1/2}(0)\sigma$ ,  $\delta_{22} = \{(1 - \rho_z^2(1))\gamma_z(0)\}^{-1/2}\sigma$ ,  $\delta_{21} = \rho_z(1)\{(1 - \rho_z^2(1))\gamma_z(0)\}^{-1/2}\sigma$ , etc. The  $\epsilon_t$  are uncorrelated with constant variance  $\sigma^2$ . For the first order autoregressive time series the transformed variables are

$$\begin{aligned} \epsilon_1 &= (1 - \rho^2)^{1/2} Z_1, \\ \epsilon_t &= Z_t - \rho Z_{t-1}, \quad t = 2, 3, \dots, n. \end{aligned}$$

As a second example of the transformation, consider the second order autoregressive process

$$Z_t = 1.53 Z_{t-1} - 0.66 Z_{t-2} + e_t,$$

where the variance of  $e_t$  is  $\sigma^2$ . Then by (2.5.8) the variance of  $Z_t$  is  $11.77\sigma^2$ . The correlations for the time series are given in Table 6.4.2. The transformation is

$$\epsilon_1 = (11.77)^{-1/2} Z_1,$$

$$\epsilon_2 = (1.7718)^{-1/2}(Z_2 - 0.9217Z_1),$$

$$\epsilon_t = e_t = Z_t - 1.53Z_{t-1} + 0.66Z_{t-2}, \quad t = 3, 4, \dots, n.$$

To define the estimated generalized least squares estimator, we can express our original model (9.7.1) in the matrix notation of Section 9.1 as

$$\mathbf{y} = \Phi \boldsymbol{\beta} + \mathbf{z}.$$

We let  $\mathbf{V}_{zz} = E\{\mathbf{z}\mathbf{z}'\}$  and  $\mathbf{z} = \mathbf{T}\mathbf{z}$ , where  $\mathbf{T}$  is the  $n \times n$  transformation matrix defined in (9.7.3). Then the estimated generalized least squares estimator is obtained by regressing  $\hat{\mathbf{T}}\mathbf{y}$  on  $\hat{\mathbf{T}}\Phi$ ,

$$\tilde{\boldsymbol{\beta}} = [\Phi' \hat{\mathbf{T}}' \hat{\mathbf{T}} \Phi]^{-1} \Phi' \hat{\mathbf{T}}' \hat{\mathbf{T}} \mathbf{y} = [\Phi' \hat{\mathbf{V}}_{zz}^{-1} \Phi]^{-1} \Phi' \hat{\mathbf{V}}_{zz}^{-1} \mathbf{y}, \quad (9.7.4)$$

where  $\hat{\mathbf{V}}_{zz}^{-1} = \hat{\mathbf{T}}' \hat{\mathbf{T}} \hat{\sigma}^{-2}$ ,  $\hat{\sigma}^2 = (n - 2p)^{-1} \sum_{j=p+1}^n (\hat{Z}_j + \sum_{j=1}^p \tilde{\alpha}_j \hat{Z}_{t-j})^2$ , and  $\hat{\mathbf{T}}$  is obtained from  $\mathbf{T}$  by replacing  $\alpha_j$  with  $\tilde{\alpha}_j$ ,  $\sigma^2$  with  $\hat{\sigma}^2$ , and so forth.

In Section 5.7 we demonstrated that the limiting distributions of the generalized least squares estimator and of the estimated generalized least squares estimator are the same under mild assumptions. By Theorem 5.7.4 and Proposition 9.7.1, the estimated generalized least squares estimator for an error process that is an autoregressive moving average will have the same limiting distribution as the generalized least squares estimator based on the true parameters, under mild assumptions. In Theorem 9.7.1, we give a direct proof of the result for autoregressive errors using the transformation  $\hat{\mathbf{T}}$ .

**Theorem 9.7.1.** Let  $Y_t$  satisfy the model (9.7.2), where the  $\varphi_{ii}$  satisfy the assumptions (9.1.7), (9.1.8), and (9.1.9). Then

$$\mathbf{D}_n(\hat{\boldsymbol{\beta}}_G - \tilde{\boldsymbol{\beta}}) = O_p(n^{-1/2}),$$

where  $\hat{\boldsymbol{\beta}}_G = [\Phi' \mathbf{V}_{zz}^{-1} \Phi]^{-1} \Phi' \mathbf{V}_{zz}^{-1} \mathbf{y}$ ,  $\tilde{\boldsymbol{\beta}}$  is defined in (9.7.4),  $d_{nii} = (\sum_{j=1}^n \varphi_{jj}^2)^{1/2}$ , and

$$\mathbf{D}_n = \text{diag}\{d_{n11}, d_{n22}, \dots, d_{nrr}\}.$$

**Proof.** We have, for  $\hat{\mathbf{T}}_t$  the  $t$ th row of  $\hat{\mathbf{T}}$  and  $\varphi_i$  the  $i$ th column of  $\Phi$ ,

$$\hat{\epsilon}_t = \hat{\mathbf{T}}_t \mathbf{z} = \epsilon_t + \sum_{j=1}^p (\tilde{\alpha}_j - \alpha_j) Z_{t-j},$$

$$\hat{w}_{ii} = \hat{\mathbf{T}}_t \varphi_i = \varphi_{ii} + \sum_{j=1}^p \alpha_j \varphi_{i-j,i} + \sum_{j=1}^p (\tilde{\alpha}_j - \alpha_j) \varphi_{i-j,i}$$

$$\stackrel{\text{def}}{=} w_{ii} + \sum_{j=1}^p (\tilde{\alpha}_j - \alpha_j) \varphi_{i-j,i}$$

for  $t = p+1, p+2, \dots, n$ , with similar expressions holding for  $t = 1, 2, \dots, p$ . By Proposition 9.7.1 and Theorem 8.2.1,  $\hat{\alpha} - \alpha = O_p(n^{-1/2})$ , and it follows that

$$\begin{aligned} d_{nii}^{-1} \sum_{t=1}^n \hat{\epsilon}_t \hat{w}_{ti} &= d_{nii}^{-1} \sum_{t=1}^n \epsilon_t w_{ti} + O_p(n^{-1/2}), \\ (d_{nii} d_{njj})^{-1} \sum_{t=1}^n \hat{w}_{ti} \hat{w}_{tj} &= (d_{nii} d_{njj})^{-1} \sum_{t=1}^n w_{ti} w_{tj} + O_p(n^{-1/2}). \end{aligned}$$

Therefore,

$$\begin{aligned} \mathbf{D}_n(\tilde{\beta} - \beta) &= (\mathbf{D}_n^{-1} \Phi' \hat{\mathbf{T}}' \hat{\mathbf{T}} \Phi \mathbf{D}_n^{-1})^{-1} \mathbf{D}_n^{-1} \Phi' \hat{\mathbf{T}}' \hat{\mathbf{T}} \mathbf{z} \\ &= (\mathbf{D}_n^{-1} \Phi' \mathbf{V}_{zz}^{-1} \Phi \mathbf{D}_n^{-1})^{-1} \mathbf{D}_n^{-1} \Phi' \mathbf{V}_{zz}^{-1} \mathbf{z} + O_p(n^{-1/2}) \\ &= \mathbf{D}_n(\hat{\beta}_G - \beta) + O_p(n^{-1/2}). \end{aligned} \quad \blacktriangle$$

Since the residual mean square for the regression of  $\hat{\mathbf{T}}\mathbf{y}$  on  $\hat{\mathbf{T}}\Phi$  converges in probability to  $\sigma^2$ , one can use the ordinary regression statistics for approximate tests and confidence intervals.

Because of the infinite autoregressive nature of invertible moving average time series, an exact transformation of the form (9.7.3) is cumbersome. However, the approximate difference equation transformation will be adequate for many purposes. For example, if

$$Z_t = e_t + b e_{t-1}, \quad |b| < 1,$$

the transformation

$$\begin{aligned} \epsilon_1 &= (1+b^2)^{-1/2} Z_1, \\ \epsilon_2 &= (1+b^2)^{1/2} (1+b^2+b^4)^{-1/2} [Z_2 - (1+b^2)^{-1} b Z_1], \\ \epsilon_t &= Z_t - b \epsilon_{t-1}, \quad t = 3, 4, \dots, n, \end{aligned} \quad (9.7.5)$$

will generally be satisfactory when  $b$  is not too close to one in absolute value.

**Example 9.7.1.** To illustrate the fitting of a regression model with time series errors, we use the data studied by Prest (1949). The data were discussed by Durbin and Watson (1951), and we use the data as they appear in that article. The data, displayed in Table 9.7.1, pertain to the consumption of spirits in the United Kingdom from 1870 to 1938. The dependent variable  $Y_t$  is the annual per capita consumption of spirits in the United Kingdom. The explanatory variables  $\varphi_{t1}$  and  $\varphi_{t2}$  are per capita income and price of spirits, respectively, both deflated by a general price index. All data are in logarithms. The model suggested by Prest can be written

$$Y_t = \beta_0 + \beta_1 \varphi_{t1} + \beta_2 \varphi_{t2} + \beta_3 \varphi_{t3} + \beta_4 \varphi_{t4} + Z_t,$$

Table 9.7.1. Annual Consumption of Spirits in the United Kingdom from 1870 to 1938

Year	Consumption $Y_t$	Income $\varphi_{t1}$	Price $\varphi_{t2}$	Year	Consumption $Y_t$	Income $\varphi_{t1}$	Price $\varphi_{t2}$
1870	1.9565	1.7669	1.9176	1905	1.9139	1.9924	1.9952
1871	1.9794	1.7766	1.9059	1906	1.9091	2.0117	1.9905
1872	2.0120	1.7764	1.8798	1907	1.9139	2.0204	1.9813
1873	2.0449	1.7942	1.8727	1908	1.8886	2.0018	1.9905
1874	2.0561	1.8156	1.8984	1909	1.7945	2.0038	1.9859
1875	2.0678	1.8083	1.9137	1910	1.7644	2.0099	2.0518
1876	2.0561	1.8083	1.9176	1911	1.7817	2.0174	2.0474
1877	2.0428	1.8067	1.9176	1912	1.7784	2.0279	2.0341
1878	2.0290	1.8166	1.9420	1913	1.7945	2.0359	2.0255
1879	1.9980	1.8041	1.9547	1914	1.7888	2.0216	2.0341
1880	1.9884	1.8053	1.9379	1915	1.8751	1.9896	1.9445
1881	1.9835	1.8242	1.9462	1916	1.7853	1.9843	1.9939
1882	1.9773	1.8395	1.9504	1917	1.6075	1.9764	2.2082
1883	1.9748	1.8464	1.9504	1918	1.5185	1.9965	2.2700
1884	1.9629	1.8492	1.9723	1919	1.6513	2.0652	2.2430
1885	1.9396	1.8668	2.0000	1920	1.6247	2.0369	2.2567
1886	1.9309	1.8783	2.0097	1921	1.5391	1.9723	2.2988
1887	1.9271	1.8914	2.0146	1922	1.4922	1.9797	2.3723
1888	1.9239	1.9166	2.0146	1923	1.4606	2.0136	2.4105
1889	1.9414	1.9363	2.0097	1924	1.4551	2.0165	2.4081
1890	1.9685	1.9548	2.0097	1925	1.4425	2.0213	2.4081
1891	1.9727	1.9453	2.0097	1926	1.4023	2.0206	2.4367
1892	1.9736	1.9292	2.0048	1927	1.3991	2.0563	2.4284
1893	1.9499	1.9209	2.0097	1928	1.3798	2.0579	2.4310
1894	1.9432	1.9510	2.0296	1929	1.3782	2.0649	2.4363
1895	1.9569	1.9776	2.0399	1930	1.3366	2.0582	2.4552
1896	1.9647	1.9814	2.0399	1931	1.3026	2.0517	2.4838
1897	1.9710	1.9819	2.0296	1932	1.2592	2.0491	2.4958
1898	1.9719	1.9828	2.0146	1933	1.2635	2.0766	2.5048
1899	1.9956	2.0076	2.0245	1934	1.2549	2.0890	2.5017
1900	2.0000	2.0000	2.0000	1935	1.2527	2.1059	2.4958
1901	1.9904	1.9939	2.0048	1936	1.2763	2.1205	2.4838
1902	1.9752	1.9933	2.0048	1937	1.2906	2.1205	2.4636
1903	1.9494	1.9797	2.0000	1938	1.2721	2.1182	2.4580
1904	1.9332	1.9772	1.9952				

SOURCE: Durbin and Watson (1951). Reproduced with permission of the Trustees of Biometrika and of the authors.

where 1869 is the origin for  $t$ ,  $\varphi_{t3} = 10^{-2}t$ ,  $\varphi_{t4} = 10^{-4}(t - 35)^2$ , and we assume  $Z_t$  is a stationary time series.

The ordinary least squares regression equation is

$$\hat{Y}_t = 2.14 + 0.69 \varphi_{t1} - 0.63 \varphi_{t2} - 0.95 \varphi_{t3} - 1.15 \varphi_{t4}. \\ (0.28) \quad (0.14) \quad (0.05) \quad (0.09) \quad (0.16)$$

The residual mean square is 0.00098. The numbers in parentheses are the standard errors computed by the ordinary regression formulas and would be a part of the output in any standard regression computer program. They are *not* proper estimators of the standard errors when the error in the equation is autocorrelated. We shall return to this point.

The Durbin-Watson  $d$  is 0.5265. Under the null hypothesis of normal independent errors, the expected value of  $d$  is

$$E\{d\} = [2(68) - 0.67](64)^{-1} = 2.1145,$$

where

$$\text{tr}\{(\Delta\Phi)'(\Delta\Phi)(\Phi'\Phi)^{-1}\} = 0.67.$$

Therefore, from equations (9.3.8) and (9.3.9),

$$\begin{aligned}\hat{\rho} &= 0.7367 + (0.5)(0.1145)(66)(65)^{-1}(1 - 0.5427) \\ &= 0.7633\end{aligned}$$

and

$$t_d = (66)^{1/2}(0.7633)(0.4174)^{-1/2} = 9.60.$$

Obviously, the hypothesis of independent errors is rejected. To investigate the nature of the autocorrelation, we fit several autoregressive models to the regression residuals. The results are summarized in Table 9.7.2. The statistics of that table are consistent with the hypothesis that the error time series is a first order autoregressive process. Therefore, we construct the transformation of (9.7.3) for a first order autoregressive process with  $\hat{\rho} = 0.7633$ . The transformation matrix is

$$\hat{T} = \begin{pmatrix} 0.6460 & 0 & 0 & \cdots & 0 \\ -0.7633 & 1 & 0 & \cdots & 0 \\ 0 & -0.7633 & 1 & \cdots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{pmatrix},$$

where  $\hat{T}$  is a  $69 \times 69$  matrix and  $0.6460 = (1 - \hat{\rho}^2)^{1/2} = [1 - (0.7633)^2]^{1/2}$ .

Regressing the transformed  $Y_t$  on the transformed  $\varphi_i$ ,  $i = 0, 1, \dots, 4$ , we obtain the estimated generalized least squares equation

$$\hat{Y}_t = 2.36 + 0.72 \varphi_{t1} - 0.80 \varphi_{t2} - 0.81 \varphi_{t3} - 0.92 \varphi_{t4}. \quad (0.30) \quad (0.15) \quad (0.07) \quad (0.11) \quad (0.27)$$

The numbers in parentheses are consistent estimates of the standard errors of the

**Table 9.7.2. Analysis of Variance for Residuals from Spirit Consumption Regression**

Source	Degrees of Freedom	Mean Square
$\hat{Z}_{t-1}$	1	0.03174
$\hat{Z}_{t-2}$ after $\hat{Z}_{t-1}$	1	0.00036
$\hat{Z}_{t-3}$ after $\hat{Z}_{t-1}, \hat{Z}_{t-2}$	1	0.00013
$\hat{Z}_{t-4}$ after $\hat{Z}_{t-1}, \hat{Z}_{t-2}, \hat{Z}_{t-3}$	1	0.00004
Error	56	0.00049

estimators. They are computed from the transformed data by the usual regression formulas. The error mean square for the transformed regression is 0.000417.

Every one of the estimated standard errors for the generalized least squares estimator based on  $\hat{\rho} = 0.7633$  is greater than the standard error calculated by the usual formulas for the simple least squares estimates. Of course, we know that the standard formulas give biased estimators for simple least squares when the errors are autocorrelated. To demonstrate the magnitude of this bias, we estimate the variance of the simple least squares estimators, assuming that the error time series is

$$Z_t = 0.7633Z_{t-1} + e_t,$$

where  $e_t$  is a sequence of uncorrelated (0, 0.000417) random variables. The covariance matrix of the simple least squares estimator is estimated by  $(\Phi'\Phi)^{-1}\Phi'\hat{V}_{zz}\Phi(\Phi'\Phi)^{-1}$ , where  $\Phi$  is the  $69 \times 5$  matrix of observations on the independent variables and  $\hat{V}_{zz}$  is the estimated covariance matrix for 69 observations from a first order autoregressive time series with  $\hat{\rho} = 0.7633$ .

Alternative estimates of the standard errors of the coefficients are compared in Table 9.7.3. The consistent estimates of the standard errors of simple least squares obtained from the matrix  $(\Phi'\Phi)^{-1}\Phi'\hat{V}_{zz}\Phi(\Phi'\Phi)^{-1}$  are approximately twice those obtained from the simple formulas. This means that the variance estimates

**Table 9.7.3. Comparison of Alternative Estimators of Standard Errors of Estimators for Spirits Example**

Coefficient	Estimator for Simple Least Squares Assuming $\rho = 0$	Consistent Estimator for Simple Least Squares	Consistent Estimator for Generalized Least Squares
$\beta_0$	0.282	0.545	0.304
$\beta_1$	0.138	0.262	0.146
$\beta_2$	0.054	0.112	0.072
$\beta_3$	0.088	0.179	0.108
$\beta_4$	0.161	0.313	0.266

computed under the assumption of zero correlation are approximately one-fourth of the consistent estimates. The estimated standard errors for the generalized least squares estimators range from about one-half (0.56 for income) to nine-tenths (0.85 for time squared) of those for simple least squares. This means that the estimated efficiency of simple least squares relative to generalized least squares ranges from about three-tenths to about seven-tenths.

The computations for regression models with autoregressive errors are available in several computer programs. The procedure AUTOREG of SAS Institute, Inc. (1989) is one convenient program. If we specify a first order autoregressive error and the Yule-Walker procedure of estimating the autoregressive process, we obtain an estimate  $\hat{\rho} = 0.7160$  with a standard error of 0.0879. The estimate of  $\rho$  differs from ours because of the different estimator used. The estimated regression equation is nearly identical to the one given above.  $\blacktriangle \blacktriangle$

In Example 9.7.1 we gave details of a two-step computational procedure for estimation of a regression equation with time series errors. Computer programs offer one the option of specifying and estimating the complete model, either by the two-step procedure or as a single maximization problem.

Assume the time series error in a linear regression model to be an autoregressive process. Thus, we write

$$Y_t = \varphi_t \boldsymbol{\beta} + Z_t, \quad t = 1, 2, \dots, n, \quad (9.7.6a)$$

$$Z_t + \sum_{i=1}^p \alpha_i Z_{t-i} = e_t, \quad (9.7.6b)$$

where  $\{e_t\}$  is a sequence of  $N(0, \sigma^2)$  random variables, and  $\varphi_t$ ,  $t = 1, 2, \dots, n$ , are  $k$ -dimensional row vectors of explanatory variables. If the process  $Z_t$  is a stationary normal process, the log likelihood is

$$L(\mathbf{Y}; \boldsymbol{\theta}) = -0.5[n \log 2\pi + \log |\mathbf{V}_{zz}| + (\mathbf{Y} - \Phi \boldsymbol{\beta})' \mathbf{V}_{zz}^{-1} (\mathbf{Y} - \Phi \boldsymbol{\beta})], \quad (9.7.7)$$

where  $\mathbf{V}_{zz} = \mathbf{V}_{zz}(\boldsymbol{\alpha}, \sigma^2)$  is the  $n \times n$  covariance matrix of  $(Z_1, Z_2, \dots, Z_n)$ ,  $\mathbf{Y}$  is the  $n$ -dimensional column vector  $(Y_1, Y_2, \dots, Y_n)'$ ,  $\Phi$  is the  $n \times k$  matrix  $(\varphi_1', \varphi_2', \dots, \varphi_n')'$ , and  $\boldsymbol{\theta}' = (\boldsymbol{\beta}', \alpha_1, \alpha_2, \dots, \alpha_p, \sigma^2)$ .

The theorems of Section 5.5 can be used to obtain the limiting distribution of the maximum likelihood estimator. The limiting distribution of the maximum likelihood estimator of  $\boldsymbol{\beta}$  is the same as the limiting distribution of the generalized least squares estimator, if the latter exists.

**Theorem 9.7.2.** Let  $Y_t$  satisfy equation (9.7.6a), where  $Z_t$  is a stationary process with the roots of the characteristic equation associated with (9.7.6b) less than one in absolute value. Assume the  $e_t$  are iid  $(0, \sigma^2)$  random variables or that the  $e_t$  satisfy the martingale difference conditions of Corollary 5.3.4. Let  $\{\varphi_t\}$  and  $\{\mathbf{M}_n\}$  be sequences such that

$$\lim_{n \rightarrow \infty} \mathbf{M}_n^{-1} \Phi' \mathbf{V}_{zz}^{-1} \Phi \mathbf{M}_n^{-1} = \mathbf{I}$$

and

$$\lim_{n \rightarrow \infty} \sup_{1 \leq t \leq n} |\mathbf{M}_n^{-1} \varphi_t|^2 = 0.$$

Let  $\hat{\theta}$  be the estimator that maximizes the log Gaussian likelihood (9.7.7), and let  $\theta^0$  be the true parameter vector. Then

$$\begin{aligned} \mathbf{M}'_n(\hat{\beta} - \beta^0) &\xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}), \\ n^{1/2}(\hat{\alpha} - \alpha^0) &\xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{A}^{-1} \sigma^2), \end{aligned}$$

$\hat{\sigma}^2 \xrightarrow{P} \sigma^2$ , and  $\mathbf{M}'_n(\hat{\beta} - \beta^0)$  is independent of  $n^{1/2}(\hat{\alpha} - \alpha^0)$  in the limit, where  $\mathbf{A} = E\{(Z_{t-1}, Z_{t-2}, \dots, Z_{t-p})'(Z_{t-1}, Z_{t-2}, \dots, Z_{t-p})\}$ .

**Proof.** Omitted. ▲

**Example 9.7.2.** We use the wheat yield data of Example 9.2.1 for the period 1908 to 1991 to illustrate the nonlinear fitting. The  $\hat{\mu}$ , from Example 9.5.1 and a plot of the data suggest that wheat yields were nearly constant for the first 25 years and for the last several, say five, years. Therefore, we consider a mean function that is constant for the first 25 years and for the last five years. A function that is continuous with continuous first derivative that meets these requirements is

$$\varphi_{12} = \begin{cases} 0, & 1 \leq t \leq 25, \\ (t - 25)^2, & 25 \leq t \leq 54, \\ 841 + 58(t - 54), & 54 \leq t \leq 70, \\ 841 + 58(t - 54) - 2.9(t - 70)^2, & 70 \leq t \leq 80, \\ 2059, & t \geq 80, \end{cases}$$

where  $t = 1$  for 1908. The first part of the function was used in Example 9.2.1. We assume the errors follow a first order autoregressive process. Then our model is

$$Y_t = \beta_0 + \beta_1 \varphi_{12} + Z_t \quad (9.7.8)$$

for  $t = 1, 2, \dots, n$ , where  $Z_t$  is the stationary process satisfying

$$Z_t + \alpha_1 Z_{t-1} = e_t,$$

and we assume the  $e_t$  are iid( $0, \sigma_e^2$ ).

If we use procedure AUTOREG of SAS/ETS<sup>0</sup> to compute the Gaussian maximum likelihood estimates, we obtain

$$\begin{array}{llll} \hat{\beta}_0 = 14.257, & \hat{\beta}_1 = 0.01075, & \hat{\alpha}_1 = -0.292, & \hat{\sigma}_e^2 = 3.427. \\ (0.380) & (0.00037) & (0.107) & \end{array}$$

The model predictions for the next three observations are

$$(\hat{Y}_{1992}, \hat{Y}_{1993}, \hat{Y}_{1994}) = (35.78, \quad 36.21, \quad 36.34). \\ (1.95) \quad (2.05) \quad (2.07)$$

The prediction for 1992 is similar to that obtained in Example 9.5.1. The model of Example 9.5.1 postulates a random walk for the mean function and uncorrelated deviations from the mean function. Hence, the predictions for all future periods are the best estimate of the yield for 1991, and the estimate for 1991 is heavily influenced by the actual observation for 1991. The predictions under the model (9.7.8) are based on the estimated trend associated with the function  $\varphi_{t_2}$ . Because the function is constant for  $t \geq 80$ , the predictions ultimately fall on a horizontal line. Because the yield for 1991 is about 2.1 bushels below the trend line, the prediction for 1992 is about  $(0.29)^2 \times 2.1 = 0.61$  bushels below the trend line, the prediction for 1993 is about  $(0.29)^3 \times 2.1 = 0.18$  bushels below the trend line, and the prediction for 1994 is about  $(0.29)^4 \times 2.1 = 0.05$  bushels below the trend line. Because the 1991 yield is below the trend line and the estimated autocorrelation is positive, the predictions for the next three years increase toward the trend line.

The standard errors of the predictions constructed in this example are smaller than those of the predictions of Example 9.5.1. In this example, the mean is assumed to be of a known functional form. In Example 9.5.1, the mean is assumed to be random walk. If the mean function is known, then it is easier to predict than the random walk. However, in using the estimated trend function for prediction, one projects the trend function into the future. Often practitioners hesitate to project polynomial trends very far into the future. In our current example, the most recent past of the mean function is assumed to be constant, and we might be less worried about our trend assumption than we would have been in constructing predictions in 1971 using the function of Example 9.2.1. To project linearly in 1971, one would need to believe that the improvements in varieties and pesticides and the use of fertilizer that led to increased yields during the observation period would continue at the same rate during the prediction period. In the case of wheat yields the linear trend predictions would have performed well for about fifteen years, from 1971 until about 1986. ▲▲

**Example 9.7.3.** As a second illustration of nonlinear estimation, we use the data on consumption of spirits of Table 9.7.1. If we use nonlinear least squares and fit the model with a first order autoregressive error to the last  $n - 1$  observations, we obtain

$$\hat{Y}_t = 2.421 + 0.716 \varphi_{t_1} - 0.818 \varphi_{t_2} - 0.846 \varphi_{t_3} - 0.560 \varphi_{t_4}, \\ (0.304) \quad (0.149) \quad (0.074) \quad (0.119) \quad (0.380)$$

$$\hat{Z}_t = 0.788 Z_{t-1}, \\ (0.078)$$

and  $\hat{\sigma}^2 = 0.00041$ , where  $\varphi_{i1} = 10^{-2}t$  and  $\varphi_{i4} = 10^{-4}(t - 35)^2$ . If we fit the full model by maximum likelihood, we obtain

$$\begin{aligned}\hat{Y}_t &= 2.391 + 0.727 \varphi_{i1} - 0.822 \varphi_{i2} - 0.773 \varphi_{i3} - 0.932 \varphi_{i4}, \\ &\quad (0.308) \quad (0.150) \quad (0.075) \quad (0.113) \quad (0.300)\end{aligned}$$

$$\hat{Z}_t = 0.812 Z_{t-1},$$

$$\quad (0.076)$$

and  $\hat{\sigma}^2 = 0.00042$ . As expected, the results are very similar to those obtained in Example 9.7.1 because the three procedures have the same asymptotic efficiency.

If we fit the model with a second order autoregressive error by maximum likelihood, we obtain

$$\begin{aligned}\hat{Y}_t &= 2.473 + 0.704 \varphi_{i1} - 0.832 \varphi_{i2} - 0.845 \varphi_{i3} - 0.470 \varphi_{i4}, \\ &\quad (0.333) \quad (0.156) \quad (0.077) \quad (0.127) \quad (0.420)\end{aligned}$$

$$\hat{Z}_t = 0.742 Z_{t-1} + 0.054 Z_{t-2},$$

$$\quad (0.140) \quad (0.139)$$

and  $\hat{\sigma}^2 = 0.00042$ . The second order coefficient in the autoregressive model is small relative to its standard error, agreeing with the results of Table 9.7.2.  $\blacktriangle\blacktriangle$

Regression models in which the current value of a time series  $Y_t$  is a function of current and lagged values of a time series  $X_t$  are sometimes called *transfer function* models. The models are often written in terms of the backshift operator. Thus, the simple model

$$Y_t = \beta_1 X_t + \beta_2 X_{t-1} + e_t$$

can be written as

$$Y_t = (\beta_1 + \beta_2 \mathcal{B}) X_t + e_t.$$

More complicated models may contain ratios of polynomials in the backshift operator. For example,

$$\begin{aligned}Y_t &= [A(\mathcal{B})]^{-1} B(\mathcal{B}) X_t + e_t, \\ &= \sum_{i=0}^{\infty} c_i X_{t-i} + e_t,\end{aligned}$$

where  $A(\mathcal{B}) = 1 - a_1 \mathcal{B} - \cdots - a_r \mathcal{B}^r$ ,  $B(\mathcal{B}) = b_1 - b_2 \mathcal{B} - \cdots - b_s \mathcal{B}^s$ , the  $c_i$  are the coefficients defined by  $[A(\mathcal{B})]^{-1} B(\mathcal{B})$ , and it is assumed that the roots of  $A(\mathcal{B}) = 0$  are greater than one so that the coefficients converge to zero. The set of coefficients  $(c_0, c_1, \dots)$  is called the *impulse response function*. Such models fall into the class of nonlinear regression models covered by the theory of Section 5.5, provided  $X_t$  is reasonably well behaved.

## 9.8 REGRESSION EQUATIONS WITH LAGGED DEPENDENT VARIABLES AND TIME SERIES ERRORS

In the classical regression problem the error in the equation is assumed to be distributed independently of the independent variables in the equation. Independent variables with this property we call *ordinary*. In our investigation of the estimation of the autoregressive parameters in Chapter 8, we noted that the autoregressive representation had the appearance of a regression problem, but that the vector of errors was not independent of the vector of lagged values of the time series. A generalization of the autoregressive model is one containing ordinary independent variables and also lagged values of the dependent variable as explanatory variables.

Consider the model

$$\begin{aligned} Y_t &= \sum_{i=1}^k \varphi_{it} \zeta_i - \sum_{j=1}^p \alpha_j Y_{t-j} + e_t \\ &= \mathbf{X}_t \boldsymbol{\theta} + e_t \end{aligned} \quad (9.8.1)$$

for  $t = 1, 2, \dots$ , where  $\mathbf{X}_t = (\varphi_{t1}, \varphi_{t2}, \dots, \varphi_{tk}, Y_{t-1}, Y_{t-2}, \dots, Y_{t-p})$ ,  $\boldsymbol{\theta} = (\zeta_1, \zeta_2, \dots, \zeta_k, -\alpha_1, -\alpha_2, \dots, -\alpha_p)$ , the  $\varphi_{it}$  are fixed constants, and  $e_t$  are uncorrelated  $(0, \sigma^2)$  random variables. The characteristic equation associated with equation (9.8.1) is

$$m^p + \sum_{j=1}^p \alpha_j m^{p-j} = 0. \quad (9.8.2)$$

Such models are common in economics. See Judge *et al.* (1985, Chapter 10). Ordinary least squares is a natural way to estimate the parameters of equation (9.8.1). Let

$$\hat{\boldsymbol{\theta}} = \left( \sum_{i=1}^n \mathbf{X}'_i \mathbf{X}_i \right)^{-1} \sum_{i=1}^n \mathbf{X}'_i \mathbf{Y}_i \quad (9.8.3)$$

denote the least squares estimator. The limiting distribution of the properly standardized  $\hat{\boldsymbol{\theta}}$  is normal under mild assumptions.

**Theorem 9.8.1.** Let  $Y_t$  satisfy equation (9.8.1), where  $Y_0, Y_{-1}, \dots, Y_{-p+1}$  are fixed. Assume

$$E\{(e_t, e_t^2) | \mathcal{A}_{t-1}\} = (0, \sigma^2) \quad \text{a.s.}$$

and

$$E\{|e_t|^{2+\nu} | \mathcal{A}_{t-1}\} < L < \infty \quad \text{a.s.}$$

for all  $t$  and some  $\nu > 0$ , where  $\mathcal{A}_{t-1}$  is the sigma-field generated by

$(\mathbf{X}_t, \mathbf{X}_{t-1}, \dots, \mathbf{X}_1)$ . Assume there exists a sequence  $\{\mathbf{M}_n\}$  of fixed nonsingular matrices such that

$$\operatorname{plim}_{n \rightarrow \infty} \mathbf{M}_n^{-1} \sum_{t=1}^n \mathbf{X}'_t \mathbf{X}_t \mathbf{M}_n^{-1'} = \lim_{n \rightarrow \infty} E \left\{ \mathbf{M}_n^{-1} \sum_{t=1}^n \mathbf{X}'_t \mathbf{X}_t \mathbf{M}_n^{-1'} \right\} = \mathbf{I} \quad (9.8.4)$$

and

$$\lim_{n \rightarrow \infty} \sup_{1 \leq t \leq n} E\{|\mathbf{M}_n^{-1} \mathbf{X}'_t|^2\} = 0. \quad (9.8.5)$$

Let  $\hat{\boldsymbol{\theta}}$  be defined by (9.8.3). Then

$$\mathbf{M}'_n (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}).$$

**Proof.** Let  $\boldsymbol{\eta}'$  be an arbitrary  $(k+p)$ -dimensional row vector, and define

$$Z_{in} = \boldsymbol{\eta}' \mathbf{M}_n^{-1} \mathbf{X}'_i e_i,$$

$$\delta_{in}^2 = E\{Z_{in}^2 | \mathcal{A}_{t-1}\} = \boldsymbol{\eta}' \mathbf{M}_n^{-1} \mathbf{X}'_i \mathbf{X}_i \mathbf{M}_n^{-1'} \boldsymbol{\eta} \sigma^2,$$

$$s_{nn}^2 = E\{V_{nn}^2\} = E \left\{ \boldsymbol{\eta}' \mathbf{M}_n^{-1} \sum_{t=1}^n \mathbf{X}'_t \mathbf{X}_t \mathbf{M}_n^{-1'} \boldsymbol{\eta} \sigma^2 \right\},$$

where  $V_{nn}^2 = \sum_{t=1}^n \delta_{in}^2$ . By assumption,

$$\operatorname{plim} s_{nn}^{-2} V_{nn}^2 = 1$$

and condition (ii) of Theorem 5.3.4 is satisfied. We have

$$s_{nn}^{-2} \sum_{t=1}^n \int_{R_t} |\boldsymbol{\eta}' \mathbf{M}_n^{-1} \mathbf{X}'_t|^2 dF_t(e) \leq s_{nn}^{-2} \sum_{t=1}^n |\boldsymbol{\eta}' \mathbf{M}_n^{-1} \mathbf{X}'_t|^2 \int_{R_0} e^2 dF_t(e),$$

where  $R_t = \{e: |e| > |\boldsymbol{\eta}' \mathbf{M}_n^{-1} \mathbf{X}'_t|^{-1} \epsilon\}$  for  $t = 1, 2, \dots, F_t(e)$  is the distribution function of  $e_t$ , and

$$R_0 = \left\{ e: |e| > \left( \sup_{1 \leq t \leq n} |\boldsymbol{\eta}' \mathbf{M}_n^{-1} \mathbf{X}'_t| \right)^{-1} \epsilon \right\}.$$

Because  $s_{nn}^{-2} V_{nn}^2 \xrightarrow{P} 1$  and because the supremum of the integrals of  $e^2$  over  $R_0$  goes to zero as  $n$  increases, condition (iii) of Theorem 5.3.4 is satisfied. Because  $\boldsymbol{\eta}$  is arbitrary, the joint normality is established.  $\blacktriangle$

In Theorem 9.8.1, conditions on the nature of the difference equation associated with (9.8.1) are imposed through the assumptions (9.8.4) and (9.8.5). If  $\{\Phi_t\}$  is a fixed sequence, where  $\Phi_t = (\varphi_{t1}, \varphi_{t2}, \dots, \varphi_{tk})$ , then the existence of a sequence of fixed matrices  $\{\mathbf{M}_{11n}\}$  such that

$$\lim_{n \rightarrow \infty} \mathbf{M}_{11n}^{-1} \sum_{i=1}^n \Phi_i' \Phi_i \mathbf{M}_{11n}^{-1} = \mathbf{I},$$

$$\lim_{n \rightarrow \infty} \sup_{1 \leq i \leq n} |\mathbf{M}_{11n}^{-1} \Phi_i|^2 = 0,$$

together with the assumption that the roots of (9.8.2) are less than one in absolute value, is sufficient for (9.8.4) and (9.8.5). See Fuller, Hasza, and Goebel (1981).

The conclusion of Theorem 9.8.1 can be obtained with martingale difference errors or with iid( $0, \sigma^2$ ) errors. The martingale difference property is critical to the proof. If the  $e_t$  are autocorrelated,  $E\{Y_{t-1} e_t\}$  may not be zero, in which case the least squares estimator is no longer consistent.

To obtain consistent estimators for the parameters in the presence of autocorrelated errors, we use the method of instrumental variables introduced in Section 5.6. The lagged values of  $\varphi_{it}$  are the natural instrumental variables. To simplify the presentation, we consider a model with two independent variables and two lagged values of  $Y_t$ ,

$$Y_t = \beta_1 \varphi_{t1} + \beta_2 \varphi_{t2} + \lambda_1 Y_{t-1} + \lambda_2 Y_{t-2} + Z_t, \quad t = 1, 2, \dots, \quad (9.8.6)$$

where  $Z_t$  is a stationary time series with properties described in Theorem 9.8.2.

There are several methods of instrumental variable estimation. In econometrics, the two common instrumental variable procedures are called two-stage least squares and limited information maximum likelihood. The two procedures have the same asymptotic efficiency. For the model (9.8.6), the first step of the two-stage procedure is the regression of  $Y_{t-1}$  and  $Y_{t-2}$  on  $\varphi_{t1}, \varphi_{t-1,1}, \varphi_{t-2,1}, \varphi_{t2}, \varphi_{t-1,2}$ , and  $\varphi_{t-2,2}$ . The predicted values  $\hat{Y}_{t-1}$  and  $\hat{Y}_{t-2}$  are obtained from these regressions, and the instrumental variable estimates are obtained by regressing  $Y_t$  on  $(\varphi_{t1}, \varphi_{t2}, \hat{Y}_{t-1}, \hat{Y}_{t-2})$ .

To apply an instrumental variable procedure to the model (9.8.6), we assume the matrix of sums of squares and products of  $(\varphi_{t1}, \varphi_{t2}, \varphi_{t-1,1}, \varphi_{t-2,1}, \varphi_{t-1,2}, \varphi_{t-2,2})$  to be nonsingular and to satisfy assumption 1 preceding Theorem 5.6.1. We require only two instrumental variables in order to estimate the parameters  $\lambda_1$  and  $\lambda_2$ , and some of the variables may be omitted if singularities appear in practice. For example, if  $\Phi = (\varphi_{00}, \varphi_{01}, \varphi_{02})$ , where  $\varphi_{00} = 1$  and  $\varphi_{01} = t$ , then the matrix with rows  $(1, t, \varphi_{t2}, t-1, \varphi_{t-1,2}, t-2, \varphi_{t-2,2})$  will be singular. One can delete variables from the regression to create a nonsingular regression problem. In the example case one can omit  $\varphi_{t-1,1} = t-1$  and  $\varphi_{t-2,1} = t-2$ .

In Theorem 9.8.2, we show that the instrumental variable estimator is consistent for the model (9.8.6). The conclusions generalize immediately to any number of  $\varphi_{it}$  and any number of lagged  $Y$ 's.

**Theorem 9.8.2.** Let  $\{Y_t: t \in (1, 2, \dots)\}$  satisfy the model (9.8.6), where

$$Z_t = \sum_{j=0}^{\infty} b_j e_{t-j},$$

the  $b_j$  are absolutely summable, the roots of  $m^2 - \lambda_1 m - \lambda_2 = 0$  are less than one in absolute value,  $(Y_{-1}, Y_0)$  are bounded in probability, and the  $e_i$  are independent  $(0, \sigma^2)$  random variables with  $E\{|e_i|^{2+\delta}\} < L^{2+\delta}$  for some finite  $L$  and some  $\delta > 0$ .

Let  $\varphi_{i1}$  and  $\varphi_{i2}$  satisfy the assumptions (9.1.8) and (9.1.9). Let

$$\mathbf{Q} = \lim_{n \rightarrow \infty} \mathbf{Q}_n$$

be nonsingular, where

$$\mathbf{Q}_n = \mathbf{D}_{1n}^{-1}(\Phi; \psi)'(\Phi; \psi)\mathbf{D}_{1n}^{-1},$$

$(\Phi; \psi)$  is the matrix whose  $t$ th row is  $(\varphi_{t1}, \varphi_{t2}, \varphi_{t-1,1}, \varphi_{t-2,1}, \varphi_{t-1,2}, \varphi_{t-2,2})$  for  $t = 3, 4, \dots, n$ , and

$$\mathbf{D}_{1n}^2 = \text{diag} \left[ \sum_{t=3}^n \varphi_{t1}^2, \sum_{t=3}^n \varphi_{t2}^2, \sum_{t=2}^{n-1} \varphi_{t1}^2, \sum_{t=1}^{n-2} \varphi_{t1}^2, \sum_{t=2}^{n-1} \varphi_{t2}^2, \sum_{t=1}^{n-2} \varphi_{t2}^2 \right].$$

Let  $\hat{\theta} = (\hat{\beta}_1, \hat{\beta}_2, \hat{\lambda}_1, \hat{\lambda}_2)$  be the instrumental variable estimator, and let

$$\mathbf{D}_{2n} = \text{diag} \left\{ \left( \sum_{t=1}^n \varphi_{t1}^2 \right)^{1/2}, \left( \sum_{t=1}^n \varphi_{t2}^2 \right)^{1/2}, S_n, S_n \right\},$$

where  $S_n^2 = \sum_{t=1}^n (\beta_1 \varphi_{t1} + \beta_2 \varphi_{t2})^2$ . Then

$$\mathbf{D}_{2n}(\hat{\theta} - \theta) = O_p(1).$$

**Proof.** By the difference equation properties of our model, we may write

$$Y_t = \beta_1 \sum_{j=0}^{t-1} w_j \varphi_{t-j,1} + \beta_2 \sum_{j=0}^{t-1} w_j \varphi_{t-j,2} + U_t + c_{1t} Y_0 + c_{2t} Y_{-1}, \quad (9.8.7)$$

where  $U_t = \sum_{j=0}^{t-1} w_j Z_{t-j}$ , the  $w_j$  are defined in terms of  $\lambda_1$  and  $\lambda_2$  by Theorem 2.6.1, and  $c_{1t}$  and  $c_{2t}$  go to zero as  $t$  goes to infinity. Equation (9.8.7) gives  $Y_{t-1}$  and  $Y_{t-2}$  as functions of lagged values of  $\varphi_{ti}$  and  $\varphi_{tj}$  and of a random term  $U_t$ . By our assumptions on the  $\varphi_{ti}$  and since  $\beta_1$  and  $\beta_2$  are not both zero, the current and lagged values of  $\varphi_{ti}$  satisfy the conditions placed on  $\Phi$  and  $\psi$  in Theorem 5.6.1. Likewise,  $Z_t$  satisfies the conditions on  $Z_t$  of Theorem 5.6.1.

The  $\mathbf{B}_n$  of assumption 4 preceding Theorem 5.6.1 is

$$\mathbf{B}_n = \mathbf{D}_{1n}^{-1}(\Phi; \psi)' \Gamma(\Phi; \psi) \mathbf{D}_{1n}^{-1},$$

where the  $ij$ th element of  $\Gamma$  is  $\gamma_Z(|i-j|)$ , and the matrix  $\mathbf{B} = \lim_{n \rightarrow \infty} \mathbf{B}_n$  is of the same form as the matrix  $\mathbf{B}$  of Theorem 9.1.1. The matrices  $\mathbf{G}_{nij}$  of assumption 4 preceding Theorem 5.6.1 are of a similar form, with the covariance matrix of  $U_{t-1}$ , of  $U_{t-2}$ , or of  $U_{t-1}$  with  $U_{t-2}$  replacing that of  $Z_t$ .

Therefore, the conditions of Theorem 5.6.1 are satisfied, and the order in probability of the error in the estimators is given by that theorem. ▲

By Theorem 9.8.2 the instrumental variable estimator is consistent when the error in the equation is a stationary moving average time series with absolutely summable coefficients. Although the instrumental variable estimators are not the most efficient estimators, they are not dependent on a specification for the error process. Hence, they can be used as preliminary estimators. In some applications, one will use the instrumental variable estimators to obtain information about the error process. Given the instrumental variable estimates  $\hat{\beta}_1, \hat{\beta}_2, \hat{\lambda}_1, \hat{\lambda}_2$  of the model (9.8.6), we can calculate the estimated residuals

$$\hat{Z}_t = Y_t - \hat{\beta}_1 \varphi_{t1} - \hat{\beta}_2 \varphi_{t2} - \hat{\lambda}_1 Y_{t-1} - \hat{\lambda}_2 Y_{t-2}. \quad (9.8.8)$$

The presence of the lagged  $Y$ -values in the equation means that the results of Theorem 9.3.1 do not hold for autocorrelations computed from these residuals. However, we are able to obtain a somewhat weaker result.

**Theorem 9.8.3.** Given the assumptions of Theorem 9.8.2, the sample covariances calculated from the instrumental variable residuals satisfy

$$\hat{\gamma}_2(h) - \hat{\gamma}_z(h) = O_p(r_n),$$

where

$$\hat{\gamma}_2(h) = \frac{1}{n} \sum_{t=3}^{n-h} \hat{Z}_t \hat{Z}_{t+h}, \quad \hat{\gamma}_z(h) = \frac{1}{n} \sum_{t=3}^{n-h} Z_t Z_{t+h}, \quad h = 0, 1, 2, \dots, n-1,$$

$\hat{Z}_t$  is defined in (9.8.8),  $r_n = \max\{n^{-1/2}, S_n^{-1}\}$ , and  $S_n$  is defined in Theorem 9.8.2.

**Proof.** Using the definition of  $\hat{Z}_t$ , we have, for  $h = 0, 1, \dots, n-1$ ,

$$\begin{aligned} \hat{\gamma}_2(h) &= \frac{1}{n} \sum_{t=3}^{n-h} \hat{Z}_t \hat{Z}_{t+h} \\ &= \frac{1}{n} \left[ \sum_{t=3}^{n-h} Z_t Z_{t+h} - (\hat{\theta} - \theta)' \sum_{t=3}^{n-h} A_t' Z_{t+h} \right. \\ &\quad - \sum_{t=3}^{n-h} Z_t A_{t+h} (\hat{\theta} - \theta) \\ &\quad \left. + (\hat{\theta} - \theta)' \sum_{t=3}^{n-h} A_t' A_{t+h} (\hat{\theta} - \theta) \right], \end{aligned}$$

where  $A_t = (\varphi_{t1}, \varphi_{t2}, Y_{t-1}, Y_{t-2})$ . Now, for example,

$$\begin{aligned} \frac{1}{n} \sum_{t=3}^{n-h} Y_{t-1} Z_{t+h} &\leq n^{-1/2} \left( \sum_{t=3}^{n-h} Y_{t-1}^2 \right)^{1/2} \left( \frac{1}{n} \sum_{t=3}^{n-h} Z_{t+h}^2 \right)^{1/2} \\ &= O_p(\max\{1, n^{-1/2} S_n\}). \end{aligned}$$

Therefore,

$$\begin{aligned} \frac{1}{n} (\hat{\theta} - \theta)' \mathbf{D}_{2n} \mathbf{D}_{2n}^{-1} \sum_{t=3}^{n-h} \mathbf{A}'_t Z_{t+h} &= O_p(\max\{n^{-1/2}, S_n^{-1}\}), \\ \frac{1}{n} (\hat{\theta} - \theta)' \mathbf{D}_{2n} \mathbf{D}_{2n}^{-1} \sum_{t=3}^{n-h} \mathbf{A}'_t \mathbf{A}_{t+h} \mathbf{D}_{2n}^{-1} \mathbf{D}_{2n} (\hat{\theta} - \theta) &= O_p(\max\{n^{-1/2}, S_n^{-1}\}), \end{aligned}$$

where  $\mathbf{D}_{2n}$  is defined in Theorem 9.8.2. ▲

Because the estimated autocorrelations are consistent estimators, the estimated autocorrelations of the residuals can be used to develop a model for the error process. Given an original model such as (9.8.6) and a specified form for the error process, it is natural to estimate the entire model by a nonlinear fitting procedure.

To introduce such models, let  $Z_t$  be the first order autoregressive process

$$Z_t = \rho Z_{t-1} + e_t, \quad (9.8.9)$$

where  $|\rho| < 1$  and the  $e_t$  are independent  $(0, \sigma^2)$  random variables. We may then write (9.8.6) as a nonlinear model with lagged values of  $Y_t$  among the explanatory variables, and whose error term  $e_t$  is a sequence of independent random variables. That is, substituting (9.8.6) into (9.8.9), we obtain

$$\begin{aligned} Y_t &= \beta_1 \varphi_{t1} + \beta_2 \varphi_{t2} + (\lambda_1 + \rho) Y_{t-1} + (\lambda_2 - \rho \lambda_1) Y_{t-2} \\ &\quad - \rho \beta_1 \varphi_{t-1,1} - \rho \beta_2 \varphi_{t-1,2} - \rho \lambda_2 Y_{t-3} + e_t \\ &\stackrel{(say)}{=} f(\mathbf{W}_t; \boldsymbol{\theta}) + e_t, \quad t = 4, 5, \dots, n, \end{aligned} \quad (9.8.10)$$

where  $\mathbf{W}_t = (\varphi_{t1}, \varphi_{t2}, \varphi_{t-1,1}, \varphi_{t-1,2}, Y_{t-1}, Y_{t-2}, Y_{t-3})$  and  $\boldsymbol{\theta}' = (\theta_1, \theta_2, \theta_3, \theta_4, \theta_5) = (\beta_1, \beta_2, \lambda_1, \lambda_2, \rho)$ . Note that we have expanded the vector of parameters to include  $\rho$ .

By Theorem 5.5.3 and its corollaries, the least squares estimator of  $\boldsymbol{\theta}$  will be consistent and, when properly standardized, asymptotically normally distributed under mild assumptions. Note that the model (9.8.10) is of the same form as (9.8.1) with coefficients that satisfy nonlinear restrictions.

**Example 9.8.1.** To illustrate the estimation of a model containing lagged values of the dependent variable, we use a model studied by Ball and St. Cyr (1966). The data, displayed in Table 9.8.1, are for production and employment in the bricks, pottery, glass, and cement industry of the United Kingdom. The

**Table 9.8.1. Production Index and Employment of the Bricks, Pottery, Glass, and Cement Industry of the United Kingdom**

Year-Quarter	Logarithm of Production Index	Logarithm of Employment (000)
1955-1	4.615	5.844
2	4.682	5.846
3	4.644	5.852
4	4.727	5.864
1956-1	4.625	5.855
2	4.700	5.844
3	4.635	5.841
4	4.663	5.838
1957-1	4.625	5.823
2	4.644	5.817
3	4.575	5.823
4	4.635	5.817
1958-1	4.595	5.802
2	4.625	5.790
3	4.564	5.790
4	4.654	5.784
1959-1	4.595	5.778
2	4.700	5.781
3	4.654	5.802
4	4.727	5.814
1960-1	4.719	5.820
2	4.787	5.826
3	4.727	5.838
4	4.787	5.841
1961-1	4.771	5.838
2	4.812	5.844
3	4.796	5.846
4	4.828	5.855
1962-1	4.762	5.852
2	4.868	5.864
3	4.804	5.864
4	4.844	5.861
1963-1	4.754	5.844
2	4.875	5.835
3	4.860	5.841
4	4.963	5.852
1964-1	4.956	5.849
2	5.004	5.858

SOURCE: Ball and St. Cyr (1966). Reproduced from *Review of Economic Studies* Vol. 33, No. 3 with the permission of the authors and editor.

quarterly data cover the period 1955–1 to 1964–2. The model studied by Ball and St. Cyr can be written as

$$Y_t = \beta_0 + \beta_1 \varphi_t + \beta_2 t + \beta_3 D_{t1} + \beta_4 D_{t2} + \beta_5 D_{t3} + \lambda Y_{t-1} + Z_t, \quad (9.8.11)$$

where

$Y_t$  = the logarithm of employment in thousands of the bricks, pottery, glass, and cement industry of the United Kingdom,

$\varphi_t$  = the logarithm of the production index of the industry,

$t$  = time with 1954–4 as the origin.

$$D_{tj} = \begin{cases} 1 & \text{if observation } t \text{ occurs in quarter } j, \\ -1 & \text{if observation } t \text{ occurs in quarter 4,} \\ 0 & \text{otherwise.} \end{cases}$$

We assume  $Z_t$  is a zero mean stationary time series,  $\beta_1 \neq 0$ , and  $|\lambda| < 1$ .

To construct the instrumental variable estimator, we use the limited information maximum likelihood procedure SYSLIN of SAS/ETS\*. In the terminology of that program  $Y_t$  and  $Y_{t-1}$  are endogenous variables and  $\varphi_t$ ,  $\varphi_{t-1}$ ,  $t$ ,  $D_{t1}$ ,  $D_{t2}$ , and  $D_{t3}$  are instruments. The instrumental variable estimated equation is

$$\hat{Y}_t = 0.27 + 0.077\varphi_t - 0.00042t - 0.00492D_{t1} \\ - 0.00336D_{t2} + 0.00728D_{t3} + 0.894Y_{t-1}.$$

The program calculates estimated standard errors under the assumption that the errors are uncorrelated. The estimated standard errors are biased if the errors are autocorrelated. Because we are considering the possibility that the errors are autocorrelated, we do not present the estimated standard errors.

An analysis of the estimated residuals  $\hat{Z}_t$ , such as that in Table 9.7.2, indicated that the data are consistent with the hypothesis that  $Z_t$  is a first order autoregressive process with a parameter of about 0.42.

To estimate the model with autoregressive errors, we use the maximum likelihood option of procedure ARIMA of SAS/ETS\*. We specify the model (9.8.11) with a first order autoregressive error. While the error in the first observation is correlated with the lagged dependent variable for that observation, we chose to include that observation in the fit. Thus, the program is used to maximize a pseudolikelihood that treats  $Y_{t-1}$  as an ordinary explanatory variable. The estimated model is

$$\hat{Y}_t = 0.88 + 0.097\varphi_t - 0.00055t - 0.00368D_{t1} \\ (0.41) \quad (0.032) \quad (0.00030) \quad (0.00186) \\ - 0.00432D_{t2} + 0.00734D_{t3} + 0.773Y_{t-1}, \\ (0.00181) \quad (0.00162) \quad (0.081)$$

and  $\hat{\rho} = 0.386$  (0.194), where the numbers in parentheses are the estimated

standard errors. Although this is a fairly small sample with a fairly large number of explanatory variables, the maximum likelihood estimates are similar to the instrumental variable estimates. Because of the small sample and the presence of time among the explanatory variables, it is a reasonable conjecture that the estimator  $\hat{\rho}$  possesses a negative bias.

There are two important reasons for recognizing the potential for autocorrelated errors in an analysis such as this one. First, the simple least squares estimators are no longer consistent when the errors are autocorrelated. Second, the estimators of the variances of the estimators may be seriously biased. A hypothesis of some interest in the study of employment was the hypothesis that  $\beta_1 + \lambda = 1$ . This corresponds to the hypothesis of constant returns to scale for labor input. In the simple least squares analysis the sum of these coefficients is 0.89 with an estimated standard error of 0.045. The resulting "t-statistic" of -2.5 suggests that there are increasing returns to scale to labor in this industry. On the other hand, the analysis recognizing the possibility of autocorrelated errors estimates the sum as 0.88, but with a standard error of 0.072. The associated "t-statistic" of -1.7 and the potential for small sample bias in the estimators could lead one to accept the hypothesis that the sum is one. ▲▲

## REFERENCES

- Section 9.1.** Anderson (1971), Grenander (1954), Grenander and Rosenblatt (1957), Hannan (1970).
- Section 9.2.** Ahlberg, Nilson, and Walsh (1967), Fuller (1969), Gallant and Fuller (1973), Greville (1969), Poirier (1973).
- Section 9.3.** Bollerslev, Chou, and Kroner (1992), Durbin and Watson (1950, 1951), Engle (1982), Granger and Teräsvirta (1993), Hannan (1969), Harvey, Ruiz, and Shephard (1994), Melino and Turnbull (1990), von Neumann (1941, 1942).
- Section 9.4.** Davis (1941), Grether and Nerlove (1970), Jorgenson (1964), Kendall and Stuart (1966), Lovell (1963), Malinvaud (1970a), Nerlove (1964), Stephenson and Farr (1972), Tintner (1940, 1952).
- Section 9.5.** Harvey (1989).
- Section 9.6.** Davis (1941), Durbin (1962, 1963), Hannan (1958), Tintner (1940, 1952), Rao and Tintner (1963).
- Section 9.7.** Amemiya (1973), Cochran and Orcutt (1949), Durbin (1960), Hildreth (1969), Malinvaud (1970a), Prest (1949).
- Section 9.8.** Aigner (1971), Amemiya and Fuller (1967), Ball and St. Cyr (1966), Dhrymes (1971), Fuller and Martin (1961), Griliches (1967), Hatanaka (1974), Malinvaud (1970a), Nagaraj and Fuller (1991, 1992), Wallis (1967).

## EXERCISES

1. Let the time series  $X_t$  be defined by

$$X_t = \beta_0 + \beta_1 t + Z_t,$$

where

$$Z_t = \rho Z_{t-1} + e_t, \quad |\rho| < 1, \quad t = 0, \pm 1, \pm 2, \dots,$$

and  $\{e_t\}$  is a sequence of independent  $(0, \sigma^2)$  random variables. Compute the covariance matrices of  $\hat{\beta}_s$  of (9.1.2) and  $\hat{\beta}_G$  of (9.1.3) for  $\rho = 0.9$  and  $n = 5$  and 10.

2. The data in the accompanying table are the forecasts of the yield of corn in Illinois released by the United States Department of Agriculture in August of the production year.
- (a) For this set of data estimate a mean function meeting the following restrictions:
- (i) The mean function is linear from 1944 through 1956, quadratic from 1956 through 1960, linear from 1960 through 1965, quadratic from 1965 through 1969, and linear from 1969 through 1974.
  - (ii) The mean function, considered as a continuous function of time, has a continuous first derivative.
- (b) Reestimate the mean function subject to the restriction that the derivative is zero for the last five years. Test the hypothesis that the slope for the last five years is zero.
- (c) Plot the data and the estimated mean function.

August Forecast of Corn Yield for Illinois

Year	Yield	Year	Yield	Year	Yield
1944	45.5	1955	58.0	1965	88.0
1945	43.0	1956	59.0	1966	82.0
1946	55.0	1957	52.0	1967	96.0
1947	45.0	1958	64.0	1968	103.0
1948	58.0	1959	63.0	1969	95.0
1949	61.0	1960	63.0	1970	93.0
1950	58.0	1961	73.0	1971	92.0
1951	56.0	1962	79.0	1972	97.0
1952	56.0	1963	81.0	1973	105.0
1953	58.0	1964	85.0	1974	86.0
1954	45.0				

SOURCE: U.S. Department of Agriculture, *Crop Production*, various issues.

3. Evaluate (9.3.5) and (9.3.6) for  $n = 10$  and the model of Exercise 9.1. The covariance between  $\sum_{t=2}^n Z_t Z_{t-1}$  and  $\sum_{t=1}^n Z_t^2$  for the first order autoregressive process can be evaluated using Theorem 6.2.1. Using these approximations and the approximation

$$E\left\{\frac{A}{B}\right\} = \frac{E\{A\}}{E\{B\}} + [E\{B\}]^{-2} \left[ \text{Cov}\{A, B\} - \left(\frac{E\{A\}}{E\{B\}}\right) \text{Var}\{B\} \right],$$

find the approximate expected value of  $\hat{r}_2(1)$  for the model and sample sizes of Exercise 9.1.

4. Using a period of three observations, construct a moving average to remove a linear trend from the third observation and hence show that the second difference operator is a multiple of the moving average constructed to remove a linear trend from the third observation in a group of three observations.

5. Define  $X_t$  by

$$X_t = \rho X_{t-1} + e_t, \quad |\rho| < 1,$$

where the  $e_t$  are independent  $(0, \sigma^2)$  random variables. Find the auto-correlation functions and spectral densities of

$$Y_t = X_t - X_{t-1} \quad \text{and} \quad W_t = X_t - 2X_{t-1} + X_{t-2}.$$

Plot the spectral densities for  $\rho = 0.5$ .

6. Consider the model

$$Y_t = \beta_0 + \beta_1 t + \beta_2 t^2 + \sum_{i=1}^4 \alpha_i \delta_{ti} + Z_t,$$

where

$$\delta_{ti} = \begin{cases} 1 & \text{for quarter } i \ (i = 1, 2, 3, 4), \\ 0 & \text{otherwise,} \end{cases}$$

$$\sum_{i=1}^4 \alpha_i = 0,$$

and  $Z_t$  is a stationary time series. Assuming that the model holds for a period of 20 observations (5 years), construct the weights needed to decompose the most recent observation in a set of 20 observations into "trend," "seasonal," and "remainder." Define trend to be  $\beta_0 + \beta_1 t + \beta_2 t^2$  for the  $t$ th observation, and seasonal for quarter  $i$  to be the  $\alpha_i$  associated with that quarter.

7. Obtain the conclusion of Proposition 9.6.3 directly by differencing the autoregressive moving average time series.
8. Prove Lemma 9.4.1.
9. Prove Lemma 9.4.2.
10. Using the data of Exercise 2 and the model of part b of that exercise, construct the statistics (9.3.8) and (9.3.9).

11. Assume that the data of Exercise 2 of Chapter 7 satisfy the model

$$Y_t = \sum_{j=1}^4 \alpha_j \delta_{tj} + \beta t + Z_t,$$

where

$$\delta_{tj} = \begin{cases} 1 & \text{for quarter } j, \\ 0 & \text{otherwise,} \end{cases}$$

$Z_t$  is the stationary autoregressive process

$$Z_t = \theta_1 Z_{t-1} + \theta_2 Z_{t-2} + e_t,$$

and  $\{e_t\}$  is a sequence of normal independent  $(0, \sigma^2)$  random variables.

- (a) Estimate the parameters of the model. Estimate the standard errors of your estimators. Do the data support the model for the error time series?
- (b) Using the model of part a, estimate gasoline consumption for the four quarters of 1974. Construct 95% confidence intervals for your predictions, using normal theory. Actual consumption during 1974 was [2223, 2540, 2596, 2529]. What happened? *Computational hint:* If a computer program is available that fits regression models with autoregressive errors, the parameters and predictions can be obtained in one fitting operation. Add four "observations" and four independent variables to the data set. The four added independent variables are defined by

$$\varphi_{i,5+i} = \begin{cases} -1 & \text{for } 1974-i, \\ 0 & \text{otherwise} \end{cases}$$

for  $i = 1, 2, 3, 4$ . The four observations on the original data for 1973 and the four added "observations" for 1974 are displayed in the accompanying table. The values of the original independent variables for the prediction period are the true values of these independent variables, while the values of  $Y_t$  for the prediction period are zero. The augmented regression model with autoregressive errors is then estimated. The coefficients for the four added independent variables are the predictions, and the estimated standard errors are appropriate for the predictions.

Year	Quarter	Dependent Variable	Independent Variables for Predictions							
			$\delta_{t1}$	$\delta_{t2}$	$\delta_{t3}$	$\delta_{t4}$	$t$	0	0	0
1973	1	2454	1	0	0	0	53	0	0	0
	2	2647	0	1	0	0	54	0	0	0
	3	2689	0	0	1	0	55	0	0	0
	4	2549	0	0	0	1	56	0	0	0
1974	1	0	1	0	0	0	57	-1	0	0
	2	0	0	1	0	0	58	0	-1	0
	3	0	0	0	1	0	59	0	0	-1
	4	0	0	0	0	1	60	0	0	-1

12. Using the Prest data of Table 9.7.1, estimate the parameters of the model

$$Y_t = \beta_0 + \beta_1 t + \beta_2(t - 35)^2 + \beta_3 \varphi_{t1} + \beta_4 \varphi_{t2} + \lambda Y_{t-1} + Z_t,$$

$$Z_t = \rho Z_{t-1} + e_t,$$

where  $\{e_t\}$  is a sequence of independent  $(0, \sigma^2)$  random variables with  $E\{e_t^4\} = \eta\sigma^4$  and  $|\rho| < 1$ ,  $|\lambda| < 1$ .

13. Let  $Y_t$  satisfy

$$Y_t = \alpha \varphi_t + Z_t,$$

$$Z_t = e_t + \beta e_{t-1}, \quad t = 1, 2, \dots,$$

where  $|\beta| < 1$ ,  $\{e_t\}$  is a sequence of independent  $(0, \sigma^2)$  random variables with  $E\{e_t^4\} = \eta\sigma^4$ , and  $\{\varphi_t\}$  satisfies the assumptions (9.1.8) and (9.1.9). Let

$$\hat{\alpha}_S = \left( \sum_{i=1}^n \varphi_i^2 \right)^{-1} \sum_{i=1}^n \varphi_i Y_i,$$

$$\hat{Z}_t = Y_t - \hat{\alpha}_S \varphi_t,$$

$$e_t(\hat{Z}; \tilde{\beta}) = \begin{cases} 0, & t = 1, \\ \hat{Z}_t - \tilde{\beta} e_{t-1}(\hat{Z}; \tilde{\beta}), & t = 2, 3, \dots, n, \end{cases}$$

$$W_t(Y; \tilde{\beta}) = \begin{cases} 0, & t = 1, \\ e_{t-1}(\hat{Z}; \tilde{\beta}) - \tilde{\beta} W_{t-1}(Y; \tilde{\beta}), & t = 2, 3, \dots, n, \end{cases}$$

and  $\tilde{\beta}$  is constructed from the first order autocorrelation of the  $\hat{Z}_t$  by the rule (8.3.1) and restricted to  $(-1, 1)$ .

(a) Show that

$$\hat{\beta} - \tilde{\beta} = O_p(n^{-1}),$$

where

$$\hat{\beta} = \tilde{\beta} + \Delta \hat{\beta},$$

$$\Delta \hat{\beta} = \frac{\sum_{t=1}^n e_t(\hat{Z}; \tilde{\beta}) W_t(\hat{Z}; \tilde{\beta})}{\sum_{t=1}^n [W_t(\hat{Z}; \tilde{\beta})]^2},$$

and  $\hat{\beta}$  is the analogous estimator constructed by replacing  $\hat{Z}$  with  $Z$ .

(b) Let  $\hat{\epsilon} = \hat{T}z$ , where

$$\hat{\epsilon}_t = \hat{\mathbf{T}}_{t,t} \mathbf{z} = \begin{cases} (1 + \hat{\beta}^2)^{-1} Z_t, & t = 1, \\ Z_t - \hat{\beta} \hat{\epsilon}_{t-1}, & t = 2, 3, \dots, n, \end{cases}$$

and  $\hat{\mathbf{T}}_{t,t}$  is the  $t$ th row of  $\hat{\mathbf{T}}$ . Show that

$$\left[ \sum_{i=1}^n \varphi_i^2 \right]^{1/2} (\hat{\alpha}_G - \tilde{\alpha}) = O_p(n^{-1/2}),$$

where

$$\tilde{\alpha} = [\Phi' \hat{\mathbf{T}}' \hat{\mathbf{T}} \Phi]^{-1} \Phi' \hat{\mathbf{T}}' \hat{\mathbf{T}} \mathbf{y},$$

$$\hat{\alpha}_G = [\Phi' \mathbf{V}^{-1} \Phi]^{-1} \Phi' \mathbf{V}^{-1} \mathbf{y},$$

and

$$\mathbf{V} = E\{\mathbf{z}\mathbf{z}'\}.$$

14. Let  $\{Y_t\}$  satisfy

$$Y_t = \sum_{i=0}^r \beta_i \varphi_{ti} + \sum_{j=1}^s \lambda_j Y_{t-j} + e_t, \quad t = 1, 2, \dots,$$

where  $\{e_t\}$  is a sequence of independent  $(0, \sigma^2)$  random variables with  $E\{e_t^4\} = \eta \sigma^4$  and the roots of  $m^s - \sum_{j=1}^s \lambda_j m^{s-j} = 0$  are less than one in absolute value. State and prove Theorem 9.8.1 for this model.

15. Add a linear term to the wheat model of Section 9.2, and fit the model with a first order autoregressive error.
16. Let  $\varphi_t = (1, \varphi_{t1}, \varphi_{t2}) = (1, t, t + (-1)')$ . Does this vector satisfy the assumptions (9.1.8) and (9.1.9) with  $\mathbf{A}_0$  nonsingular? Is there a linear transformation of  $\varphi_t$  that satisfies the two assumptions with  $\mathbf{A}_0$  nonsingular?
17. Let the model (9.7.1) hold with  $\varphi_t$  that satisfy the assumptions (9.1.8) and (9.1.9). Show that the conditions of Theorem 5.7.4 are satisfied for  $\{Z_t\}$  that is a stationary autoregressive moving average with iid  $(0, \sigma^2)$  errors.
18. Using the estimates of Example 9.5.1, construct the vector  $\mathbf{H}_m$  to estimate  $\mu_m$  using  $(Y_{m-3}, Y_{m-2}, \dots, Y_{m+2}, Y_{m+3})$ . What is the estimated variance of  $\hat{\mu}_m$  constructed with the vector?
19. The U.S. wheat yields for 1992 and 1993 are 39.4 and 38.3. Are these yields consistent with the model of Example 9.7.2? Fit the model of Example 9.7.2 to the data for 1908 through 1993, and predict yields for 1994, 1995, and 1996. Fit the model in which (9.7.8) is replaced with

$$Y_t = \beta_0 + \beta_1 \varphi_{t2} + \beta_2 \varphi_{t3} + Z_t,$$

where

$$\varphi_{t3} = \begin{cases} 0, & 1 \leq t \leq 70, \\ (t - 70)^2, & 70 \leq t \leq 80, \\ 100 + 20(t - 80), & t \geq 80. \end{cases}$$

Use the fitted model to predict yields for 1994, 1995, and 1996.

20. The U.S. wheat yields for 1992 and 1993 are given in Exercise 19 as 39.4 and 38.3, respectively. Fit the structural model of Example 9.5.1 to the data for 1908 through 1993. Predict yields for 1994, 1995, and 1996.
21. Let  $Z_t$  be a stationary time series with absolutely summable covariance function. Show that the assumptions (9.1.8) and (9.1.9) are sufficient for the limit (9.1.7) to exist. Give an example of a  $Z_t$  and  $\varphi_t$  such that  $B$  of (9.1.7) is singular.
22. Let  $Z_t$  satisfy the ARCH model of (9.3.21). Show that  $Z_t$  satisfies the assumptions of Proposition 9.3.3.
23. Consider the model

$$Z_t = \theta Z_{t-1} + a_t, \quad |\theta| < 1, \\ a_t = (k_1 + k_2 a_{t-1}^2)^{1/2} e_t,$$

where the  $e_t$  are  $N(0, 1)$  random variables and  $0 \leq k_2 \leq 0.3$ . Let  $\hat{\theta}_{OLS}$  and  $(\hat{k}_{1,OLS}, \hat{k}_{2,OLS})$  denote the least squares estimators obtained by regressing  $Z_t$  on  $Z_{t-1}$  and  $\hat{a}^2$  on  $(1, \hat{a}_{t-1}^2)$ , respectively, where  $\hat{a}_t = Z_t - \hat{\theta}_{OLS} Z_{t-1}$ . Define  $\hat{h}_t = (\hat{k}_{1,OLS} + \hat{k}_{2,OLS} \hat{a}_{t-1}^2)$ . Let  $(\hat{k}_{1,GLS}, \hat{k}_{2,GLS})$  be the estimator of  $(k_1, k_2)$  obtained by regressing  $\hat{h}_t^{-1} \hat{a}^2$  on  $(\hat{h}_t^{-1}, \hat{h}_t^{-1} \hat{a}_{t-1}^2)$ . Define  $\tilde{h}_t = (\hat{k}_{1,GLS} + \hat{k}_{2,GLS} \hat{a}_{t-1}^2)$ .

Let  $\hat{\theta}_{GLS}$  denote the estimated generalized least squares estimator obtained by regressing  $\tilde{h}_t^{-1/2} Z_t$  on  $\tilde{h}_t^{-1/2} Z_{t-1}$ .

- Find the asymptotic distribution of

$$n^{1/2}(\hat{\theta}_{OLS} - \theta, \hat{k}_{1,OLS} - k_1, \hat{k}_{2,OLS} - k_2)'.$$

- Find the asymptotic distribution of

$$n^{1/2}(\hat{\theta}_{GLS} - \theta, \hat{k}_{1,GLS} - k_1, \hat{k}_{2,GLS} - k_2)'.$$

- Consider the conditional maximum likelihood estimators  $\hat{\theta}_{ML}$ ,  $\hat{k}_{1,ML}$ , and  $\hat{k}_{2,ML}$  obtained by minimizing

$$L_n(\theta, k_1, k_2) = \sum_{t=2}^n \log h_t + \sum_{t=2}^n h_t^{-1} (Z_t - \theta Z_{t-1})^2,$$

where  $h_t = k_1 + k_2(Z_{t-1} - \theta Z_{t-2})^2$ . Find the asymptotic distribution of  $n^{1/2}(\hat{\theta}_{\text{ML}} - \theta, \hat{k}_{1,\text{ML}} - k_1, \hat{k}_{2,\text{ML}} - k_2)'$ . Compare the limiting distributions of  $\hat{\theta}_{\text{OLS}}$ ,  $\hat{\theta}_{\text{GLS}}$ , and  $\hat{\theta}_{\text{ML}}$ .

**CHAPTER 10**

# Unit Root and Explosive Time Series

In Chapter 9, we studied time series in which the mean was a function of time. A second important class of nonstationary time series are those that satisfy nonstationary stochastic difference equations. We begin our discussion with least squares estimators for the univariate unit root autoregressive process. Alternative estimation procedures for unit processes are discussed in Section 10.1.3. Tests based on the methods of Section 10.1.3 have been demonstrated to be more powerful against the unknown mean stationary alternative than tests based on ordinary least squares regression. The remaining sections are devoted to autoregressive processes with a root greater than one, multivariate unit root autoregressive processes, and time series with a moving average unit root.

**10.1. UNIT ROOT AUTOREGRESSIVE TIME SERIES****10.1.1. The Autoregressive Process with a Unit Root**

Consider the first order time series

$$\begin{aligned} Y_t &= \rho Y_{t-1} + e_t, \quad t = 1, 2, \dots, \\ Y_0 &= 0, \end{aligned} \tag{10.1.1}$$

where the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables. In Section 2.1, we showed that the solution to the difference equation is

$$Y_t = \sum_{j=0}^{t-1} \rho^j e_{t-j}. \tag{10.1.2}$$

One case of particular interest is obtained when  $\rho = 1$ . Then  $Y_t$  is the sum of  $t$  independent  $(0, \sigma^2)$  random variables. The time series (10.1.1) with  $\rho = 1$  is sometimes called a *random walk*.

By the results of Section 8.1, the maximum likelihood estimator of  $\rho$  conditional on  $Y_1$ , for any real valued  $\rho$ , is

$$\hat{\rho} = \left( \sum_{t=2}^n Y_{t-1}^2 \right)^{-1} \sum_{t=2}^n Y_t Y_{t-1}, \quad (10.1.3)$$

and

$$\hat{\rho} - \rho = \left( \sum_{t=2}^n Y_{t-1}^2 \right)^{-1} \sum_{t=2}^n Y_{t-1} e_t. \quad (10.1.4)$$

If  $\rho = 1$ , then  $Y_t = \sum_{j=1}^t e_j$  and the mean and variance of the numerator and the denominator of (10.1.4) can be computed. We have

$$\begin{aligned} E\left\{ \sum_{t=2}^n Y_{t-1}^2 \right\} &= E\left\{ \sum_{t=1}^{n-1} \left( \sum_{i=1}^t e_i \right)^2 \right\} \\ &= \frac{1}{2} n(n-1)\sigma^2, \end{aligned} \quad (10.1.5)$$

$$\text{Var}\left\{ \sum_{t=2}^n Y_{t-1}^2 \right\} = \frac{1}{3} n(n-1)(n^2-n+1)\sigma^4, \quad (10.1.6)$$

$$\text{Var}\left\{ \sum_{t=2}^n Y_{t-1} e_t \right\} = \sum_{t=1}^{n-1} t\sigma^4 = \frac{1}{2} n(n-1)\sigma^4, \quad (10.1.7)$$

$$\text{Cov}\left\{ \sum_{t=2}^n Y_{t-1} e_t, \sum_{t=2}^n Y_{t-1}^2 \right\} = \frac{1}{3} n(n-1)(n-2)\sigma^4. \quad (10.1.8)$$

Normality was used in obtaining (10.1.6) and (10.1.8). The other results remain true for  $e_t$  that are iid( $0, \sigma^2$ ).

These results differ in a notable way from those obtained for  $|\rho| < 1$ . When  $|\rho| < 1$ , every entry in the covariance matrix of  $(\Sigma_{t=2}^n Y_{t-1} e_t, \Sigma_{t=2}^n Y_{t-1}^2)$  is of order  $n$ , and the limit of the expected value of  $n^{-1} \Sigma_{t=2}^n Y_{t-1}^2$  is  $\sigma^2(1-\rho^2)^{-1}$ . If  $\rho = 1$ , the variance of the numerator of (10.1.4) is order  $n^2$  instead of order  $n$ . Also, the mean of the denominator of (10.1.4) is of order  $n^2$ , and the variance is order  $n^4$ . The denominator divided by  $0.5n(n-1)\sigma^2$  gives a random variable with mean one and variance  $4(n^2-n+1)[3(n^2-n)]^{-1}$ . Because the variance converges to  $4/3$ , the normalized denominator does not converge to a constant in mean square.

The moment results and the fact that  $\Sigma Y_t^2$  is a quadratic form in  $(e_1, e_2, \dots, e_n)$  can be used to show that the error in  $\hat{\rho}$ , as an estimator of one, is  $O_p(n^{-1})$ . Thus, for  $\rho = 1$ , the least squares estimator converges in probability to the true value more rapidly than does the estimator for  $|\rho| < 1$ .

**Lemma 10.1.1.** Let  $\hat{\rho}$  be defined by (10.1.3), and assume  $Y_t$  satisfies (10.1.2) with  $\rho = 1$ , where the  $e_t$  are normal independent  $(0, \sigma^2)$  random variables. Then

$$\hat{\rho} - 1 = O_p(n^{-1}).$$

**Proof.** The denominator of  $\hat{\rho}$  can be written as

$$\sum_{i=2}^n Y_{i-1}^2 = \mathbf{e}' \mathbf{A}_n \mathbf{e},$$

where  $\mathbf{e}' = (e_1, e_2, \dots, e_{n-1})$  and

$$\mathbf{A}_n = \begin{pmatrix} n-1 & n-2 & n-3 & \cdots & 1 \\ n-2 & n-2 & n-3 & \cdots & 1 \\ n-3 & n-3 & n-3 & \cdots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 1 & \cdots & 1 \end{pmatrix}. \quad (10.1.9)$$

The matrix  $\mathbf{A}_n$  can be diagonalized by the  $(n-1) \times (n-1)$  orthonormal matrix  $\mathbf{M}_n$  whose  $ij$ th element is

$$m_{ni} = 2(2n-1)^{-1/2} \cos[(4n-2)^{-1}(2j-1)(2i-1)\pi]. \quad (10.1.10)$$

Then

$$\mathbf{M}_n \mathbf{A}_n \mathbf{M}_n' = \Lambda_n = \text{diag}(\lambda_{1,n}, \lambda_{2,n}, \dots, \lambda_{n-1,n}),$$

where

$$\lambda_{in} = \frac{1}{4} \sec^2[(n-i)\pi(2n-1)^{-1}] \quad (10.1.11)$$

for  $i = 1, 2, \dots, n-1$ . See Rutherford (1946). Hence, the denominator of  $\hat{\rho} - 1$  can be written as

$$\mathbf{e}' \mathbf{A}_n \mathbf{e} = \sum_{i=1}^{n-1} \lambda_{in} U_{in}^2, \quad (10.1.12)$$

where

$$U_{in} = \sum_{i=1}^{n-1} m_{ni} e_i, \quad i = 1, 2, \dots, n-1,$$

and the  $U_{in}$  are  $\text{NI}(0, \sigma^2)$  random variables. By the moment results (10.1.5) and (10.1.6)

$$\sum_{i=1}^{n-1} \lambda_{in} = \frac{1}{2} n(n-1)$$

$$\sum_{i=1}^{n-1} \lambda_{in}^2 = \frac{1}{6} n(n-1)(n^2 - n + 1).$$

If we fix  $i$  and take the limit as  $n \rightarrow \infty$ , we have

$$\lim_{n \rightarrow \infty} n^{-2} \lambda_{in} = \gamma_i^2 = 4[(2i-1)\pi]^{-2}$$

for  $i = 1, 2, \dots$ . Therefore, the largest root,  $\lambda_{in}$ , increases at the rate  $n^2$  as  $n$  increases. Given  $\epsilon > 0$ ,

$$\begin{aligned} P\left\{\sum_{i=1}^{n-1} \lambda_{in} U_{in}^2 < \frac{1}{2}n(n-1)\epsilon\right\} &\leq P\{\lambda_{in} U_{in}^2 < 0.5n(n-1)\epsilon\} \\ &\leq P\{U_{in}^2 < 2\epsilon\}. \end{aligned}$$

Because  $U_{in}^2$  is a chi-square random variable, given  $\Delta > 0$ , there exists an  $\epsilon > 0$  such that  $P\{U_{in}^2 < 2\epsilon\} < \Delta$ . It follows that, given  $\Delta > 0$ , there exists a  $K\epsilon^{-1}$  such that

$$P\{(\mathbf{e}' \mathbf{A}_n \mathbf{e})^{-1} n(n-1) > K\epsilon^{-1}\} < \Delta$$

for all  $n$  and, by (10.1.6),

$$n(n-1) \left[ \sum_{t=2}^n Y_{t-1}^2 \right]^{-1} = O_p(1).$$

Because  $n^{-2} \sum_{t=2}^n Y_{t-1} e_t = O_p(n^{-1})$ , it follows that  $\hat{\rho} - 1 = O_p(n^{-1})$ . ▲

The numerator of  $\hat{\rho} - 1$  can be written in an alternative informative manner as

$$\sum_{t=2}^n Y_{t-1} e_t = \sum_{t=2}^n \sum_{j=1}^{t-1} e_j e_t = \frac{1}{2} \left( Y_n^2 - \sum_{t=1}^n e_t^2 \right). \quad (10.1.13)$$

Hence, as  $n \rightarrow \infty$ ,

$$2(n\sigma^2)^{-1} \sum_{t=2}^n Y_{t-1} e_t + 1 \xrightarrow{\mathcal{L}} \chi_1^2,$$

where  $\chi_1^2$  is the chi-square distribution with one degree of freedom. The probability that a one-degree-of-freedom chi-square random variable is less than one is 0.6826. Therefore, because the denominator is always positive, the probability that  $\hat{\rho} < 1$ , given  $\rho = 1$ , approaches 0.6826 as  $n$  gets large. Although a chi-square distribution is skewed to the right, the high correlation between the numerator and denominator of  $\hat{\rho} - 1$  strongly dampens the skewness. In fact, the distribution of  $\hat{\rho}$  displays skewness to the left.

To establish the limiting distribution of the least squares estimator of  $\rho$ , given that  $\rho = 1$ , we will use the following lemma.

**Lemma 10.1.2.** Let  $\{U_{in}; 1 \leq t \leq n, n \geq 1\}$  denote a triangular array of random variables defined on the probability space  $(\Omega, \mathcal{A}, P)$ . Assume

$$E[(U_{in}, U_{in}^2, U_{in}U_{sn})] = (0, \sigma^2, 0)$$

for  $1 \leq t \neq s \leq n$ . Let  $\{w_i; i = 1, 2, \dots\}$  be a sequence of real numbers, and let  $\{w_{in}; i = 1, 2, \dots, n; n = 1, 2, \dots\}$  be a triangular array of real numbers. If

$$\lim_{n \rightarrow \infty} \sum_{i=1}^n w_{in}^2 = \sum_{i=1}^{\infty} w_i^2 < \infty$$

and

$$\lim_{n \rightarrow \infty} w_{in} = w_i, \quad i = 1, 2, \dots,$$

then

$$\sum_{i=1}^n w_{in} U_{in} = \sum_{i=1}^n w_i U_{in} + o_p(1).$$

**Proof.** Let  $\epsilon > 0$  be given. Then we can choose an  $M$  such that

$$\sigma^2 \sum_{i=M+1}^{\infty} w_i^2 < \frac{\epsilon}{9}$$

and

$$\sigma^2 \left| \sum_{i=1}^n w_{in}^2 - \sum_{i=1}^{\infty} w_i^2 \right| < \frac{\epsilon}{9}$$

for all  $n > M$ . Furthermore, given  $M$ , we can choose  $N_0 > M$  such that  $n > N_0$  implies

$$\sigma^2 \sum_{i=1}^M (w_{in} - w_i)^2 < \frac{\epsilon}{9}$$

and

$$\sigma^2 \sum_{i=M+1}^n w_{in}^2 < \frac{3\epsilon}{9}.$$

Hence, for all  $n > N_0$ ,

$$V \left\{ \sum_{i=1}^n w_{in} U_{in} - \sum_{i=1}^n w_i U_{in} \right\} < \epsilon. \quad \blacktriangle$$

We now give the primary result for the distribution of the least squares estimator of the autoregressive parameter when the true process is a random walk.

**Theorem 10.1.1.** Let

$$Y_t = Y_{t-1} + e_t, \quad t = 1, 2, \dots,$$

where  $Y_0 = 0$ . Assume  $\{e_t\}_{t=1}^n$  satisfies

$$\begin{aligned} E\{(e_t, e_t^2) | \mathcal{A}_{t-1}\} &= (0, \sigma^2) \quad \text{a.s.}, \\ E\{|e_t|^{2+\delta} | \mathcal{A}_{t-1}\} &< M < \infty \quad \text{a.s.} \end{aligned}$$

for some  $\delta > 0$ , where  $\mathcal{A}_t$  is the sigma-field generated by  $\{e_1, e_2, \dots, e_t\}$ . Then

$$n(\hat{\rho} - 1) \xrightarrow{\mathcal{D}} (2G)^{-1}(T^2 - 1),$$

where  $\hat{\rho}$  is defined in (10.1.3),

$$\begin{aligned} (G, T) &= \left( \sum_{i=1}^{\infty} \gamma_i^2 U_i^2, \sum_{i=1}^{\infty} 2^{1/2} \gamma_i U_i \right), \\ \gamma_i &= (-1)^{i+1} 2[(2i-1)\pi]^{-1}, \end{aligned}$$

$\{U_i\}_{i=1}^{\infty}$  is a sequence of NI(0, 1) random variables, and  $(G, T)$  is defined as a limit in mean square.

**Proof.** The estimator  $\hat{\rho}$  is scale invariant. Hence, there is no loss of generality in assuming  $\sigma^2 = 1$ . Let

$$(G_n, T_n) = \left( n^{-2} \sum_{i=2}^n Y_{i-1}^2, n^{-1/2} \sum_{i=1}^{n-1} e_i \right).$$

From the definition of  $\hat{\rho}$  and from (10.1.13), we have

$$\begin{aligned} n(\hat{\rho} - 1) &= \left( n^{-2} \sum_{i=2}^n Y_{i-1}^2 \right)^{-1} n^{-1} \sum_{i=2}^n Y_{i-1} e_i \\ &= (2G_n)^{-1}(T_n^2 - 1) + o_p(1), \end{aligned}$$

where  $n^{-1} \sum_{i=1}^n e_i^2 = \sigma^2 + o_p(1)$  by Corollary 5.3.5.

To establish the limit distribution result, we show that  $(G_n, T_n)$  converges in distribution to  $(G, T)$ . Let

$$\mathbf{U}_n = (U_{1n}, U_{2n}, \dots, U_{n-1,n})' = \mathbf{M}_n \mathbf{e}_n,$$

where  $\mathbf{e}_n = (e_1, e_2, \dots, e_{n-1})'$  and  $\mathbf{M}_n$  is the matrix with  $(i, j)$ th element given by (10.1.10). Then

$$T_n = n^{-1/2} \mathbf{J}'_n \mathbf{e}_n = n^{-1/2} \mathbf{J}'_n \mathbf{M}_n^{-1} \mathbf{U}_n,$$

where  $\mathbf{J}_n = (1, 1, \dots, 1)'$ . Let  $k_{in}$  be the  $i$ th element of  $n^{-1/2} \mathbf{J}'_n \mathbf{M}_n^{-1}$ . The element  $k_{in}$  is the covariance between  $T_n$  and  $U_{in}$ . Using this fact and the results of Jolley (1961, p. 78), it can be shown that, for fixed  $i$ ,

$$\lim_{n \rightarrow \infty} k_{in} = 2^{1/2} \gamma_i.$$

See Dickey (1976, p. 29). Also,  $2 \sum_{i=1}^{\infty} \gamma_i^2 = 1$ ; see Jolley (1961, p. 56). Let

$$(G_n, T_n) = \left( \sum_{i=1}^{n-1} n^{-2} \lambda_{in} U_{in}^2, \sum_{i=1}^{n-1} k_{in} U_{in} \right),$$

$$(\hat{G}_n, \hat{T}_n) = \left( \sum_{i=1}^{n-1} \gamma_i^2 U_{in}^2, \sum_{i=1}^{n-1} 2^{1/2} \gamma_i U_{in} \right),$$

where  $\lambda_{in}$  is defined in (10.1.11). Then  $T_n - \hat{T}_n$  converges in probability to zero by Lemma 10.1.2. Also,

$$G_n - \hat{G}_n = \sum_{i=1}^{n-1} (n^{-2} \lambda_{in} - \gamma_i^2) U_{in}^2 = o_p(1)$$

because

$$\lim_{n \rightarrow \infty} \sum_{i=1}^{n-1} |n^{-2} \lambda_{in} - \gamma_i^2| = 0.$$

The entries in  $n^{1/2} \mathbf{M}_n$  are bounded, and the sum of squares of each row of  $\mathbf{M}_n$  is one. Therefore, by Corollary 5.3.4, for fixed  $l$ ,

$$(U_{1n}, U_{2n}, \dots, U_{ln})' \xrightarrow{\mathcal{L}} (U_1, U_2, \dots, U_l)',$$

where  $(U_1, U_2, \dots, U_l) \sim N(\mathbf{0}, \mathbf{I})$ . Note that

$$(\hat{G}_n, \hat{T}_n) = \left( \sum_{i=1}^l \gamma_i^2 U_{in}^2, \sum_{i=1}^l 2^{1/2} \gamma_i U_{in} \right) + \left( \sum_{i=l+1}^{n-1} \gamma_i^2 U_{in}^2, \sum_{i=l+1}^{n-1} 2^{1/2} \gamma_i U_{in} \right).$$

Because the sequence  $\{\gamma_i^2\}_{i=1}^{\infty}$  is summable, the vector

$$\left( \sum_{i=l+1}^{n-1} \gamma_i^2 U_{in}^2, \sum_{i=l+1}^{n-1} 2^{1/2} \gamma_i U_{in} \right)$$

converges to zero in probability as  $l$  tends to infinity, uniformly in  $n$ . Therefore, by Lemma 6.3.1,  $(G_n, T_n)$  converges in distribution to  $(G, T)$  and the result is established.  $\blacktriangle$

We gave the proof of Theorem 10.1.1 for  $e_t$  that are martingale differences. The result also holds for  $e_t$  that are iid( $0, \sigma^2$ ) random variables.

The limiting distribution of Theorem 10.1.1 can be obtained by using Theorems 5.3.5 and 5.3.6. By Corollary 5.3.6, with  $\sigma^2 = 1$ ,

$$(G_n, T_n) \xrightarrow{\mathcal{L}} (G, T),$$

where

$$(G, T) = \left[ \int_0^1 W^2(t) dt, W(1) \right] \quad (10.1.14)$$

and  $W(t)$  is the standard Wiener process defined in Section 5.3.

In Theorem 10.1.1, we assumed  $Y_0 = 0$  to simplify the presentation. Because

$$\text{plim} \left[ \sum_{t=1}^n \left( \sum_{j=1}^t e_j \right)^2 \right]^{-1} \sum_{t=1}^n \left( Y_0 + \sum_{j=1}^t e_j \right)^2 = 1$$

for any finite  $Y_0$ , the conclusion of Theorem 10.1.1 is true for any finite  $Y_0$ .

The first part of Table 10.A.1 contains percentiles for the empirical distribution of  $\hat{\rho}$  given that  $\rho = 1$ . Normally distributed errors were used to construct the table for  $n = 25, 50, 100, 250$ , and  $500$ . The total number of observations available is  $n$ , while  $n - 1$  sums of squares and products are used in calculating the regression coefficient. The entries for  $n = \infty$  are the percentiles for the limiting distribution of Theorem 10.1.1.

The table may also be used for the distribution of  $\hat{\rho}$  for  $\rho = -1$ , since, for symmetrically distributed  $e_t$ , the distribution for  $\rho = -1$  is the mirror image of the distribution for  $\rho = 1$ . If the  $e_t$  are not symmetrically distributed, the limiting distribution for  $\rho = -1$  remains the mirror image of the limiting distribution for  $\rho = 1$ .

**Corollary 10.1.1.1.** Let the model of Theorem 8.5.1 hold. Then

$$\lim_{n \rightarrow \infty} (P\{\hat{\rho} - \rho > a \mid \rho = 1\} - P\{\hat{\rho} - \rho < -a \mid \rho = -1\}) = 0$$

for all real  $a$ . If, in addition, the  $e_t$  are independently and identically distributed with distribution that is symmetric about zero, then

$$P\{\hat{\rho} - \rho > a \mid \rho = 1\} = P\{\hat{\rho} - \rho < -a \mid \rho = -1\}$$

for all real  $a$  and all  $n \geq 2$ .

**Proof.** Let  $X_t = \sum_{j=0}^{t-1} (-1)^j e_{t-j}$ . In a manner similar to that used to obtain (10.1.13), we have

$$\sum_{t=2}^n X_{t-1} e_t = \sum_{t=2}^n e_t \sum_{j=1}^{t-1} (-1)^{j-1} e_{t-j} = 0.5 \left( \sum_{t=1}^n e_t^2 - X_n^2 \right).$$

Therefore, for the estimator of  $\rho$  when  $\rho = -1$ ,  $P\{\hat{\rho} + 1 < a\}$  is

$$\begin{aligned} P\left\{ \sum_{t=2}^n X_{t-1} e_t + a \sum_{t=2}^n X_{t-1}^2 < 0 \right\} \\ = P\left\{ 0.5 \left( X_n^2 - \sum_{t=1}^n e_t^2 \right) - a \sum_{t=1}^{n-1} \left[ \sum_{j=0}^{t-1} (-1)^j e_{t-j} \right]^2 > 0 \right\}. \end{aligned}$$

Although the sign of  $e_i$  in the weighted sum  $\sum_{j=0}^{t-1} (-1)^j e_{i-j}$  is not the same for all  $t > i$ , the sign is always opposite of that for  $e_{i-1}$  and  $e_{i+1}$ , and it follows that

$$\sum_{t=1}^{n-1} X_t^2 = \sum_{t=1}^{n-1} \left[ \sum_{j=1}^t (-1)^j e_j \right]^2.$$

If the  $e_t$  are iid( $0, \sigma^2$ ) and if the distribution of  $e_t$ ,  $t = 1, 2, \dots$ , is symmetric, then the distributional properties of the sequence  $(-e_1, e_2, -e_3, \dots)$  are the same as the distributional properties of the sequence  $(e_1, e_2, e_3, \dots)$ , and we conclude that, for symmetric  $e_t$  and for any  $a$ ,

$$\begin{aligned} P\left\{ \frac{1}{2} \left[ \left( \sum_{t=1}^n e_t \right)^2 - \sum_{t=1}^n e_t^2 \right] - a \sum_{t=1}^{n-1} \left( \sum_{j=1}^t e_j \right)^2 > 0 \right\} \\ = P\left\{ \frac{1}{2} \left[ \left( \sum_{t=1}^n (-1)^t e_t \right)^2 - \sum_{t=1}^n e_t^2 \right] - a \sum_{t=1}^{n-1} \left( \sum_{j=1}^t (-1)^j e_j \right)^2 > 0 \right\}. \end{aligned}$$

Also, the large sample distribution of  $\hat{\rho}$  given  $\rho = -1$  is the mirror image of the distribution for  $\rho = 1$  for  $e_t$  that are not symmetric, because the  $U_t$  have a limiting normal distribution for the sequence  $(-e_1, e_2, -e_3, \dots)$  as well as for the sequence  $(e_1, e_2, e_3, \dots)$ .  $\blacktriangle$

A natural statistic to use in testing the hypothesis that  $\rho = 1$  is the test statistic one would calculate in ordinary linear regression:

$$\hat{\tau} = \left[ s^{-2} \sum_{t=2}^n Y_{t-1}^2 \right]^{1/2} (\hat{\rho} - 1), \quad (10.1.15)$$

where

$$s^2 = (n-2)^{-1} \sum_{t=2}^n \hat{e}_t^2 = (n-2)^{-1} \sum_{t=2}^n (Y_t - \hat{\rho} Y_{t-1})^2.$$

**Corollary 10.1.1.2.** Let the model of Theorem 10.1.1 hold. Then

$$\hat{\tau} \xrightarrow{\mathcal{L}} (4G)^{-1/2} (T^2 - 1),$$

where  $\hat{\tau}$  is defined in (10.1.15).

**Proof.** We have

$$\begin{aligned} s^2 &= (n-2)^{-1} \left[ \sum_{t=2}^n e_t^2 - (\hat{\rho} - \rho) \sum_{t=2}^n Y_{t-1} e_t \right] \\ &= (n-2)^{-1} \sum_{t=2}^n e_t^2 + O_p(n^{-1}). \end{aligned}$$

Therefore,  $s^2 \xrightarrow{P} \sigma^2$  by Corollary 5.3.8. Because

$$\begin{aligned}\hat{\tau} &= 2^{-1}(s^2 G_n)^{-1/2} \left( T_n^2 - n^{-1} \sum_{t=1}^n e_t^2 \right) \\ &= (4G_n)^{-1/2} (T_n^2 - 1) + o_p(1),\end{aligned}$$

we have the result. ▲

Percentage points for the empirical distribution of  $\hat{\tau}$  are given in Table 10.A.2. The percentiles for  $n = \infty$  are the percentiles of the distribution of Corollary 10.1.1.2.

To extend the results for the first order process with  $\rho = 1$  to the  $p$ th order autoregressive process, we consider the time series

$$Y_t = \sum_{j=1}^p Z_j, \quad t = 1, 2, \dots, \quad (10.1.16)$$

where  $\{Z_t; t \in (0, \pm 1, \pm 2, \dots)\}$  is a  $(p-1)$ st order stationary autoregressive time series with the representation

$$Z_t = \sum_{i=2}^p \theta_i Z_{t-i+1} + e_t, \quad (10.1.17)$$

the  $e_t$  are martingale difference random variables with the properties defined in Theorem 10.1.1, and the absolute value of the largest root of

$$m^{p-1} - \sum_{i=2}^p \theta_i m^{p-i} = 0 \quad (10.1.18)$$

is less than some  $\lambda < 1$ . By (10.1.16) and (10.1.17), we may write

$$Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} = e_t, \quad t = p+1, p+2, \dots, \quad (10.1.19)$$

where  $p-1$  of the roots of

$$m^p + \sum_{j=1}^p \alpha_j m^{p-j} = 0$$

are the  $p-1$  roots of (10.1.18), and the remaining root is one. In a problem where a root of one is suspected, it is operationally desirable and theoretically convenient to write the autoregressive equation so that the unit root is isolated as a coefficient. To this end, we write

$$Y_t = \theta_1 Y_{t-1} + \sum_{j=2}^p \theta_j (Y_{t-j+1} - Y_{t-j}) + e_t, \quad (10.1.20)$$

for  $t = p+1, p+2, \dots$ , where  $p \geq 2$ ,  $\theta_i = \sum_{j=1}^p \alpha_j$ ,  $i = 2, 3, \dots, p$ , and  $\theta_1 = -\sum_{j=1}^p \alpha_j$ . If there is a unit root,  $\theta_1 = 1$ .

If one knew that  $\theta_1 = 1$ , one would regress  $Y_t - Y_{t-1} = Z_t$  on  $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p+1}$  and obtain an estimator of  $(\theta_2, \theta_3, \dots, \theta_p)$ , say  $(\tilde{\theta}_2, \tilde{\theta}_3, \dots, \tilde{\theta}_p)$ , with the properties described in Section 8.2. It is interesting that, given  $\theta_1 = 1$ , the regression of  $Y_t$  on  $Y_{t-1}, Z_{t-1}, \dots, Z_{t-p+1}$  yields an estimator of  $\theta_1$ , say  $\hat{\theta}_1$ , such that the large sample distribution of  $nc(\hat{\theta}_1 - 1)$  is the same as that of  $n(\hat{\rho} - 1)$ , where  $\hat{\rho}$  is the estimator (10.1.3) and  $c$  is a constant defined in Theorem 10.1.2. The large sample distribution of  $(\tilde{\theta}_2, \tilde{\theta}_3, \dots, \tilde{\theta}_p)$  is the same as that of the regression coefficients in the regression of  $Z_t$  on  $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p+1}$ .

**Theorem 10.1.2.** Let  $Y_t$  be defined by (10.1.16). Suppose the  $e_t$  satisfy the assumptions of Theorem 10.1.1. Let  $\tilde{\theta} = (\tilde{\theta}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_p)'$  be the vector of regression coefficients obtained by regressing  $Y_t$  on  $Y_{t-1}, Z_{t-1}, \dots, Z_{t-p+1}$ ,  $t = p+1, p+2, \dots, n$ , and define  $D_n$  to be the diagonal matrix

$$D_n = \text{diag}\{n, n^{1/2}, \dots, n^{1/2}\}.$$

Define  $c = \sum_{i=0}^{\infty} w_i = (1 - \sum_{i=2}^p \theta_i)^{-1}$ , where

$$Z_t = \sum_{i=0}^{\infty} w_i e_{t-i},$$

and the  $w_i$  satisfy the homogeneous difference equation with characteristic equation (10.1.18) and initial conditions  $w_0 = 1$  and  $w_i = 0$  for  $i < 0$ . Let

$$\tilde{\theta} - \theta = \left[ \left( c \sum_{i=2}^n W_{i-1}^2 \right)^{-1} \sum_{i=2}^n W_{i-1} e_i, \tilde{\theta}_2 - \theta_2, \tilde{\theta}_3 - \theta_3, \dots, \tilde{\theta}_p - \theta_p \right]',$$

where  $W_i = \sum_{j=1}^i e_j$ , and  $(\tilde{\theta}_2, \tilde{\theta}_3, \dots, \tilde{\theta}_p)'$  is the vector of regression coefficients obtained by regressing  $Z_t$  on  $Z_{t-1}, Z_{t-2}, \dots, Z_{t-p+1}$ . Then

$$\plim_{n \rightarrow \infty} D_n(\tilde{\theta} - \theta) = 0.$$

Furthermore,  $n(\hat{\theta}_1 - 1)$  is independent of  $n^{1/2}(\tilde{\theta}_2 - \theta_2, \dots, \tilde{\theta}_p - \theta_p)$  in the limit.

**Proof.** Using the definition of  $Z_t$  as the weighted sum of past  $e_i$ , we can write

$$\begin{aligned} Y_t &= \sum_{j=1}^t \left( \sum_{i=0}^{\infty} w_i \right) e_j - \sum_{j=1}^t \sum_{i=j}^{\infty} w_i e_{t-j+1} + \sum_{j=0}^{\infty} e_{-j} \sum_{i=j+1}^{j+t} w_i \\ &= c W_t + Q_t + R_t, \end{aligned}$$

where  $Q_t = \sum_{j=1}^t g_j e_{t-j+1}$ ,  $g_j = -\sum_{i=j}^{\infty} w_i$ , and  $R_t = \sum_{j=0}^{\infty} e_{-j} \sum_{i=j+1}^{j+t} w_i$ . Since the absolute value of the largest root of (10.1.18) is less than  $\lambda < 1$ , by Exercise 2.24,  $|w_i|$  is less than  $M_1 \lambda^i$  and  $|g_j|$  is less than  $M_2 \lambda^j$  for some finite  $M_1$  and  $M_2$ . It

follows that  $E\{Q_t Q_{t+h}\}$  and  $E\{Z_t Z_{t+h}\}$  are bounded by a multiple of  $\lambda^{|h|}$ . Also, the expectations of  $W_t Q_j$ ,  $Y_t Z_j$ ,  $n^{-1} Y_t Y_j$ , and  $R_t^2$  are bounded. Using these results, it can be shown that

$$n^{-2} \left[ \sum_{t=p+1}^n Y_{t-1}^2 - c^2 \sum_{t=p+1}^n W_{t-1}^2 \right] = o_p(1), \quad (10.1.21)$$

and

$$n^{-3/2} \left[ \sum_{t=p+1}^n Y_{t-1} Z_{t-j} \right] = o_p(1), \quad j = 1, 2, \dots, p-1.$$

Let  $\mathbf{A}_n$  be the matrix of sums of squares and cross products of  $Y_{t-1}$ ,  $Z_{t-1}$ ,  $Z_{t-2}, \dots, Z_{t-p+1}$ , and let  $\mathbf{B}_n$  be the matrix

$$\mathbf{B}_n = \begin{pmatrix} c^2 \sum_{t=p+1}^n W_{t-1}^2 & \mathbf{0} \\ \mathbf{0} & \mathbf{C}_n \end{pmatrix},$$

where  $\mathbf{C}_n$  is the  $(p-1) \times (p-1)$  matrix of sums of squares and cross products of  $Z_{t-1}$ ,  $Z_{t-2}, \dots, Z_{t-p+1}$ . The lower right  $(p-1) \times (p-1)$  submatrix of  $\mathbf{A}_n$  is  $\mathbf{C}_n$ . Furthermore, the  $ij$ th element of  $n^{-1} \mathbf{C}_n$  is converging in probability to  $\gamma_2(i-j)$ . By the convergence results for  $n^{-2} \sum W_{t-1}^2$ , we have  $\mathbf{D}_n^{-1} \mathbf{B}_n \mathbf{D}_n^{-1} = o_p(1)$  and  $\mathbf{D}_n \mathbf{B}_n^{-1} \mathbf{D}_n = o_p(1)$ . Therefore, using equation (10.1.21),

$$\mathbf{D}_n (\mathbf{A}_n^{-1} - \mathbf{B}_n^{-1}) \mathbf{D}_n = o_p(1).$$

Similarly,

$$\begin{aligned} n^{-1} \left( \sum_{t=p+1}^n Y_{t-1} e_t - c \sum_{t=p+1}^n W_{t-1} e_t \right) &= o_p(1), \\ \frac{n^{-1} \sum_{t=p+1}^n Y_{t-1} e_t}{n^{-2} \sum_{t=p+1}^n Y_{t-1}^2} &= \frac{n c \sum_{t=p+1}^n W_{t-1} e_t}{c^2 \sum_{t=p+1}^n W_{t-1}^2} + o_p(1), \end{aligned}$$

and the probability limit result is established.

From (10.1.13),

$$n^{-1} \sum_{t=p+1}^n W_{t-1} e_t = 0.5(n^{1/2} \bar{\epsilon})^2 - 0.5\sigma^2 + O_p(n^{-1/2}).$$

Let  $\mathbf{U}_{n,l} = (U_{1n}, U_{2n}, \dots, U_{ln})$ ,  $\mathbf{X}_t = (Z_{t-1}, Z_{t-2}, \dots, Z_{t-p+1})$ , and  $\mathbf{A} = E\{\mathbf{X}_t' \mathbf{X}_t\}$ , where  $U_{in}$  is defined in (10.1.12). Now

$$n^{-1/2} \left( \sum_{t=1}^l b_{1t} m_{nt} + b_2 + \sum_{j=1}^{p-1} b_{3j} Z_{t-j} \right) e_t,$$

where  $m_{nt}$  is defined in (10.1.10) and  $b_{11}, \dots, b_{1l}, b_2, b_{31}, \dots, b_{3,p-1}$  are fixed constants, satisfies the assumptions of Theorem 5.3.4. Therefore,

$$\left[ \mathbf{U}_{n,l}, n^{-1/2} \sum_{t=1}^n e_t, \left( n^{-1/2} \mathbf{A}^{-1/2} \sum_{t=1}^n \mathbf{X}'_t e_t \right)' \right]' \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I}\sigma^2)$$

for every fixed  $l$ . It follows, by Lemma 6.3.1, that

$$\left[ G_n, T_n, \left( n^{-1/2} \mathbf{A}^{-1/2} \sum_{t=1}^n \mathbf{X}'_t e_t \right)' \right]$$

converges in distribution to  $(G, T, \xi')$ , where  $(G_n, T_n)$  is defined in the proof of Theorem 10.1.1,  $\xi$  is a  $N(\mathbf{0}, \mathbf{I}\sigma^2)$  random variable, and  $(G, T)$  is independent of  $\xi$ .  $\blacktriangle$

**Corollary 10.1.2.1.** Let the assumptions of Theorem 10.1.2 hold. Let

$$\begin{aligned} \mathbf{K}_n^2 &= \text{diag} \left\{ \sum_{t=p+1}^n Y_{t-1}^2, \sum_{t=p+1}^n Z_{t-1}^2, \sum_{t=p+1}^n Z_{t-2}^2, \dots, \sum_{t=p+1}^n Z_{t-p+1}^2 \right\}, \\ \tilde{\mathbf{K}}_n^2 &= \text{diag} \left\{ c^2 \sum_{t=p+1}^n W_{t-1}^2, \sum_{t=p+1}^n Z_{t-1}^2, \sum_{t=p+1}^n Z_{t-2}^2, \dots, \sum_{t=p+1}^n Z_{t-p+1}^2 \right\}. \end{aligned}$$

Then

$$\plim_{n \rightarrow \infty} [\mathbf{K}_n(\hat{\theta} - \theta) - \tilde{\mathbf{K}}_n(\tilde{\theta} - \theta)] = \mathbf{0},$$

where  $W_t$ ,  $c$ ,  $\hat{\theta}$ , and  $\tilde{\theta}$  are defined in Theorem 10.1.1.

**Proof.** By the proof of Theorem 10.1.2,

$$n^{-1} \left( \sum_{t=p+1}^n Y_{t-1}^2 \right)^{1/2} = n^{-1} \left( c^2 \sum_{t=p+1}^n W_{t-1}^2 \right)^{1/2} + o_p(1),$$

and  $n^{-1} (\sum_{t=p+1}^n Y_{t-1}^2)^{1/2} = O_p(1)$ . Because

$$\plim_{n \rightarrow \infty} n^{-1} \sum_{t=p+1}^n Z_{t-1}^2 = \gamma_Z(0)$$

by Theorem 6.3.5, the result follows.  $\blacktriangle$

On the basis of Corollary 10.1.2.1, Table 10.A.2 can be used to investigate the

hypothesis that one of the roots in a  $p$ th order autoregressive process is unity. That is, the pivotal statistic for  $\theta_1$ , computed by an ordinary least squares program,

$$\hat{\tau} = [\hat{V}\{\hat{\theta}_1\}]^{-1/2}(\hat{\theta}_1 - 1), \quad (10.1.22)$$

where  $\hat{V}\{\hat{\theta}_1\}$  is the ordinary least squares estimator of the variance of  $\hat{\theta}_1$ , has the limiting distribution given in Corollary 10.1.1.2. Also, the regression pivots can be used to test hypotheses about the coefficients of the stationary process. For example, the hypothesis that the original process is of order  $p - 1$  is equivalent to the hypothesis that  $\alpha_p = 0$ , which is equivalent to the hypothesis that  $\theta_p = 0$ . A test statistic for this hypothesis is the ordinary regression pivotal statistic

$$t_p = [\hat{V}\{\hat{\theta}_p\}]^{-1/2}\hat{\theta}_p,$$

where  $\hat{V}\{\hat{\theta}_p\}$  is the ordinary least squares estimator of the variance of  $\hat{\theta}_p$ . Under the assumptions of Theorem 10.1.2,  $t_p$  is asymptotically normal when  $\theta_p = 0$ .

We note that the coefficient for  $Y_{t-1}$  in the regression of  $Y_t$  on  $Y_{t-1}, Z_{t-1}, \dots, Z_{t-p+1}$  is not the largest root of the fitted autoregressive equation. To see this, consider the second order process. The fitted equation is

$$\hat{Y}_t = (\hat{m}_1 + \hat{m}_2)Y_{t-1} - \hat{m}_1\hat{m}_2Y_{t-2},$$

where  $(\hat{m}_1, \hat{m}_2)$  are the two roots of the estimated characteristic equation. It follows that the estimated equation in  $Y_{t-1}$  and  $Z_{t-1}$  is

$$\hat{Y}_t = [\hat{m}_1 + \hat{m}_2(1 - \hat{m}_1)]Y_{t-1} + \hat{m}_1\hat{m}_2(Y_{t-1} - Y_{t-2}).$$

Also, the error in the coefficient of  $Y_{t-1}$  as an estimator of one is

$$\hat{\theta}_1 - 1 = (\hat{m}_1 - 1)(1 - \hat{m}_2) = (\hat{m}_1 - 1)(1 - m_2) + O_p(n^{-3/2}).$$

Because  $c = (1 - m_2)^{-1}$ , the limiting distribution of  $n(\hat{m}_1 - 1)$  is the same as the limiting distribution of  $n(\hat{\rho} - 1)$  given in Table 10.A.1. We state the generalization of this property as a corollary.

**Corollary 10.1.2.2.** Assume that the model (10.1.16)–(10.1.17) holds, and express  $Y_t$  in the form (10.1.19), where one of the roots of

$$m^p + \sum_{i=1}^p \alpha_i m^{p-i} = 0$$

is equal to one and the other roots are less than one in absolute value. Let  $\tilde{\alpha}' = (\tilde{\alpha}_1, \tilde{\alpha}_2, \dots, \tilde{\alpha}_p)$  be the least squares estimator of  $\alpha$  estimated subject to the restriction that  $\tilde{m}_1$  is real, where  $(\tilde{m}_1, \tilde{m}_2, \dots, \tilde{m}_p)$  are the roots of the estimated characteristic equation with  $\tilde{m}_1 \geq |\tilde{m}_2| \geq \dots \geq |\tilde{m}_p|$ . Then

$$n(\tilde{m}_1 - 1) \xrightarrow{\mathcal{L}} (2G)^{-1}(T^2 - 1),$$

where  $(G, T)$  is defined in Theorem 10.1.1.

**Proof.** By (10.1.16) and (10.1.17), we have

$$m^p + \sum_{i=1}^p \alpha_i m^{p-i} = (m-1) \left( m^{p-1} - \sum_{i=1}^{p-1} \theta_{i+1} m^{p-i-1} \right),$$

$\hat{\theta}_1 - 1 = O_p(n^{-1})$ , and  $\hat{\theta}_i - \theta_i = O_p(n^{-1/2})$  for  $i = 2, \dots, p$ , where the  $\hat{\theta}_i$ ,  $i = 1, 2, \dots, p$ , are the regression coefficients of Theorem 10.1.2. Therefore, by Corollary 5.8.1, the roots of

$$m^p + \sum_{i=1}^p \hat{\alpha}_i m^{p-i} = 0,$$

where the relation between  $\hat{\theta}_i$  and  $\hat{\alpha}_i$  is defined by (10.1.20), converge in probability to the roots of the equation defined with  $(\alpha_1, \dots, \alpha_p)$ . The estimated polynomial for the ordinary least squares estimator evaluated at  $m = 1$  is

$$1 + \sum_{i=1}^p \hat{\alpha}_i = \prod_{i=1}^p (1 - \hat{m}_i) = (1 - \hat{m}_1) \prod_{i=2}^p (1 - \hat{m}_i).$$

Using the fact that  $\hat{\theta}_1 = -\sum_{i=1}^p \hat{\alpha}_i$ , and the limiting result for the roots,

$$\begin{aligned} \hat{\theta}_1 - 1 &= (\hat{m}_1 - 1) \left( \prod_{i=2}^p (1 - m_i) + o_p(1) \right) \\ &= (\hat{m}_1 - 1) \left( 1 - \sum_{i=2}^p \theta_i + o_p(1) \right). \end{aligned}$$

This equality also demonstrates that given  $\epsilon > 0$ , there is some  $N_\epsilon$  such that for  $n > N_\epsilon$ , the probability that  $\hat{m}_1$ , computed with the ordinary least squares estimator of  $\hat{\alpha}$ , is real is greater than  $1 - \epsilon$ . The conclusion follows because we demonstrated in the proof of Theorem 10.1.2 that

$$n(\hat{\theta}_1 - 1) \xrightarrow{\mathcal{L}} (2G)^{-1} \left( 1 - \sum_{i=2}^p \theta_i \right) (T^2 - 1). \quad \blacktriangle$$

For the model (10.1.1) with  $|\rho| < 1$ , the limiting behavior of the estimator of  $\rho$  is the same whether the mean is known or estimated. The result is no longer true when  $\rho = 1$ . Consider the estimator

$$\hat{\rho}_\mu = \frac{\sum_{t=2}^n (Y_t - \bar{y}_{(0)})(Y_{t-1} - \bar{y}_{(-1)})}{\sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})^2}, \quad (10.1.23)$$

where

$$[\bar{y}_{(0)}, \bar{y}_{(-1)}] = (n-1)^{-1} \sum_{t=2}^n (Y_t, Y_{t-1}).$$

When  $\rho = 1$ , we have  $(3^{-1}n)^{-1/2} \bar{y} \xrightarrow{\mathcal{D}} N(0, 1)$  and this random variable makes a contribution to the limiting distribution of  $n(\hat{\rho}_\mu - 1)$ . Also see the moment results of Exercise 10.4. The limit random variable associated with  $\bar{y}$  is denoted by  $H$  in Theorem 10.1.3.

**Theorem 10.1.3.** Let the assumptions of Theorem 10.1.1 hold. Let  $\hat{\rho}_\mu$  be defined by (10.1.23), and let

$$\hat{\tau}_\mu = [\hat{V}\{\hat{\rho}_\mu\}]^{-1/2}(\hat{\rho}_\mu - 1), \quad (10.1.24)$$

where

$$\hat{V}\{\hat{\rho}_\mu\} = \left[ \sum_{t=2}^n (Y_t - \bar{y}_{(-1)})^2 \right]^{-1} s^2$$

and

$$s^2 = (n-3)^{-1} \sum_{t=2}^n [Y_t - \bar{y}_{(0)} - \hat{\rho}_\mu(Y_{t-1} - \bar{y}_{(-1)})]^2.$$

Then

$$n(\hat{\rho}_\mu - 1) \xrightarrow{\mathcal{D}} (G - H^2)^{-1}[0.5(T^2 - 1) - TH]$$

and

$$\hat{\tau}_\mu \xrightarrow{\mathcal{D}} (G - H^2)^{-1/2}[0.5(T^2 - 1) - TH],$$

where

$$H = \sum_{i=1}^{\infty} 2^{1/2} \gamma_i^2 U_i = \int_0^1 W(t) dt,$$

$U_i \sim NI(0, 1)$ ,  $W(t)$  is the Wiener process, and  $G$ ,  $T$ , and  $\gamma$  are defined in Theorem 10.1.1.

**Proof.** The method of proof is the same as that of Theorem 10.1.1. Let

$$H_n = n^{-3/2} \sum_{t=2}^n Y_{t-1},$$

and define  $\mathbf{U}_n$  and  $\mathbf{M}_n$  as in the proof of Theorem 10.1.1. Then

$$H_n = \mathbf{L}'_n \mathbf{e}_n = \mathbf{L}'_n \mathbf{M}_n^{-1} \mathbf{U}_n,$$

where  $\mathbf{L}'_n = n^{-3/2}(n-1, n-2, \dots, 1)$ . Let  $q_{in}$  be the  $i$ th element of  $\mathbf{L}'_n \mathbf{M}_n^{-1}$ . This element is the covariance between  $H_n$  and  $\mathbf{U}_n$ , and it can be shown that for fixed  $i$

$$\lim_{n \rightarrow \infty} q_{in} = 2^{1/2} \gamma_i^2.$$

The remainder of the proof follows that of Theorem 10.1.1. ▲

The second part of Table 10.A.1 contains empirical percentiles for the distribution of  $n(\hat{\rho}_\mu - 1)$ , and the second part of Table 10.A.2 contains the empirical percentiles of the corresponding studentized statistic given that  $\rho = 1$ . The values for  $n = \infty$  are the percentiles of the limiting distribution of Theorem 10.1.3.

It can be demonstrated that  $\text{plim } n(\hat{\rho}_\mu - \rho) = 0$  when  $\rho = -1$ . That is, estimating the mean does not alter the limiting distribution when  $\rho = -1$ . Therefore, the first part of Tables 10.A.1 and 10.A.2 can be used to approximate the distributions of  $\hat{\rho}_\mu$  and  $\hat{\tau}_\mu$  when  $\rho = -1$ .

The distributions of Theorem 10.1.3 are obtained under the model (10.1.1) with  $\rho = 1$ . This point can be emphasized by writing an extended model as

$$Y_t = (1 - \rho)\mu + \rho Y_{t-1} + e_t,$$

where  $\rho$  and  $\mu$  are unknown parameters. The null model for the test  $\hat{\tau}_\mu$  is  $\rho = 1$ . The extended model then reduces to (10.1.1). The alternative model with  $\rho \neq 1$  permits a nonzero value for  $\theta_0 = (1 - \rho)\mu$ . Thus, the test based on  $\hat{\tau}_\mu$  is invariant to the mean  $\mu$  of the alternative model.

The limiting distribution of  $n^{3/2}(\hat{\theta}_1 - 1)$ , where  $\hat{\theta}_1$  is the least squares estimator and  $\theta_1 = 1$ , for the model

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + e_t$$

with  $\theta_0 \neq 0$  is normal, and is discussed in Section 10.1.2.

It follows from Theorem 10.1.3 that the distribution of the ordinary least squares estimator of  $\theta_0$  for the model (10.1.1) with  $\rho = 1$  is not normal. Let  $\Delta Y_t = Y_t - Y_{t-1}$ , and consider the equation

$$\Delta Y_t = \theta_0 + (\theta_1 - 1)Y_{t-1} + e_t, \quad (10.1.25)$$

with the unknown  $\theta_0$  equal to zero. The estimated regression equation can be written

$$\Delta \hat{Y}_t = \Delta \bar{Y} + (\hat{\theta}_1 - 1)(Y_{t-1} - \bar{y}_{(-1)}), \quad (10.1.26)$$

where  $\bar{y}_{(-1)}$  is defined in (10.1.23),  $\Delta \bar{Y} = (n-1)^{-1} \sum_{t=2}^n \Delta Y_t$ , and

$$\hat{\theta}_1 - 1 = \left( \sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})^2 \right)^{-1} \sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})(\Delta Y_t - \Delta \bar{Y}).$$

Under the null model,  $\Delta Y_t = e_t$ , and  $\Delta \bar{Y} = \bar{e}$ . It follows that

$$\hat{\theta}_0 = \bar{e} - (\hat{\theta}_1 - 1)\bar{y}_{(-1)}$$

and

$$n^{1/2} \hat{\theta}_0 \xrightarrow{\mathcal{L}} T - (G - H^2)^{-1}[0.5(T^2 - 1) - TH]H, \quad (10.1.27)$$

where  $G$ ,  $T$ , and  $H$  are defined in Theorem 10.1.1 and Theorem 10.1.3. Dickey (1976) and Dickey and Fuller (1981) have discussed the distribution of  $n^{1/2}\hat{\theta}_0$ . Under the normal distribution null model,  $\Delta \bar{Y}$  of (10.1.26) is distributed as a normal random variable. Also, because  $\hat{\theta}_1$  is a quadratic function of  $(Y_1, \dots, Y_n)$ ,  $\Delta \bar{Y}$  is uncorrelated with  $\hat{\theta}_1$  for normal  $e_t$ . Therefore, estimating the equation in the form (10.1.26) yields two uncorrelated estimators whose null distributions are easily interpreted.

The distributional results for the estimator of the parameters of the first order autoregressive model with intercept extend to the  $p$ th order process in much the same manner as the results for the model with no intercept. It follows from Theorem 10.1.4 that the regression pivotal calculated to test  $\theta_1 = 1$  has the tabled distribution  $\hat{\tau}_\mu$  when the fitted model has an intercept.

**Theorem 10.1.4.** Let  $Y_t$  be defined by (10.1.16). Suppose the assumptions of Theorem 10.1.2 hold. Let  $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)'$  be the vector of regression coefficients obtained by regressing  $Y_t$  on  $(1, Y_{t-1}, Z_{t-1}, \dots, Z_{t-p+1})$ ,  $t = p + 1, p + 2, \dots, n$ . Define  $D_n$  to be the diagonal matrix

$$D_n = \text{diag}(n^{1/2}, n, n^{1/2}, n^{1/2}, \dots, n^{1/2}),$$

and let  $\tilde{\theta} - \theta$  be the vector such that  $\tilde{\theta}_0 = \hat{\theta}_0$ ;

$$\tilde{\theta}_1 - \theta_1 = \left[ c \sum_{t=2}^n (W_{t-1} - \bar{w}_{(-1)})^2 \right]^{-1} \sum_{t=2}^n (W_{t-1} - \bar{w}_{(-1)})e_t,$$

where  $W_t$  and  $c$  are defined in Theorem 10.1.2, and  $\bar{w}_{(-1)} = (n-1)^{-1} \sum_{t=1}^{n-1} W_t$ , and  $(\tilde{\theta}_2, \tilde{\theta}_3, \dots, \tilde{\theta}_p)$  is the vector of regression coefficients obtained by regressing  $Z_t$  on  $(Z_{t-1}, Z_{t-2}, \dots, Z_{t-p+1})$ . Let  $\tau_\mu^*$  be the ordinary regression pivotal computed by dividing  $\tilde{\theta}_1 - 1$  by the regression estimated standard error of  $\tilde{\theta}_1$ . Then

$$\plim_{n \rightarrow \infty} D_n(\tilde{\theta} - \theta) = \mathbf{0},$$

$n(\hat{\theta}_1 - \theta_1)$  is independent of  $n^{1/2}[\hat{\theta}_2 - \theta_2, \dots, \hat{\theta}_p - \theta_p]$  in the limit, and the limiting distribution of  $\tau_\mu^*$  is the limiting distribution of  $\hat{\tau}_\mu$  given in Theorem 10.1.3.

**Proof.** Omitted. ▲

There has been extensive research on models with unit roots. We give only a few references. Hasza and Fuller (1982) studied models with two unit roots. Dickey, Hasza, and Fuller (1984) gave results for seasonal models, and Tiao and Tsay (1983) gave results for models with complex roots. Said and Dickey (1984) show that fitting a high order autoregressive model is an appropriate way to test for a unit root in a model of unknown form. Chan and Wei (1988) and Pantula (1989) give results that cover models with a number of unit roots. Phillips (1986, 1987a-c, 1988) and his co-workers have given many results on both univariate and vector processes. Shin (1990) and Yap and Reinsel (1995a,b) give tests for a unit root in the autoregressive portion of an autoregressive moving average. Ahtola and Tiao (1984), Chan (1988), Chan and Wei (1987), and Phillips (1987b) have studied the behavior of estimators when  $\rho$  is close to one but not equal to one.

**Example 10.1.1.** In this example we analyze the one-year treasury bill interest rate for the period January 1960 through August 1979. This is one of three interest rate series studied by Stock and Watson (1988). The data are given in Table 10.B.1 of Appendix 10.B. We postulate the time series to be an autoregressive process, and we are interested in testing the hypothesis that the process has a unit root. In this example, we use the ordinary least squares method of fitting. If we fit a sixth order autoregressive process, the fitted equation is

$$\begin{aligned}\hat{Y}_t = & 0.104 + 1.332 Y_{t-1} - 0.453 Y_{t-2} + 0.127 Y_{t-3} \\& (0.064) \quad (0.067) \quad (0.111) \quad (0.115) \\& + 0.043 Y_{t-4} - 0.079 Y_{t-5} + 0.014 Y_{t-6} \\& (0.115) \quad (0.111) \quad (0.067)\end{aligned}$$

and the regression residual mean square is 0.0833. There are 230 observations in the regression, and the residual mean square has 223 degrees of freedom. The coefficients for  $Y_{t-3}$ ,  $Y_{t-4}$ ,  $Y_{t-5}$ , and  $Y_{t-6}$  are small relative to the least squares standard errors. If we drop  $Y_{t-6}$  from the regression, the  $t$ -statistic for  $Y_{t-5}$  is  $-1.036$ . If we drop  $Y_{t-5}$  from the regression, the  $t$ -statistic for  $Y_{t-4}$  is  $-0.884$ , and if we drop  $Y_{t-4}$  from the regression, the  $t$ -statistic for  $Y_{t-3}$  is  $1.427$ . By Theorem 8.2.1, the  $t$ -statistics are approximately normally distributed for a stationary process. By Theorem 10.1.4, the  $t$ -statistics are approximately normally distributed if the process contains a single unit root. Thus, the testing procedure is appropriate in either case. The  $F$ -test for the coefficients of  $Y_{t-4}$ ,  $Y_{t-5}$ , and  $Y_{t-6}$ , computed as the sum of the squared  $t$ -statistics divided by 3, is 0.63. We compute the  $F$ -statistic in this way because different numbers of observations were used in the different regressions. The value of  $F$  is such that one easily accepts the hypothesis that the three coefficients are zero. The  $t$ -statistic for the coefficient of  $Y_{t-3}$  is such that one could either drop the coefficient or retain it in the analysis. We choose to retain the coefficient, and we proceed with the analysis using the third order model.

We are interested in testing for a unit root against the alternative of a stationary process, but we are willing to entertain the possibility of two unit roots. Dickey and Pantula (1987) suggest a sequential testing procedure in such situations. One begins with the greatest number of unit roots one is willing to consider. In our case, we begin by testing for a second unit root under the maintained hypothesis of a single unit root. We are willing to assume there is no intercept in the model for the differences. The estimated regression is

$$\Delta^2 \hat{Y}_t = -0.763 \Delta Y_{t-1} + 0.103 \Delta^2 Y_{t-1}, \\ (0.076) \quad (0.065)$$

where the numbers in parentheses are the regression estimated standard errors. Because the  $\tau$ -statistic  $\hat{\tau} = (0.076)^{-1}(-0.763) = -10.09$  is less than the 1% tabular value of  $-2.58$  in the first part of Table 10.A.2, we reject the hypothesis of a second unit root.

The regression equation for the test of a unit root in the original time series is

$$\Delta \hat{Y}_t = 0.082 - 0.012 Y_{t-1} + 0.343 \Delta Y_{t-1} - 0.094 \Delta Y_{t-2}. \\ (0.062) \quad (0.011) \quad (0.065) \quad (0.066)$$

Because the  $\hat{\tau}_\mu$  statistic of  $-1.09$  for the coefficient of  $Y_{t-1}$  is greater than the 10% tabular value of  $-2.57$  from the second part of Table 10.A.2, we would accept the hypothesis of a unit root at that level. The analysis of these data is extended in Example 10.1.2.  $\blacktriangle \blacktriangle$

### 10.1.2. Random Walk with Drift

In this section, we obtain the limiting distribution of the least squares estimator for  $(\theta_0, \theta_1)$  of the model

$$Y_t = \begin{cases} \theta_0 + \theta_1 Y_{t-1} + e_t, & t = 1, 2, \dots, \\ Y_0, & t = 0, \end{cases} \quad (10.1.28)$$

under the assumption that  $\theta_0 \neq 0$  and  $\theta_1 = 1$ . The model (10.1.28) with  $\theta_0 \neq 0$  is sometimes called a random walk with *drift*, where  $\theta_0$  is the drift parameter. Under (10.1.28)

$$Y_t = Y_0 + \theta_0 t + W_t, \quad (10.1.29)$$

where  $W_t = \sum_{j=1}^t e_j$ . When  $\theta_0 \neq 0$ , the time trend will dominate the long-run behavior of  $Y_t$  of (10.1.29). This is because  $W_t = O_p(t^{1/2})$  and hence  $W_t$  is small in probability relative to  $t$ . The least squares estimator of  $\theta = (\theta_0, \theta_1)'$  is  $\hat{\theta} = (\hat{\theta}_0, \hat{\theta}_1)'$ , where

$$\begin{aligned}\hat{\theta}_1 &= \left[ \sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})^2 \right]^{-1} \sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)}) Y_t, \\ \hat{\theta}_0 &= \bar{y}_{(0)} - \hat{\theta}_1 \bar{y}_{(-1)}, \\ \bar{y}_{(i)} &= (n-1)^{-1} \sum_{t=2}^n Y_{t+i}, \quad i = 0, -1.\end{aligned}\tag{10.1.30}$$

We give the limiting distribution of the least squares estimator under the assumption that  $\theta_0 \neq 0$  and  $\theta_1 = 1$ .

**Theorem 10.1.5.** Let the process  $Y_t$  satisfy (10.1.28) where  $\{e_t\}$  satisfies the conditions of Theorem 10.1.1 or is a sequence of independent identically distributed  $(0, \sigma^2)$  random variables. Assume  $\theta_0 \neq 0$  and  $\theta_1 = 1$ . Let the least squares estimator be given by (10.1.30). Then

$$\begin{pmatrix} n^{1/2}(\hat{\theta}_0 - \theta_0) \\ n^{3/2}(\hat{\theta}_1 - \theta_1) \end{pmatrix} \xrightarrow{\mathcal{L}} N \left[ \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & 2^{-1}\theta_0 \\ 2^{-1}\theta_0 & 3^{-1}\theta_0^2 \end{pmatrix}^{-1} \sigma^2 \right].$$

**Proof.** We have

$$\begin{aligned}\sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})^2 &= \theta_0^2 \sum_{t=2}^n (t - \bar{t})^2 + 2\theta_0 \sum_{t=2}^n (t - \bar{t})(W_{t-1} - \bar{w}_{(-1)}) \\ &\quad + \sum_{t=2}^n (W_{t-1} - \bar{w}_{(-1)})^2\end{aligned}$$

and

$$\sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)}) e_t = \theta_0 \sum_{t=2}^n (t - \bar{t}) e_t + \sum_{t=2}^n (W_{t-1} - \bar{w}_{(-1)}) e_t,$$

where  $\bar{t} = 0.5(n+1)$  and  $\bar{w}_{(-1)} = (n-1)^{-1} \sum_{t=2}^n W_{t-1}$ . By our previous results,

$$\sum_{t=2}^n (W_{t-1} - \bar{w}_{(-1)}) [n^{-2}(W_{t-1} - \bar{w}_{(-1)}), n^{-5/2}(t - \bar{t}), n^{-1}e_t] = O_p(1).$$

Therefore, using  $\sum_{t=2}^n (t - \bar{t})^2 = (12)^{-1}n(n-1)(n-2)$ ,

$$n^{3/2}(\hat{\theta}_1 - 1) = 12n^{-3/2}\theta_0^{-1} \sum_{t=2}^n (t - \bar{t}) e_t + O_p(n^{-1/2}).$$

Similarly,

$$\hat{\theta}_0 - \theta_0 = n^{-1} \sum_{t=1}^n e_t - 6n^{-2} \sum_{t=1}^n (t - \bar{t}) e_t + O_p(n^{-1})$$

and

$$n^{1/2}(\hat{\theta}_0 - \theta_0) \xrightarrow{\mathcal{L}} N(0, 4\sigma^2).$$

The joint limiting distribution follows by Theorem 6.3.4. ▲

The regression estimated equation for the model (10.1.28) can be written in the form

$$\Delta Y_t = \Delta \bar{Y} + (\hat{\theta}_1 - 1)(Y_{t-1} - \bar{y}_{t-1}), \quad (10.1.31)$$

introduced in (10.1.26). This form has the advantage that  $\Delta \bar{Y}$  is asymptotically normally distributed with mean  $\theta_0$  for all  $\theta_0$  when  $\theta_1 = 1$ .

A time series of the form (10.1.28) with  $\theta_0 \neq 0$  will tend, over time, to increase if  $\theta_0 > 0$  and to decrease if  $\theta_0 < 0$ . A natural alternative model for a time series displaying such behavior is as the sum of a stationary time series and a time trend. Let

$$Y_t = \phi_0 + \phi_1 t + X_t, \quad (10.1.32)$$

where  $X_t$  is an autoregressive time series satisfying

$$X_t = \theta_1 X_{t-1} + \sum_{j=2}^p \theta_j (X_{t-j+1} - X_{t-j}) + e_t. \quad (10.1.33)$$

By substituting (10.1.33) into (10.1.32), we obtain

$$Y_t = \psi_0 + \psi_1 t + \theta_1 Y_{t-1} + \sum_{j=2}^p \theta_j (Y_{t-j+1} - Y_{t-j}) + e_t, \quad (10.1.34)$$

where  $\psi_0 = \phi_0(1 - \theta_1) + \theta_1 \phi_1 - \phi_1 \sum_{j=2}^p \theta_j$  and  $\psi_1 = \phi_1(1 - \theta_1)$ . Thus, one way to investigate the hypothesis that  $Y_t$  contains a unit root is to fit the regression model (10.1.34) by ordinary least squares and test the hypothesis that  $\theta_1 = 1$ . The distribution of the test statistic differs from that obtained when time is not included in the regression.

**Theorem 10.1.6.** Let  $Y_t$  satisfy (10.1.34) where  $\{e_t\}_{t=1}^n$  is such that

$$\begin{aligned} E[(e_t, e_t^2) | \mathcal{A}_{t-1}] &= (0, \sigma^2) \quad \text{a.s.,} \\ E[|e_t|^{2+\delta} | \mathcal{A}_{t-1}] &< M < \infty \quad \text{a.s.} \end{aligned}$$

for some  $\delta > 0$  and  $t = 1, 2, \dots$ , where  $\mathcal{A}_t$  is the sigma-field generated by  $(e_1, e_2, \dots, e_t)$ . Assume  $\theta_1^0 = 1$ .

Let  $(\hat{\psi}_0, \hat{\psi}_1, \hat{\theta}_1, \hat{\theta}_2, \dots, \hat{\theta}_p)$  be the ordinary least squares regression coefficients, and let

$$\hat{\tau}_\tau = [\hat{V}\{\hat{\theta}_1\}]^{-1/2}(\hat{\theta}_1 - 1)$$

be the ordinary regression pivotal for  $\theta_1$ . Then

$$\hat{\tau}_r \xrightarrow{d} 0.5(G - H^2 - 3K^2)^{-1/2}[(T - 2H)(T - 6K) - 1],$$

where

$$K = \sum_{i=1}^{\infty} 2^{1/2}(2\gamma_i^3 - \gamma_i^2)U_i = 2 \int_0^1 tW(t) dt - H,$$

$W(t)$  is the Wiener process, and  $T$ ,  $G$ ,  $H$ ,  $\gamma_i$ , and  $U_i$  are defined in Theorem 10.1.1 and Theorem 10.1.3. Also, the limiting distribution of

$$n^{1/2}[(\hat{\theta}_2 - \theta_2^0), (\hat{\theta}_3 - \theta_3^0), \dots, (\hat{\theta}_p - \theta_p^0)]$$

is the same as the limiting distribution of the least squares estimator obtained by regressing  $\Delta Y_t$  on  $(\Delta Y_{t-1}, \dots, \Delta Y_{t-p+1})$ .

**Proof.** Omitted. See Dickey and Fuller (1979). ▲

In (10.1.34), we expressed  $Y_t$  as a function of the vector  $(1, t, Y_{t-1}, \Delta Y_{t-1}, \dots, \Delta Y_{t-p+1})$  and computed the estimates by ordinary least squares. A second estimation procedure that has been heavily used in practice is to adjust the observations by subtracting the mean functions as estimated by ordinary least squares. The conclusions of Theorem 10.1.4 and of Theorem 10.1.6 also hold for the coefficients obtained in the regression of  $y_t$  on  $(y_{t-1}, \Delta y_{t-1}, \dots, \Delta y_{t-p})$ , where  $y_t$  is the deviation of  $Y_t$  from the estimated mean function.

### 10.1.3. Alternative Estimators

In the stationary case, we presented a number of alternative estimators of the parameters of the autoregressive process. The maximum likelihood estimator associated with (8.1.9), the least squares estimator (8.2.6), and the weighted symmetric estimator associated with (8.2.14) all have the same limiting distribution when  $Y_t$  is a stationary autoregressive process. However, the limiting distributions differ if the process has a unit root.

We begin our discussion with the first order model and let

$$Y_t = \begin{cases} \rho Y_{t-1} + e_t, & t = 2, 3, \dots, \\ 0, & t = 0, \end{cases} \quad (10.1.35)$$

where the  $e_t$  are independent  $(0, \sigma^2)$  random variables. The simple symmetric estimator for  $\rho$  minimizes (8.2.14) with  $w_t = 0.5$  and is

$$\tilde{\rho}_s = \left[ \frac{1}{2}(Y_1^2 + Y_n^2) + \sum_{t=2}^{n-1} Y_t^2 \right]^{-1} \sum_{t=2}^n Y_{t-1} Y_t. \quad (10.1.36)$$

The error in the simple symmetric estimator is

$$\tilde{\rho}_s - \rho = \left[ (Y_n^2 - Y_1^2) + 2 \sum_{t=2}^n Y_{t-1}^2 \right]^{-1} \left[ 2 \sum_{t=2}^n Y_{t-1} e_t - \rho(Y_n^2 - Y_1^2) \right], \quad (10.1.37)$$

and if  $\rho = 1$ , the error reduces to

$$\tilde{\rho}_s - 1 = - \left[ (Y_n^2 - Y_1^2) + 2 \sum_{t=2}^n Y_{t-1}^2 \right]^{-1} \sum_{t=2}^n e_t^2, \quad (10.1.38)$$

where we have used (10.1.13).

An estimator of  $\sigma^2$  associated with (10.1.36) is  $\tilde{\sigma}_s^2 = (n-2)^{-1} Q_s(\tilde{\rho}_s)$ , where  $Q_s$  is (8.2.14) with  $w_t = 0.5$ . Then

$$\tilde{\sigma}_s^2 = (n-2)^{-1} \left[ \sum_{t=2}^n (Y_t - \tilde{\rho}_s Y_{t-1})^2 + 0.5(1 - \tilde{\rho}_s^2)(Y_1^2 - Y_n^2) \right]. \quad (10.1.39)$$

If this estimator of variance is used to construct a pivotal statistic, the pivotal reduces to a function of  $\tilde{\rho}_s$ . Letting

$$\hat{\tau}_s = \left\{ \tilde{\sigma}_s^2 [0.5(Y_1^2 + Y_n^2) + \sum_{t=2}^{n-1} Y_t^2]^{-1} \right\}^{-1/2} (\tilde{\rho}_s - 1), \quad (10.1.40)$$

we have

$$\hat{\tau}_s = -(n-2)^{1/2} (1 + \tilde{\rho}_s)^{-1/2} (1 - \tilde{\rho}_s)^{1/2}. \quad (10.1.41)$$

The estimator constructed with the weights of (8.2.15) is called the weighted symmetric estimator and can be written as

$$\tilde{\rho}_w = \left( \sum_{t=2}^{n-1} Y_t^2 + n^{-1} \sum_{t=1}^n Y_t^2 \right)^{-1} \sum_{t=2}^n Y_{t-1} Y_t. \quad (10.1.42)$$

The error in the weighted symmetric estimator when  $\rho = 1$  is

$$\tilde{\rho}_w - 1 = \left( \sum_{t=2}^{n-1} Y_t^2 + n^{-1} \sum_{t=1}^n Y_t^2 \right)^{-1} \left( \sum_{t=2}^n Y_{t-1} e_t + Y_1^2 - n^{-1} \sum_{t=1}^n Y_t^2 \right). \quad (10.1.43)$$

The pivotal statistic for the weighted symmetric estimator is

$$\hat{\tau}_w = \left[ \tilde{\sigma}_w^2 \left( \sum_{t=2}^{n-1} Y_t^2 + n^{-1} \sum_{t=1}^n Y_t^2 \right)^{-1} \right]^{-1/2} (\tilde{\rho}_w - 1), \quad (10.1.44)$$

where  $\tilde{\sigma}_w^2$  is the estimator of  $\sigma^2$  constructed by dividing  $Q_w(\tilde{\rho}_w)$  by  $n-2$ . The

limiting distributions of the symmetric estimators follow from (10.1.38) and (10.1.43).

**Theorem 10.1.7.** Let the assumptions of Theorem 10.1.1 hold, and let  $\tilde{\rho}_s$  be defined by (10.1.36),  $\hat{\tau}_s$  by (10.1.40),  $\tilde{\rho}_w$  by (10.1.42), and  $\hat{\tau}_w$  by (10.1.44). Then

$$\begin{aligned} n(\tilde{\rho}_s - 1) &\xrightarrow{\mathcal{L}} -(2G)^{-1}, \\ \hat{\tau}_s &\xrightarrow{\mathcal{L}} -0.5G^{-1/2}, \\ n(\tilde{\rho}_w - 1) &\xrightarrow{\mathcal{L}} 0.5G^{-1}(T^2 - 1) - 1, \\ \hat{\tau}_w &\xrightarrow{\mathcal{L}} 0.5G^{-1/2}(T^2 - 1) - G^{1/2}, \end{aligned}$$

where  $G$  and  $T$  are defined in Theorem 10.1.1.

**Proof.** From (10.1.38) we have

$$\begin{aligned} n(\tilde{\rho}_s - 1) &= -0.5 \left[ (2n^2)^{-1}(Y_n^2 + Y_1^2 + 2 \sum_{t=2}^{n-1} Y_t^2) \right]^{-1} \sigma^2 + o_p(1) \\ &= -0.5 \left[ n^{-2} \sum_{t=1}^n Y_t^2 \right]^{-1} \sigma^2 + o_p(1), \end{aligned}$$

and the first result follows from Theorem 10.1.1. The limiting distribution for  $\hat{\tau}_s$  is a consequence of (10.1.41).

Using (10.1.43), and the limiting distributions of  $n^{-2} \sum_{t=1}^n Y_t^2$ , and of  $n^{-1} \sum_{t=2}^n Y_{t-1} e_t$ , derived in Theorem 10.1.1, we obtain the results for  $n(\tilde{\rho}_w - 1)$  and  $\hat{\tau}_w$ . ▲

The first part of Table 10.A.3 contains percentiles of the distribution of  $\hat{\tau}_s$ , and the first part of Table 10.A.4 contains the percentiles of the distribution of  $\hat{\tau}_w$ . The percentiles were generated using the Monte Carlo method. The percentiles for the limiting distributions were constructed by approximating the infinite sums that define  $(G, T, H, K)$  with finite sums. The raw Monte Carlo percentiles were smoothed using a function of  $n^{-1}$ . The standard errors of the estimates of Table 10.A.3 are generally less than 0.01, and those of Table 10.A.4 are generally less than 0.007.

The difference  $\tilde{\rho}_s - 1$  is always negative, and this is reflected in the percentiles of Table 10.A.3. The percentiles of  $n(\tilde{\rho}_s - 1)$  are defined in terms of those of Table 10.A.3 by equation (10.1.41). The distribution of  $\tilde{\rho}_s$  has somewhat smaller tails than the distribution of the ordinary least squares estimator. For example, the distance between the 0.025 and 0.975 percentiles of the limiting distribution is 11.03 for  $\tilde{\rho}_s$ , and is 12.10 for  $\hat{\rho}$ . About 10% of the values of  $\tilde{\rho}_w$  are greater than one, and the 90th percentile of  $\hat{\tau}_w$  is close to zero.

The distributions of the symmetric estimators are a function of other variables included in the regression model, as is the distribution of the ordinary least squares estimator. Let  $\hat{f}_{yt}$  be an estimated "mean function" for the time series  $Y_t$ . Examples of  $\hat{f}_{yt}$  are the sample mean

$$\hat{f}_{yt1} = \bar{y}_n,$$

the linear regression trend function

$$\hat{f}_{yt2} = b_0 + b_1 t, \quad (10.1.45)$$

where  $(b_0, b_1)$  is the vector of regression coefficients obtained in the ordinary least squares regression of  $Y_t$  on  $(1, t)$ , and the quadratic regression trend function

$$\hat{f}_{yt3} = b_0 + b_1 t + b_2 t^2,$$

where  $(b_0, b_1, b_2)$  is the vector of regression coefficients obtained in the ordinary least squares regression of  $Y_t$  on  $(1, t, t^2)$ .

The limiting distributions of the  $\tau$ -statistics are given in Theorem 10.1.8. The limiting distributions of  $n(\tilde{\rho}_{sj} - 1)$ ,  $g = s, w$  and  $j = 1, 2$ , are obtained by squaring the denominators in the expressions. For example, the limiting distribution of  $n(\hat{\rho}_{s1} - 1)$  is the distribution of  $-0.5(G - H^2)^{-1}$ .

**Theorem 10.1.8.** Let the assumptions of Theorem 10.1.7 hold. Let  $\tilde{\rho}_{sj}$  be the estimator obtained by replacing  $Y_t$  with  $Y_t - \hat{f}_{ytj}$  in (10.1.36), and let  $\hat{\tau}_{sj}$  be the pivotal obtained by replacing  $Y_t$  with  $Y_t - \hat{f}_{ytj}$  in (10.1.40). Let  $\tilde{\rho}_{wj}$  and  $\hat{\tau}_{wj}$  be constructed in an analogous manner using (10.1.42) and (10.1.44), respectively. Then

$$\begin{aligned}\hat{\tau}_{s1} &\xrightarrow{\mathcal{L}} -0.5(G - H^2)^{-1/2}, \\ \hat{\tau}_{s2} &\xrightarrow{\mathcal{L}} -0.5(G - H^2 - 3K^2)^{-1/2}, \\ \hat{\tau}_{w1} &\xrightarrow{\mathcal{L}} (G - H^2)^{-1/2}[0.5(T^2 - 1) - TH - G + 2H^2], \\ \hat{\tau}_{w2} &\xrightarrow{\mathcal{L}} \frac{0.5[(T - 2H)(T - 6K) - 1] + (H - 3K)^2 - (G - H^2 - 3K^2)}{(G - H^2 - 3K^2)^{1/2}},\end{aligned}$$

where  $G$ ,  $T$ ,  $H$ , and  $K$  are defined in Theorem 10.1.6.

**Proof.** The limiting distributions of the simple symmetric test statistics are obtained by replacing  $Y_t$  with  $Y_t - \hat{f}_{ytj}$  in (10.1.40), where  $\tilde{\rho}_s - 1$  is given in (10.1.38), and the limiting distributions of the weighted symmetric test statistics are obtained by replacing  $Y_t$  with  $Y_t - \hat{f}_{ytj}$  in (10.1.44), where  $\tilde{\rho}_w - 1$  is given in

(10.1.43). We have

$$\begin{aligned} n^{-1}[(Y_1 - \bar{y}_n)^2, (Y_1 - \hat{f}_{y,2})^2] &\xrightarrow{\mathcal{L}} [H^2, (H - 3K)^2], \\ \left[ \bar{e}_n \bar{y}_n, n^{-1} \sum_{t=2}^n \hat{f}_{y,t-1,2} \hat{f}_{y,t2} \right] &\xrightarrow{\mathcal{L}} [HT, 0.5((T-2H)(T-6K)-1)], \\ n^{-2} \left[ \sum_{t=2}^n (Y_t - \bar{y}_n)^2, \sum_{t=1}^n (Y_t - \hat{f}_{y,2})^2 \right] &\xrightarrow{\mathcal{L}} [G - H^2, G - H^2 - 3K^2]. \end{aligned}$$

The limiting distributions of the test statistics follow from the joint limiting distribution of the components. ▲

The distributions of the pivotal statistics for the mean adjusted case are given in the second parts of Tables 10.A.3 and 10.A.4. The distributions for the linear trend adjusted case are given in the third parts of the tables. In addition, the distribution for the simple symmetric pivotal is given for the quadratic trend adjusted case.

The distributions of the symmetric estimators generalize to higher order processes in the same manner as the ordinary least squares estimators. Table 10.1.1 contains a data arrangement that can be used for the weighted regression estimation of the autoregressive model written as

$$Y_t = \theta_1 Y_{t-1} + \theta_2 Z_{t-1} + \theta_3 Z_{t-2} + \cdots + \theta_p Z_{t-p+1} + e_t, \quad (10.1.46)$$

where  $Z_t = Y_t - Y_{t-1}$ . If the model contains a mean function, the procedure is adjusted as described in Section 8.2.2. The hypothesis of a unit root is tested by testing the hypothesis that  $\theta_1 = 1$ .

**Table 10.1.1. Data Arrangement for Regression Estimation of Autoregressive Parameters by the Weighted Symmetric Procedure**

Weight	Dependent Variable	Parameter				
		$\theta_1$	$\theta_2$	...	$\theta_p$	
$w_{p+1}$	$Y_{p+1}$	$Y_p$	$Y_p - Y_{p-1}$	...	$Y_2 - Y_1$	
$w_{p+2}$	$Y_{p+2}$	$Y_{p+1}$	$Y_{p+1} - Y_p$	...	$Y_3 - Y_2$	
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$w_{n-p}$	$Y_n$	$Y_{n-1}$	$Y_{n-1} - Y_{n-2}$	...	$Y_{n-p+1} - Y_{n-p}$	
$1 - w_{n-p-1}$	$Y_{n-p}$	$Y_{n-p+1}$	$Y_{n-p+1} - Y_{n-p+2}$	...	$Y_{n-1} - Y_n$	
$1 - w_{n-p}$	$Y_{n-p-1}$	$Y_{n-p}$	$Y_{n-p} - Y_{n-p+1}$	...	$Y_{n-2} - Y_{n-1}$	
⋮	⋮	⋮	⋮	⋮	⋮	⋮
$1 - w_2$	$Y_1$	$Y_2$	$Y_2 - Y_3$	...	$Y_p - Y_{p+1}$	

**Theorem 10.1.9.** Let the assumptions of Theorem 10.1.2 hold. Let  $w_{ts} = 0.5$ , and let  $w_{tw}$  be defined in (8.2.15). Let  $\hat{\theta}_{gj}$  be the value of  $\theta$  that minimizes

$$\begin{aligned} & \sum_{t=p+1}^n w_{tg} (y_{ij} - \theta_1 y_{t-1,j} - \theta_2 z_{t-1,j} - \cdots - \theta_p z_{t-p+1,j})^2 \\ & + \sum_{t=1}^{n-p} (1 - w_{t+1,g}) (y_{ij} - \theta_1 y_{t+1,j} + \theta_2 z_{t+2,j} + \cdots + \theta_p z_{t+p,j})^2, \end{aligned} \quad (10.1.47)$$

where  $y_{ij} = Y_t - \hat{f}_{yij}$ ,  $z_{ij} = y_{ij} - y_{t-1,j}$ ,  $\hat{f}_{y11} = \bar{y}$ , and  $\hat{f}_{y12}$  is the least squares estimated linear trend function. Let  $\hat{\tau}_{gj}$ ,  $g = s, w$ ,  $j = 1, 2$ , be the pivotal for  $\theta_1 = 1$  associated with (10.1.47). Then the limiting distributions of the  $\hat{\tau}_{gj}$  are those given in Theorem 10.1.8.  $\blacktriangle$

**Proof.** Omitted.  $\blacktriangle$

The estimator of  $\rho$  for the autoregressive process (10.1.35) constructed by maximizing the normal stationary likelihood has a different limiting distribution than either that of ordinary least squares or those of the symmetric estimators when  $\rho = 1$ . The limiting distribution of that estimator, which we call the maximum likelihood estimator, has been given by Gonzalez-Farias (1992).

**Theorem 10.1.10.** Let the assumptions of Theorem 10.1.7 hold. Let  $\hat{\rho}_m$  be the value of  $\rho$  that maximizes

$$\begin{aligned} L(Y; \rho, \sigma^2) = & -0.5n \log 2\pi - 0.5n \log \sigma^2 + 0.5 \log(1 - \rho^2) \\ & - (2\sigma^2)^{-1} \left[ (1 - \rho^2) Y_1^2 + \sum_{t=2}^n (Y_t - \rho Y_{t-1})^2 \right]. \end{aligned} \quad (10.1.48)$$

Then

$$n(\hat{\rho}_m - 1) \xrightarrow{\mathcal{L}} 0.25G^{-1}\{(T^2 - 1) - [(T^2 - 1)^2 + 8G]^{1/2}\},$$

where  $G$  and  $T$  are defined in Theorem 10.1.1.

**Proof.** From (8.1.10), the cubic equation defining  $(\rho - 1) = \delta$  can be written

$$f_n(\delta) = \delta^3 + (3 + c_1)\delta^2 + (3 + 2c_1 + c_2)\delta + (1 + c_1 + c_2 + c_3) = 0, \quad (10.1.49)$$

where  $c_1 = -\bar{\rho}$ ,  $c_3 = (n - 2)^{-1}n\bar{\rho}$ ,

$$\begin{aligned} \bar{\rho} &= \left[ (n - 2)^{-1}(n - 1) \sum_{t=2}^{n-1} Y_t^2 \right]^{-1} \sum_{t=2}^n Y_{t-1} Y_t, \\ c_2 &= -(n - 1)^{-1}(n + 1) - \hat{\eta}, \end{aligned}$$

and

$$\hat{\eta} = (n - 1)^{-1} \left( \sum_{t=2}^{n-1} Y_{t-1}^2 \right)^{-1} (Y_1^2 + Y_n^2).$$

We observe that

$$\begin{aligned} 3 + c_1 &= 3 - \tilde{\rho} = 2 + O_p(n^{-1}), \\ 3 + 2c_1 + c_2 &= -2(\tilde{\rho} - 1 + n^{-1}) + O_p(n^{-2}), \\ 1 + c_1 + c_2 + c_3 &= 2n^{-1}(\tilde{\rho} - 1 + n^{-1}) - \hat{\eta} + O_p(n^{-3}). \end{aligned}$$

It follows that the roots of (10.1.49) converge in probability to 0, 0, and  $-2$ . Hence, the root in  $(-2, 0)$  is converging in probability to zero. Also,

$$\begin{aligned} n^2 \hat{\eta} &\xrightarrow{\mathcal{L}} G^{-1} T^2, \\ n(\tilde{\rho} - 1) &\xrightarrow{\mathcal{L}} G^{-1} [0.5(T^2 - 1) - G], \\ n^2(1 + c_1 + c_2 + c_3) &\xrightarrow{\mathcal{L}} -G^{-1}. \end{aligned}$$

Let

$$g_n(\delta) = a_{0n}\delta^2 + a_{1n}\delta + a_{2n},$$

where  $a_{0n} = 3 + c_1$ ,  $a_{1n} = 3 + 2c_1 + c_2$ , and  $a_{2n} = 1 + c_1 + c_2 + c_3$ . Because  $n^2 a_{2n} \xrightarrow{\mathcal{L}} -G^{-1}$ , the roots of  $g_n(\delta) = 0$  are real and of opposite sign with probability approaching one as  $n$  increases. Letting  $\tilde{\delta}_2$  denote the negative root of  $g_n(\delta) = 0$ , we have

$$\begin{aligned} n\tilde{\delta}_2 &= n(2a_{0n})^{-1}[-a_{1n} - (a_{1n}^2 - 4a_{0n}a_{2n})^{1/2}] \\ &= 0.5n\{\tilde{\rho} - 1 + n^{-1} - [(\tilde{\rho} - 1 - n^{-1})^2 + 2\hat{\eta} - 4n^{-2}]^{1/2}\} + O_p(n^{-1}) \\ &\xrightarrow{\mathcal{L}} 0.25G^{-1}\{(T^2 - 1) - [(T^2 - 1)^2 + 8G]^{1/2}\}. \end{aligned} \quad (10.1.50)$$

Also,

$$\begin{aligned} 0 &= f_n(\hat{\delta}_m) = f_n(\tilde{\delta}_2) + f'_n(\tilde{\delta}_2)(\hat{\delta}_m - \tilde{\delta}_2) + f''_n(\delta^*)(\hat{\delta}_m - \tilde{\delta}_2)^2 \\ &= \tilde{\delta}_2^3 + (3\tilde{\delta}_2^2 + 2a_{0,n}\tilde{\delta}_2 + a_{1,n})(\hat{\delta}_m - \tilde{\delta}_2) + (6\delta^* + 2a_{0,n})(\hat{\delta}_m - \tilde{\delta}_2)^2, \end{aligned}$$

where  $\hat{\delta}_m$  is the unique root of  $f_n(\delta) = 0$  in  $(-2, 0)$ ,  $f'_n(\cdot)$  is the derivative of  $f_n(\cdot)$ ,  $f''_n(\cdot)$  is the second derivative of  $f_n(\cdot)$ , and  $\delta^*$  is between  $\hat{\delta}_m$  and  $\tilde{\delta}_2$ . Because

$$n^2 f_n(\tilde{\delta}_2) = n^2 \tilde{\delta}_2^3 < 0$$

and

$$n^2 f_n(0) = n^2 a_{2,n} < 0$$

for  $n$  large,  $\hat{\delta}_m < \hat{\delta}_2$  for  $n$  large. It follows that

$$n(\hat{\delta}_m - \hat{\delta}_2) = \frac{-nb + [n^2 b^2 - 4n^2 \hat{\delta}_2^3 (2a_{0,n} + 6\delta^*)]^{1/2}}{2(2a_{0,n} + 6\delta^*)} \xrightarrow{\mathcal{L}} 0,$$

where  $b = 2a_{0,n}\hat{\delta}_2 + a_{1,n} + 3\hat{\delta}_2^2$ . Therefore, the limiting distribution of  $n\hat{\delta}_m$  is the same as the limiting distribution of  $n\hat{\delta}_2$ .  $\blacktriangle$

From (10.1.50), we see that the value of  $\hat{\rho}_m - 1$  is close to the value of  $\tilde{\rho}_w - 1$  unless  $\tilde{\rho}_w - 1$  is positive, or unless  $\tilde{\rho}_w - 1$  is negative and  $2(n^2 \hat{\eta} - 2)$  is large relative to  $\tilde{\rho}_w - 1$ . The limiting distribution of  $2(n^2 \hat{\eta} - 2)$  is the distribution of  $2G^{-1}(T^2 - 2G)$ , where  $E\{T^2 - 2G\} = 0$ . Empirical studies show that the 5% and 10% points of the distribution of  $n(\hat{\rho}_m - 1)$  are very similar to those of the distribution of  $n(\tilde{\rho}_w - 1)$ .

If  $\mu$  is unknown and the estimator of  $(\mu, \rho, \sigma^2)$  is obtained by maximizing the likelihood (8.1.5), the limiting distribution of the estimator of  $\rho$  differs from that of Theorem 10.1.10. Let  $\tilde{\rho}_{m1}$  denote the estimator of  $\rho$  that maximizes the likelihood (8.1.5). If  $Y_t$  satisfies the assumptions of Theorem 10.1.7, Gonzalez-Farias (1992) has shown that

$$n(\tilde{\rho}_{m1} - 1) \xrightarrow{\mathcal{L}} \zeta,$$

where  $\zeta$  is the negative root of

$$a_4 \zeta^4 + a_3 \zeta^3 + a_2 \zeta^2 + a_1 \zeta + a_0 = 0,$$

$$[a_0, a_3, a_4] = [-4, -T^2 + 1 - 2TH + H^2 - 8(G - H^2), 2(G - H^2)],$$

$$a_1 = -4T^2 + 8 + 8TH - 8H^2 + 2(T - 2H)^2,$$

$$a_2 = 8(G - H^2) + 4T - 5 - 8TH + H^2,$$

and  $G$ ,  $H$ , and  $T$  are defined in Theorem 10.1.6. Gonzalez-Farias (1992) also showed empirically that the limiting distribution of the estimator of  $\rho$  that maximizes (8.1.5) is nearly indistinguishable from the limiting distribution of the estimator that maximizes the known mean likelihood (10.1.48) where  $\mu$  is replaced with  $\bar{y}$ .

**Corollary 10.1.10.** Let the assumptions of Theorem 10.1.7 hold. Let  $\hat{\rho}_{m1}$  be the value of  $\rho$  that maximizes (10.1.48) with  $y_{t1} = Y_t - \bar{y} = Y_t - \hat{f}_{yt1}$  replacing  $Y_t$ , and let  $\hat{\rho}_{m2}$  be the value of  $\rho$  that maximizes (10.1.48) with  $y_{t2} = Y_t - \hat{f}_{yt2}$  replacing  $Y_t$ , where  $\hat{f}_{yt2}$  is the ordinary least squares estimator of the linear trend function. Then

$$n(\hat{\rho}_{mj} - 1) \xrightarrow{d} 0.5\{\zeta_j + 1 - [(\zeta_j - 1)^2 + 2\eta_j - 4]^{1/2}\},$$

where  $\zeta_j$  is the limiting distribution of  $n(\tilde{\rho}_{wj} - 1)$  obtained by multiplying the distribution of  $\hat{\tau}_{wj}$  of Theorem 10.1.8 by  $(G - H^2)^{-1/2}$  for  $j = 1$  and by  $(G - H^2 - 3K^2)^{-1/2}$  for  $j = 2$ ,

$$\eta_1 = (G - H^2)^{-1}[H^2 + (T - H)^2],$$

and

$$\eta_2 = (G - H^2 - 3K^2)^{-1}[(H - 3K)^2 + (T - H - 3K)^2].$$

**Proof.** The results follow from (10.1.50).  $\blacktriangle$

A pivotal statistic for the test of  $\rho = 1$  can be constructed by dividing  $\hat{\rho}_{mj} - 1$  by the estimated standard error of  $\hat{\rho}_{mj}$ . For higher order processes, the test is constructed by testing the hypothesis that the sum of the autoregressive coefficients is one.

There are several methods available for computing the estimated covariance matrix of the estimated coefficients. One procedure is to write the concentrated log likelihood of (8.4.5) in the form

$$\sum_{i=1}^n g_i^2(\mathbf{Y}; \boldsymbol{\theta}), \quad (10.1.51)$$

where  $g_i(\mathbf{Y}; \boldsymbol{\theta}) = c(\boldsymbol{\theta})z_i(\mathbf{Y}; \boldsymbol{\theta})$ ,  $z_i(\mathbf{Y}; \boldsymbol{\theta})$  are uncorrelated  $(0, \sigma^2)$  random variables when the function is evaluated at the true  $\boldsymbol{\theta}$ , and  $c(\boldsymbol{\theta})$  is a constant depending on  $\boldsymbol{\theta}$ . Then the maximum likelihood estimators are obtained by minimizing (10.1.51), and an estimator of the covariance matrix of the maximum likelihood estimator  $\hat{\boldsymbol{\theta}}_m$  is

$$\hat{\mathbf{V}}\{\hat{\boldsymbol{\theta}}_m\} = \left[ \sum_{i=1}^n \mathbf{h}'_i(\mathbf{Y}; \hat{\boldsymbol{\theta}}_m) \mathbf{h}_i(\mathbf{Y}; \hat{\boldsymbol{\theta}}_m) \right]^{-1} \hat{\sigma}_m^2, \quad (10.1.52)$$

where  $\mathbf{h}'_i(\mathbf{Y}; \hat{\boldsymbol{\theta}}_m)$  is the vector of derivatives of  $g_i(\mathbf{Y}; \boldsymbol{\theta})$  with respect to  $\boldsymbol{\theta}$  evaluated at the maximum likelihood estimator,

$$\hat{\sigma}_m^2 = (n - r)^{-1} \sum_{i=1}^n z_i^2(\mathbf{Y}; \boldsymbol{\theta}) \quad (10.1.53)$$

is the maximum likelihood estimator of  $\sigma^2$  adjusted for degrees of freedom, and  $r$  is the number of parameters estimated.

Table 10.A.5 contains the percentiles of the test statistics based on the maximum likelihood estimator and the estimated covariance matrix (10.1.52). This table was constructed by the same procedures used to construct Table 10.A.4.

We have presented a number of test statistics for the unit root problem. The ordinary least squares estimators are easy to compute and can be computed with an

ordinary regression program. They may be used in exploratory procedures even when other programs are available. The simple symmetric estimator is next in ease of computation and can also be computed with an ordinary least squares regression program.

If the objective is to test the hypothesis of a unit root against the alternative of a stationary process with unknown mean, the maximum likelihood and weighted symmetric procedures are recommended. Monte Carlo studies have demonstrated that these two procedures have nearly identical power against the stationary unknown mean alternative and that they are much more powerful than ordinary least squares in that case. The power of the test based on the simple symmetric estimator falls between that of ordinary least squares and that of maximum likelihood, being closer to that of maximum likelihood. See for example, Pantula, Gonzalez-Farias, and Fuller (1994) and Park and Fuller (1995).

If a test for a unit root against the alternative of a model with known mean is desired, the ordinary least squares procedure is recommended. See Park and Fuller (1995).

**Example 10.1.2.** We continue the analysis of the one-year treasury bill interest rate begun in Example 10.1.1. The simple symmetric estimator of the third order process is

$$\hat{\Delta Y}_t = 0.106 - 0.020 Y_{t-1} + 0.354 \Delta Y_{t-1} - 0.089 \Delta Y_{t-2}. \\ (0.062) \quad (0.011) \quad (0.065) \quad (0.066)$$

The regression equation was computed using an ordinary regression program and the data arrangement of Table 10.1.1 with an intercept included in the regression. The estimated error variance is 0.083. The value of the statistic for a test of a unit root is  $-1.76$ , which is greater than the 0.10 tabular value of  $-2.34$  given in Table 10.A.3.

If we use the weights of (8.2.15), we obtain

$$\hat{\Delta Y}_t = 0.090 - 0.015 Y_{t-1} + 0.360 \Delta Y_{t-1} - 0.092 \Delta Y_{t-2} \\ (0.062) \quad (0.011) \quad (0.065) \quad (0.066)$$

as the weighted symmetric least squares estimated equation. The test statistic for a unit root is  $-1.36$ , which is greater than the 0.10 tabular value of  $-2.23$  given in Table 10.A.4.

The maximum likelihood estimates for the third order process are

$$(-\hat{\alpha}_1, -\hat{\alpha}_2, -\hat{\alpha}_3) = (1.3436, -0.4510, 0.0915), \\ (0.0650) \quad (0.1053) \quad (0.0656)$$

where we used procedure ARIMA of SAS/ETS<sup>\*</sup> to compute the maximum likelihood estimator. This program uses the Marquardt algorithm and equation (10.1.52) to compute the estimated covariance matrix of the coefficients. The sum

of the estimated autoregressive coefficients is 0.9841, and the estimated standard error of the sum of the coefficients is 0.0119. Therefore, the maximum likelihood pivotal for testing the hypothesis of a unit root is  $-1.34$ . The computed value is greater than the 0.10 tabular value of  $-2.32$  given in the second part of Table 10.A.5.

In this example, all procedures lead to essentially the same conclusions. As stated in the text, the weighted symmetric and maximum likelihood procedures are the recommended procedures when the mean is estimated. In this example, and in the majority of practical situations, the weighted symmetric estimator and the maximum likelihood estimator will both reject the hypothesis of a unit root or both accept the hypothesis when the test is against the stationary alternative.  $\blacktriangle \blacktriangle$

From the tabled distribution of  $n(\hat{\rho} - 1)$  of Table 10.A.1, it is clear that the least squares estimator of the unit root parameter of an autoregressive process is biased. Also, the estimator of the autoregressive parameter is biased in the stationary case. See Section 8.2. In the unit root case, the bias is order  $n^{-1}$ , the same order as the standard deviation of the estimator.

There are several possible situations associated with the use of data to investigate an autoregressive process. One may be willing, on the basis of subject matter knowledge, to specify that the process is stationary. Then the use of an estimation procedure, such as maximum likelihood based on the stationary likelihood, that restricts all roots of the process to be less than one in absolute value, is appropriate. In a second situation, one may have strong theoretical reasons for believing a unit root is present. If so, one may choose to test for a unit root and, if the test is consistent with the hypothesis of a unit root, estimate the process subject to the restriction of a unit root. In the third situation, a unit root is a distinct possibility, but not to the extent that it is a model to be maintained unless rejected by the data. In this third case, it is natural to adopt an estimator that is less biased in the vicinity of the unit root than are the ordinary estimators.

Assume that the model is a  $p$ th order autoregressive process. Assume we wish to restrict the estimator of the largest root to the interval  $(-1, 1]$ . Under the assumption that only a single root of one is possible and that there is no drift if a root is equal to one, an estimation procedure is the following.

### 1. Estimate the model

$$Y_t = \theta_0 + \theta_1 Y_{t-1} + \sum_{j=2}^p \theta_j \Delta Y_{t-j+1} + e_t \quad (10.1.54)$$

by weighted symmetric least squares, and compute the statistic  $\hat{\tau}_{w1}$  for the test that  $\theta_1 = 1$ . This can be done by fitting the weighted least squares regression of  $y_t$  on  $(y_{t-1}, \Delta Y_{t-1}, \dots, \Delta Y_{t-p})$ , where  $y_t = Y_t - \bar{y}_n$ .

### 2. Construct a new estimator of $\theta_1$

$$\tilde{\theta}_1 = \hat{\theta}_1 + c(\hat{\tau}_{w1})[\hat{V}\{\hat{\theta}_1\}]^{0.5}, \quad (10.1.55)$$

where  $\hat{\theta}_1$  is the estimator from step 1,

$$c(\hat{\tau}_{w1}) = \begin{cases} -\hat{\tau}_{w1} & \text{if } \hat{\tau}_{w1} \geq -1.2, \\ 0.035672(\hat{\tau}_{w1} + 7.0)^2 & \text{if } -7.00 < \hat{\tau}_{w1} \leq -1.2, \\ 0 & \text{if } \hat{\tau}_{w1} \leq -7.00, \end{cases}$$

and  $\hat{V}\{\hat{\theta}_1\}$  is the estimated variance of the weighted least squares estimator.

3. Estimate  $(\theta_2, \dots, \theta_p)$  by the regression of  $y_t - \hat{\theta}_1 y_{t-1}$  on  $\Delta Y_{t-j}$ ,  $j = 1, 2, \dots, p-1$ . Denote the estimators by  $(\tilde{\theta}_2, \dots, \tilde{\theta}_p)$ .

In the presence of a unit root ( $\theta_1 = 1$ ), the outlined estimation procedure is approximately median unbiased for  $\theta_1$  because the median of  $\hat{\tau}_{w1}$  is about -1.20 when  $\theta_1 = 1$ . Of course, if  $\theta_1$  is restricted to the interval  $(-1, 1]$ , it is not possible to construct a reasonable estimator of  $\theta_1$  that is mean unbiased for all  $\theta_1 \in (-1, 1]$ . The function  $c(\hat{\tau}_{w1})$  of (10.1.55) was chosen to be a smooth function of  $\hat{\tau}_{w1}$  with value of 1.20 at  $\hat{\tau}_{w1} = -1.20$ . The modified estimator differs from the weighted least squares estimator if  $\hat{\tau}_{w1} > -7.00$ . This set of  $\hat{\tau}_{w1}$  corresponds, approximately, to  $\hat{\theta}_1 > (n+49)^{-1}(n-49)$ . The weighted least squares estimator  $\hat{\theta}_1$  is biased for most values of  $\theta_1$ , but the bias is small relative to the standard error for  $\theta_1$  considerably less than one and  $n$  large.

The empirical properties of the weighted symmetric estimator and the estimator (10.1.55) are compared for the first order process in Table 10.1.2. The true model is

$$Y_t = \rho Y_{t-1} + e_t,$$

where  $e_t \sim N(0, 1)$ . The generated process is stationary if  $\rho < 1$ , and  $Y_0 = 0$  if  $\rho = 1$ . The weighted least squares estimator, denoted by  $\hat{\rho}_w$ , is defined by (10.1.42) with  $Y_t$  replaced by  $Y_t - \bar{y}$ , provided the estimator is less than one, and is equal to one otherwise. The entries in the table are based on 3000 samples.

The estimator  $\tilde{\rho}$  defined by (10.1.55) has a uniformly smaller median bias than the weighted least squares estimator  $\hat{\rho}_w$ . The mean bias is also uniformly smaller. However, the empirical variance of  $\hat{\rho}_w$  is smaller than that of  $\tilde{\rho}$  for  $\rho$  differing from one by more than  $8n^{-1}$ , and the mean square error of  $\hat{\rho}_w$  is smaller than that of  $\tilde{\rho}$  for  $\rho$  differing from one by more than four standard errors of the estimator. In those situations, the mean square error of  $\tilde{\rho}$  exceeds that of  $\hat{\rho}_w$  by modest percentages. The largest percentage difference in the table is about 8%. On the other hand, for all three values of  $n$ , the modified estimator has a mean square error less than one half that of  $\hat{\rho}_w$  for  $\rho = 1$ . For a fixed  $\rho$ , such as 0.75, the mean square error multiplied by  $n$  declines as  $n$  increases for both estimators, primarily because of the decrease in bias.

The last two columns of Table 10.1.2 contain the empirical 2.5 percentiles of the statistics

$$\tilde{t}_1 = [\hat{V}(\hat{\rho}_w)]^{-1/2}(\tilde{\rho} - \rho)$$

**Table 10.1.2. Empirical Properties of Estimator (10.1.55) and Weighted Least Squares Estimator for First Order Process**

$\rho$	Median		$nMSE$		Percentiles	
	$\hat{\rho}_w$	$\tilde{\rho}$	$\hat{\rho}_w$	$\tilde{\rho}$	$\hat{t}_{1,025}$	$\tilde{t}_{1,025}$
$n = 50$						
0.15	0.122	0.126	1.046	1.127	-2.25	-2.25
0.50	0.456	0.491	0.998	1.064	-2.36	-2.31
0.75	0.695	0.757	0.857	0.748	-2.52	-2.28
0.85	0.791	0.860	0.782	0.567	-2.58	-2.22
0.90	0.838	0.909	0.747	0.467	-2.64	-2.20
0.95	0.884	0.955	0.720	0.357	-2.71	-2.19
0.98	0.912	0.981	0.710	0.309	-2.82	-2.25
1.00	0.933	1.000	0.728	0.311	-2.83	-2.22
$n = 100$						
0.50	0.479	0.483	0.836	0.887	-2.13	-2.13
0.75	0.724	0.747	0.630	0.624	-2.27	-2.16
0.85	0.822	0.855	0.521	0.452	-2.38	-2.13
0.90	0.870	0.905	0.459	0.352	-2.45	-2.11
0.95	0.919	0.956	0.395	0.241	-2.57	-2.10
0.98	0.948	0.984	0.363	0.170	-2.71	-2.15
0.99	0.957	0.992	0.360	0.156	-2.75	-2.16
1.00	0.965	1.000	0.379	0.159	-2.80	-2.17
$n = 200$						
0.75	0.738	0.742	0.562	0.590	-2.29	-2.28
0.85	0.836	0.847	0.421	0.421	-2.33	-2.25
0.90	0.886	0.901	0.342	0.313	-2.40	-2.20
0.95	0.935	0.953	0.262	0.199	-2.52	-2.18
0.98	0.964	0.983	0.216	0.125	-2.69	-2.21
0.99	0.974	0.992	0.204	0.101	-2.74	-2.19
1.00	0.983	1.000	0.203	0.091	-2.80	-2.17

and

$$\hat{t}_1 = [\hat{V}(\hat{\rho}_w)]^{-1/2}(\hat{\rho}_w - \rho),$$

where  $\hat{\rho}_w$  is the weighted symmetric estimator and  $\tilde{\rho}$  is the estimator (10.1.55). Because  $\tilde{\rho}$  has small median bias, the pivotal statistic  $\hat{t}_1$  has a median close to zero for all values of  $\rho$ .

Because the 2.5 percentile for  $\hat{t}_{w1}$  is -2.80 for  $n = 200$  and  $\rho = 1$ , the 2.5

percentile of  $\tilde{t}_1$  is  $-2.17$  for  $n = 200$  and  $\rho = 1$ . The estimator  $\hat{\rho}_w$  has a negative bias for positive  $\rho$ , and the 2.5 percentile is less than  $-2.0$  for  $\rho$ 's that are a considerable distance from one. The percentiles for  $\hat{t}_1$  show a steady decrease to about  $-2.8$  at  $\rho = 1.00$ .

The empirical 97.5 percentiles of  $\tilde{t}_1$  and  $\hat{t}_1$  are compared in Table 10.1.3. The percentiles of  $\hat{t}_1$  decline as  $\rho$  moves towards one. The maximum value of the 97.5 percentile for  $\tilde{t}_1$  is approximately equal to  $2.30$  for  $1 - \rho = 12n^{-1}$ . At least 2.5% of the estimates  $\tilde{\rho}$  are equal to one for all three sample sizes when  $1 - \rho = 6n^{-1}$ . The majority of the 97.5 percentiles of  $\tilde{t}_1$  fall below  $2.30$ , and the majority of the 2.5 percentiles are above  $-2.30$ . Therefore, the interval

$$[\tilde{\rho} - 2.30(\hat{V}\{\hat{\rho}_w\})^{1/2}, \tilde{\rho} + 2.30(\hat{V}\{\hat{\rho}_w\})^{1/2}] \quad (10.1.56)$$

will have coverage greater than 95% for most parameter values and for sample sizes in the 50 to 200 range. Because  $\rho$  is restricted to  $(-1, 1]$ , the actual confidence interval is the intersection of (10.1.56) and  $(-1, 1]$ .

**Example 10.1.3.** We continue the analysis of Example 10.1.2 under the assumption that the largest root of the autoregressive process is in the interval  $(-1, 1]$ . The statistic  $\hat{\tau}_{w1}$  from Example 10.1.2 is  $-1.36$ . The modified estimator of (10.1.55) is

$$\hat{\theta}_1 = \hat{\theta}_{w1} + 0.013 = 0.997.$$

Regressing  $y_t - 0.997y_{t-1}$  on  $\Delta Y_{t-1}$  and  $\Delta Y_{t-2}$ , we obtain the estimated equation,

$$\hat{y}_t = 0.997 y_{t-1} + 0.355 \Delta Y_{t-1} - 0.103 \Delta Y_{t-2}, \\ (0.013) \quad (0.065) \quad (0.066)$$

where  $y_t = Y_t - \bar{y}_n$ . The standard errors for the coefficients of  $\Delta Y_{t-1}$  and  $\Delta Y_{t-2}$  are those computed for the original weighted regression of Example 10.1.2. The standard error for the coefficient of  $Y_{t-1}$  is the standard error from Example 10.1.2 increased by 15%. With these standard errors, normal tables can be used for approximate tests and confidence intervals for the parameters of the equation. An

Table 10.1.3. Empirical 97.5 Percentiles of  $\tilde{t}$  and  $\hat{t}$

$1 - \rho$	$n = 50$		$n = 100$		$n = 200$	
	$\hat{t}_{1,975}$	$\tilde{t}_{1,975}$	$\hat{t}_{1,975}$	$\tilde{t}_{1,975}$	$\hat{t}_{1,975}$	$\tilde{t}_{1,975}$
$25n^{-1}$	1.49	2.02	1.58	2.17	1.55	2.17
$12n^{-1}$	1.26	2.27	1.34	2.29	1.36	2.32
$6n^{-1}$	1.04	2.04 <sup>†</sup>	1.11	2.08 <sup>†</sup>	1.05	2.04 <sup>†</sup>

<sup>†</sup> More than 2.5% of the estimates equal one.

approximate 95% confidence interval for  $\theta_1$  under the assumption that the largest root is in  $(-1, 1]$  is  $[0.971, 1]$ .

The characteristic equation associated with the modified estimates is

$$m^3 - 1.352m^2 + 0.458m - 0.103 = 0,$$

and the largest root is 0.996. Using a Taylor expansion and the estimated covariance matrix with the variance of the coefficient of  $y_{t-1}$  increased by 15%, the standard error of the largest root is 0.013. An approximate 95% confidence interval for the largest root is  $[0.970, 1]$ .  $\blacktriangle \blacktriangle$

#### 10.1.4. Prediction for Unit Root Autoregressions

It is an important result that prediction procedures appropriate for stationary time series are also appropriate for time series with an autoregressive unit root. As in the stationary case, estimation of the parameters adds a term that is  $O_p(n^{-1/2})$  to the prediction error when the autoregression contains a unit root.

**Theorem 10.1.11.** Let  $Y_t$  satisfy

$$\begin{aligned} Y_t &= \alpha_0 - \sum_{j=1}^p \alpha_j Y_{t-j} + e_t \\ &= \alpha_0 + \theta_1 Y_{t-1} + \sum_{j=2}^p \theta_j (Y_{t-j+1} - Y_{t-j}) + e_t, \end{aligned}$$

where  $\theta_1 = -\sum_{j=1}^p \alpha_j$ ,  $\theta_j = \sum_{i=j}^p \alpha_i$ ,  $j = 2, 3, \dots, p$ , and  $e_t$  are iid( $0, \sigma^2$ ) random variables. Let  $Y_1, Y_2, \dots, Y_p$  be fixed, and let  $\hat{\alpha}$  and  $\hat{\theta}$  be the least squares estimators of  $\alpha = (\alpha_0, \alpha_1, \dots, \alpha_p)'$  and  $\theta = (\theta_1, \theta_2, \dots, \theta_p)'$ . Assume that one of the roots of the characteristic equation is one and that the other roots are less than one in absolute value. Let  $\hat{Y}_{n+s}$  be the predictor of  $Y_{n+s}$  based upon  $\hat{\alpha}$ . Then

$$Y_{n+s} - \hat{Y}_{n+s} = \sum_{j=0}^{s-1} w_j e_{n+s-j} + O_p(n^{-1/2}),$$

where the  $w_j$  satisfy the difference equation

$$w_j + \sum_{i=1}^p \alpha_i w_{j-i} = 0$$

subject to the initial conditions  $w_0 = 1$  and

$$w_j + \sum_{i=1}^j \alpha_i w_{j-i} = 0, \quad j = 1, 2, \dots, p-1.$$

**Proof.** By Theorems 10.1.4 and 10.1.5,  $\hat{\alpha} - \alpha = O_p(n^{-1/2})$  and

$$\hat{\theta}_1 - \theta_1 = \begin{cases} O_p(n^{-1}) & \text{if } \alpha_0 = 0, \\ O_p(n^{-3/2}) & \text{if } \alpha_0 \neq 0. \end{cases}$$

It follows that

$$\begin{aligned} Y_{n+1} - \hat{Y}_{n+1} &= \alpha_0 - \hat{\alpha}_0 + (\theta_1 - \hat{\theta}_1)Y_n + \sum_{j=2}^p (\theta_j - \hat{\theta}_j)\Delta Y_{n-j+2} + e_{n+1} \\ &= e_{n+1} + O_p(n^{-1/2}) \end{aligned}$$

because  $\Delta Y_t = O_p(1)$ ,  $Y_n = O_p(n^{1/2})$  if  $\alpha_0 = 0$ , and  $Y_n = O_p(n)$  if  $\alpha_0 \neq 0$ . Repeated substitution into the expression

$$\hat{Y}_{n+s} = \hat{\alpha}_0 + \hat{\theta}_1 \hat{Y}_{n+s-1} + \sum_{j=2}^p \hat{\theta}_j (\hat{Y}_{n-j+s+1} - \hat{Y}_{n-j+s})$$

yields the result for  $s > 1$ . ▲

It follows from Theorem 10.1.11 that the estimator of the prediction variance given in (8.5.14) for the stationary process can also be used for a process with a unit root.

**Example 10.1.4.** By Theorem 8.5.1 and Theorem 10.1.11, the prediction variance formula that ignores the effect of estimation error can be used for models whose autoregressive part has roots less than or equal to one in absolute value. In Examples 10.1.1 and 10.1.2, we found the behavior of the one-year treasury bill interest rate to be consistent with the hypothesis of a unit root. Least squares or maximum likelihood can be used to construct predictors that are appropriate whether or not the process has a unit root.

The one-, two-, and three-period forecasts computed with the third order process estimated by stationary maximum likelihood are 9.01, 8.95, and 8.87, respectively. The standard errors estimated using the approximation that ignores the estimation error are 0.291, 0.486, and 0.626 for the one-, two-, and three-period forecasts. The corresponding approximate standard errors computed by the least squares prediction formulas are 0.292, 0.493, and 0.641. The least squares standard errors were computed by fitting the model with three explanatory variables for the three predictions as described in Example 8.5.1. ▲▲

## 10.2. EXPLOSIVE AUTOREGRESSIVE TIME SERIES

We begin our study of explosive autoregressive time series with the first order process. Let

$$Y_t = \begin{cases} \theta_1 Y_{t-1} + e_t, & t = 1, 2, \dots, \\ 0, & t = 0, \end{cases} \quad (10.2.1)$$

where  $|\theta_1| > 1$  and  $\{e_t\}_{t=-\infty}^{\infty}$  is a sequence of i.i.d. random variables with mean zero and positive variance  $\sigma^2$ . Equation (10.2.1) is understood to mean that  $Y_t$  is created by adding  $e_t$  to  $\theta_1 Y_{t-1}$ . As with other autoregressive processes, repeated substitution enables us to express  $Y_t$  as a weighted average of the  $e_i$ ,

$$Y_t = \sum_{j=0}^{t-1} \theta_1^j e_{t-j}. \quad (10.2.2)$$

It follows that, for  $|\theta_1| > 1$ , the variance of  $Y_t$  is

$$V\{Y_t\} = (\theta_1^2 - 1)^{-1} (\theta_1^{2t} - 1) \sigma^2 \quad (10.2.3)$$

and the variance increases exponentially in  $t$ .

In expression (10.2.2), the term in  $e_1$  is  $\theta_1^{t-1} e_1$ . When  $|\theta_1| > 1$ , the weight  $\theta_1^{t-1}$  increases exponentially as  $t$  increases. Hence, the variance contribution from  $e_1$  is of order  $\theta_1^{2t}$ . If  $|\theta_1| < 1$ , the weight applied to the first observation declines to zero as  $t$  increases. If  $|\theta_1| = 1$ , the weight of the first observation becomes small relative to the standard deviation of the process.

To further understand the importance of the first few  $e_i$  in the model with  $|\theta_1| > 1$ , we observe that

$$\begin{aligned} Y_t &= \sum_{j=0}^{t-1} \theta_1^j e_{t-j} = \sum_{i=1}^t \theta_1^{t-i} e_i \\ &= \theta_1^t \sum_{i=1}^t \theta_1^{-i} e_i. \end{aligned} \quad (10.2.4)$$

Letting  $X_t = \sum_{i=1}^t \theta_1^{-i} e_i$ , we have

$$X_t \rightarrow X \quad \text{a.s.} \quad (10.2.5)$$

as  $t$  increases, where  $X = \sum_{i=1}^{\infty} \theta_1^{-i} e_i$ . Thus for large  $t$ , the behavior of  $Y_t$  is essentially that of an exponential multiple of the random variable  $X$ .

The model (10.2.1) extended to include an intercept term and a general initial value is

$$Y_t = \begin{cases} \theta_0 + \theta_1 Y_{t-1} + e_t, & t = 1, 2, \dots, \\ y_0, & t = 0. \end{cases} \quad (10.2.6)$$

The least squares estimators of  $\theta_0$  and  $\theta_1$  are

$$\begin{aligned} \hat{\theta}_1 &= \left[ \sum_{t=1}^n (Y_{t-1} - \bar{y}_{(-1)})^2 \right]^{-1} \sum_{t=1}^n (Y_{t-1} - \bar{y}_{(-1)})(Y_t - \bar{y}_{(0)}), \\ \hat{\theta}_0 &= \bar{y}_{(0)} - \hat{\theta}_1 \bar{y}_{(-1)}, \end{aligned} \quad (10.2.7)$$

where

$$\bar{y}_{(t)} = n^{-1} \sum_{i=1}^n Y_{t+i}, \quad i = 0, -1.$$

If  $y_0$  is fixed and  $e_i \sim NI(0, \sigma^2)$ , the least squares estimators are the maximum likelihood estimators.

Rubin (1950) showed that  $\text{plim } \hat{\theta}_1 = \theta_1$  for the least squares estimator of the model (10.2.1). White (1958, 1959) considered the asymptotic properties of the least squares statistics for (10.2.1) under the assumption that  $e_i \sim NI(0, \sigma^2)$ . White used moment generating function techniques to obtain his results. Anderson (1959) extended White's results using the representation (10.2.4). Hasza (1977) studied estimators for the general model (10.2.6), and Theorem 10.2.1 is taken from that work.

**Theorem 10.2.1.** Let  $\{Y_t\}_{t=0}^\infty$  be defined by (10.2.6) with  $|\theta_1| > 1$  and  $\{e_t\}$  a sequence of iid( $0, \sigma^2$ ) random variables. Let the least squares estimators be defined by (10.2.7). The limiting distribution of  $(\theta_1^2 - 1)^{-1} \theta_1^n (\hat{\theta}_1 - \theta_1)$  is that of  $W_1 W_2^{-1}$ , where  $W_1$  and  $W_2$  are independent random variables,

$$(W_1, W_2) = \left( \sum_{i=1}^{\infty} \theta_1^{-i} e_{-i}, \delta_0 + \sum_{j=1}^{\infty} \theta_1^{-j} e_j \right),$$

and  $\delta_0 = y_0 + \theta_0(\theta_1 - 1)^{-1}$ . The limiting distribution of  $n^{1/2}(\hat{\theta}_0 - \theta_0)$  is that of a  $N(0, \sigma^2)$  random variable, and  $n^{1/2}(\hat{\theta}_0 - \theta_0)$  is independent of  $\theta_1^n (\hat{\theta}_1 - \theta_1)$  in the limit. If  $e_t \sim NI(0, \sigma^2)$ ,  $y_0 = 0$ , and  $\theta_0 = 0$ , then  $(\theta_1^2 - 1)^{-1} \theta_1^n (\hat{\theta}_1 - \theta_1)$  converges in distribution to a Cauchy random variable.

**Proof.** We may write  $Y_t$  as

$$\begin{aligned} Y_t &= \theta_1^t \left( y_0 + \theta_0 \sum_{j=1}^t \theta_1^{-j} + \sum_{j=1}^t \theta_1^{-j} e_j \right) \\ &= \theta_1^t X_t, \quad t = 0, 1, \dots, \end{aligned} \tag{10.2.8}$$

where

$$(X_t, \delta_0) = \left( \delta_0 + \sum_{j=1}^t \theta_1^{-j} e_j, y_0 + \theta_0 \sum_{j=1}^t \theta_1^{-j} \right).$$

Now  $X_t$  converges a.s. as  $t \rightarrow \infty$  to a random variable

$$X = \delta_0 + \sum_{j=1}^{\infty} \theta_1^{-j} e_j,$$

where  $\delta_0 = y_0 + \theta_0(\theta_1 - 1)^{-1}$ . Then  $\theta_1^{-t} Y_t \rightarrow X$  a.s. as  $t$  increases. Note that

$$X - X_t = \sum_{j=t+1}^{\infty} (\theta_0 \theta_1^{-j} + \theta_1^{-j} e_j) = O_p(|\theta_1|^{-t}).$$

The error in the least squares estimator of  $\theta_1$  is

$$\hat{\theta}_1 - \theta_1 = \frac{\sum_{t=1}^n Y_{t-1} e_t - n \bar{y}_{(-1)} \bar{e}}{\sum_{t=1}^n Y_{t-1}^2 - n \bar{y}_{(-1)}^2}, \quad (10.2.9)$$

where

$$\begin{aligned} \theta_1^{-2n} \sum_{t=1}^n Y_{t-1}^2 &= \sum_{j=0}^{n-1} \theta_1^{-2j-2} X^2 + 2 \sum_{t=1}^n \theta_1^{2(t-n)-2} X (X_{t-1} - X) \\ &\quad + \sum_{t=1}^n \theta_1^{2(t-n)-2} (X_{t-1} - X)^2. \end{aligned} \quad (10.2.10)$$

Using

$$\begin{aligned} E\{(X_{t-1} - X)^2\} &= (1 - \theta_1^{-2})^{-1} \theta_1^{-2t} \sigma^2 + \theta_0^2 (\theta_1 - 1)^{-2} \theta_1^{-2(t-1)} = O(|\theta_1|^{-2t}), \\ \text{Var}\left\{\sum_{t=1}^n \theta_1^{2(t-n)-2} \sum_{j=t}^{\infty} \theta_1^{-j} e_j\right\} &= O(|\theta_1|^{-2n}), \end{aligned}$$

and

$$\sum_{t=1}^n \theta_1^{2(t-n)-2} (X_{t-1} - X)^2 = O_p(n|\theta_1|^{-2n}), \quad (10.2.11)$$

we have

$$\theta_1^{-2n} \sum_{t=1}^n Y_{t-1}^2 = (\theta_1^2 - 1)^{-1} X^2 + O_p(|\theta_1|^{-n}). \quad (10.2.12)$$

By similar arguments

$$\begin{aligned} \theta_1^{-n} \sum_{t=1}^n Y_{t-1} e_t &= X \sum_{t=1}^n \theta_1^{t-n-1} e_t + O_p(n|\theta_1|^{-n}), \\ \theta_1^{-n} \bar{y}_{(-1)} &= n^{-1} (\theta_1 - 1)^{-1} X + O_p(|\theta_1|^{-n}), \end{aligned} \quad (10.2.13)$$

and

$$[n^{1/2} \theta_1^{-n} \bar{y}_{(-1)} \bar{e}, n \theta_1^{-2n} \bar{y}_{(-1)}^2] = O_p(n^{-1}).$$

Therefore,

$$\theta_1^{-2n} \sum_{t=1}^n (Y_{t-1} - \bar{y}_{(-1)})^2 = (\theta_1^2 - 1)^{-1} X^2 + O_p(n^{-1})$$

and

$$\theta_1^{-n} \sum_{t=1}^n (Y_{t-1} - \bar{y}_{(-1)})(e_t - \bar{e}) = X \sum_{t=1}^n \theta_1^{t-n-1} e_t + O_p(n^{-1/2}).$$

It follows that

$$(\theta_1^2 - 1)^{-1} \theta_1^n (\hat{\theta}_1 - \theta_1) = X^{-1} Z_n + O_p(n^{-1/2}), \quad (10.2.14)$$

where  $Z_n = \sum_{t=1}^n \theta_1^{t-n-1} e_t = \sum_{j=1}^n \theta_1^{-j} e_{n-j+1}$ . Because

$$Z_n = \sum_{j < n/2} \theta_1^{-j} e_{n-j+1} + O_p(|\theta_1|^{-n})$$

and

$$X = \delta_0 + \sum_{j < n/2} \theta_1^{-j} e_j + O_p(|\theta_1|^{-n}),$$

$Z_n$  and  $X$  are asymptotically independent.

Therefore,  $(\theta_1^2 - 1)^{-1} \theta_1^n (\hat{\theta}_1 - \theta_1)$  has a nondegenerate limiting distribution. If the  $e_t$ 's are  $\text{NI}(0, \sigma^2)$  random variables, then  $Z_n$  and  $X$ , both being linear combinations of the  $e_t$ 's, are normally distributed. The mean of the  $Z_n$  is zero and the mean of  $X$  is  $\delta_0$ , so that the limiting distribution of  $(\theta_1^2 - 1)^{-1} \theta_1^n (\hat{\theta}_1 - \theta_1)$  is that of  $W_1 W_2^{-1}$ , where  $(W_1, W_2)$  is a normal vector. If  $\delta_0 = 0$  and the  $e_t$  are normal, then the limiting distribution of  $(\theta_1^2 - 1)^{-1} \theta_1^n (\hat{\theta}_1 - \theta_1)$  is that of the ratio of two independent standard normal variables, which is Cauchy.

Now, from (10.2.7),

$$\hat{\theta}_0 - \theta_0 = \bar{e}_{(0)} - (\hat{\theta}_1 - \theta_1) \bar{y}_{(-1)} = \bar{e}_{(0)} + O_p(n^{-1})$$

by (10.2.13) and (10.2.14). Also,  $n^{1/2} \bar{e}_{(0)}$  is independent of  $X$  and of  $Z_n$  in the limit, because the elements of

$$\mathbf{L}_n = \left( n^{-1/2} \sum_{t=1}^{n-k_n} e_t, \sum_{j=1}^{k_n} \theta_1^{-j} e_{n-j+1} \right)$$

are independent and the vector  $\mathbf{L}_n - (n^{1/2} \bar{e}_0, Z_n)$  converges in probability to zero for  $n^{0.1} < k_n < n^{0.2}$ , for example. ▲

The distributions of other regression statistics follow from Theorem 10.2.1. Let the regression residual mean square be

$$\hat{\sigma}^2 = (n-2)^{-1} \sum_{t=1}^n [Y_t - \bar{y}_{(0)} - \hat{\theta}_1(Y_{t-1} - \bar{y}_{(-1)})]^2. \quad (10.2.15)$$

**Corollary 10.2.1.1.** Suppose the assumptions of Theorem 10.2.1 hold. Then

$$\operatorname{plim}_{n \rightarrow \infty} \hat{\sigma}^2 = \sigma^2.$$

**Proof.** Using the least squares definition of  $\hat{\theta}_1$ ,

$$\hat{\sigma}^2 = (n-2)^{-1} \left[ \sum_{t=1}^n (e_t - \bar{e})^2 - (\hat{\theta}_1 - \theta_1)^2 \sum_{t=1}^n (Y_{t-1} - \bar{y}_{(-1)})^2 \right].$$

We have shown that  $(\hat{\theta}_1 - \theta_1) = O_p(|\theta_1|^{-n})$  and that  $\sum_{t=1}^n (Y_{t-1} - \bar{y}_{(-1)})^2 = O_p(|\theta_1|^{-2n})$ . Because the  $e_t$  are iid( $0, \sigma^2$ ), the mean square of the  $e_t$  converges in probability to  $\sigma^2$ .  $\blacktriangle$

**Corollary 10.2.1.2.** Suppose that  $e_t \sim NI(0, \sigma^2)$ . Then, under the model of Theorem 10.2.1,

$$\hat{\tau} = (\hat{\theta}_1 - \theta_1) \left[ \hat{\sigma}^{-2} \sum_{t=1}^n (Y_{t-1} - \bar{y}_{(-1)})^2 \right]^{1/2} \xrightarrow{\mathcal{L}} N(0, 1). \quad (10.2.16)$$

**Proof.** We have

$$\begin{aligned} & (\hat{\theta}_1 - \theta_1) \left( \sum_{t=1}^n (Y_{t-1} - \bar{y}_{(-1)})^2 \right)^{1/2} \\ &= [(\theta_1^2 - 1)X^{-1}Z_n][(\theta_1^2 - 1)^{-1/2}|X|] + O_p(n|\theta_1|^{-n}) \\ &= (\operatorname{sgn} X)(\theta_1^2 - 1)^{1/2}Z_n + O_p(n|\theta_1|^{-n}). \end{aligned}$$

The random variables  $X$  and  $Z_n$  are asymptotically independent, and the variance of  $Z_n$  converges to  $(\theta_1^2 - 1)^{-1}\sigma^2$ . The result follows, because  $\hat{\sigma}^2 \xrightarrow{P} \sigma^2$  by Corollary 10.2.1.1.  $\blacktriangle$

Because the pivotal  $\hat{\tau}$  is approximately  $N(0, 1)$  in large samples,  $\hat{\tau}$  may be used to test hypotheses concerning  $\theta_1$  and to set confidence intervals for  $\theta_1$  when the original errors are normally distributed.

We now extend these results to the  $p$ th order process with a single root larger than one in absolute value. Let

$$Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} = \begin{cases} \theta_0 + e_t, & t = 1, 2, \dots, \\ y_t, & t = -p+1, -p+2, \dots, 0, \end{cases} \quad (10.2.17)$$

where  $(y_{-p+1}, y_{-p+2}, \dots, y_0)$  is an initial vector and  $\{e_t\}_{t=-\infty}^{\infty}$  is a sequence of independent identically distributed  $(0, \sigma^2)$  random variables. Let  $|m_1| > |m_2| \geq |m_3| \geq \dots \geq |m_p|$  be the roots of the characteristic equation

$$m_p + \sum_{j=1}^p \theta_j m^{p-j} = 0. \quad (10.2.18)$$

Assume  $|m_1| > 1$  and all other roots are less than one in absolute value. We can also write the first equation of model (10.2.17) as

$$Y_t = \theta_0 + m_1 Y_{t-1} + \sum_{j=2}^p \theta_j (Y_{t-j+1} - m_1 Y_{t-j}) + e_t, \quad (10.2.19)$$

where the roots of

$$m^{p-1} - \sum_{j=2}^p \theta_j m^{p-j} = \prod_{j=1}^{p-1} (m - m_{j+1}) = 0 \quad (10.2.20)$$

are the  $p - 1$  roots of (10.2.18) that are less than one in absolute value. We consider a conceptual regression with the true  $m_1$ , denoted by  $m_1^0$ , used to define the regression variables. Let

$$\hat{\theta} = \left( \sum_{t=1}^n \mathbf{L}_t' \mathbf{L}_t \right)^{-1} \sum_{t=1}^n \mathbf{L}_t' Y_t, \quad (10.2.21)$$

where

$$\mathbf{L}_t = (1, Y_{t-1}, Y_{t-1} - m_1^0 Y_{t-2}, \dots, Y_{t-p+1} - m_1^0 Y_{t-p}),$$

be the least squares estimator of  $\theta = (\theta_0, \theta_1, \dots, \theta_p)'$ , where  $\theta_1 = m_1$ . The conclusion of Theorem 10.2.2 is analogous to that of Theorem 10.1.2 in that the limiting distribution of the coefficient of  $Y_{t-1}$  is closely related to the limiting distribution in the first order case and the other coefficients have the limiting distribution associated with the stationary process.

Two alternative assumptions for the initial conditions can be considered. In the first, the vector  $(y_{-p+1}, y_{-p+2}, \dots, y_0)$  is treated as fixed. In the second, we let

$$Y_t = \begin{cases} \theta_0 + \theta_1 Y_{t-1} + Z_t, & t = 1, 2, \dots, \\ y_0, & t = 0, \end{cases} \quad (10.2.22)$$

where  $y_0$  is fixed and  $\{Z_t\}_{t=-\infty}^\infty$  is a stationary autoregressive process with characteristic equation (10.2.20). The limiting distributions are the same under the two formulations.

**Theorem 10.2.2.** Suppose the model (10.2.17), (10.2.22) holds with  $|\theta_1^0| > 1$  and the roots of (10.2.20) less than one in absolute value, where  $\theta^0$  is the true parameter vector. Let  $\hat{\theta}$  be defined by (10.2.21), and let

$$\xi_n = n^{1/2} [(\hat{\theta}_0 - \theta_0^0), (\hat{\theta}_2 - \theta_2^0), \dots, (\hat{\theta}_p - \theta_p^0)]'.$$

Then  $d_{n+1}^{1/2}(\hat{\theta}_1 - \theta_1^0)$  converges to a random variable with mean zero and variance  $\sigma^2$  and

$$\xi_n \xrightarrow{\mathcal{L}} N(0, \mathbf{A}^{-1} \sigma^2),$$

where  $\mathbf{A} = E\{(1, Z_{t-1}, \dots, Z_{t-p})'(1, Z_{t-1}, \dots, Z_{t-p})\}$  and  $d_{n+1} = \sum_{t=2}^n Y_{t-1}^2$ . Furthermore,  $\xi_n$  and  $d_{n+1}^{1/2}(\hat{\theta}_1 - \theta_1^0)$  are independent in the limit. If  $e_t \sim NI(0, \sigma^2)$ , then

$$d_{n+1}^{1/2}(\hat{\theta}_1 - \theta_1^0) \xrightarrow{\mathcal{L}} N(0, \sigma^2).$$

**Proof.** We have

$$\mathbf{D}_n^{1/2}(\hat{\theta} - \theta^0) = \left( \mathbf{D}_n^{-1/2} \sum_{t=1}^n \mathbf{L}'_t \mathbf{L}_t \mathbf{D}_n^{-1/2} \right)^{-1} \mathbf{D}_n^{-1/2} \sum_{t=1}^n \mathbf{L}'_t e_t,$$

where

$$\mathbf{D}_n = \text{diag} \left[ n, \left( \sum_{t=2}^n Y_{t-1}^2 \right), n, \dots, n \right].$$

By the arguments used in Theorem 10.2.1

$$Y_t = (m_1^0)' \left[ y_0 + \theta_0^0 \sum_{j=1}^t (m_1^0)^{-j} + \sum_{j=1}^t (m_1^0)^{-j} Z_j \right] = (m_1^0)' U_t$$

and  $U_t$  converges to

$$U = y_0 + \theta_0^0(m_1^0 - 1)^{-1} + \sum_{j=1}^{\infty} (m_1^0)^{-j} Z_j.$$

To simplify the notation, we now write  $m_1$  for  $m_1^0$ . It follows from the arguments used to obtain (10.2.12) and (10.2.14) that

$$\frac{\sum_{t=1}^n Y_{t-1} e_t}{(\sum_{t=1}^n Y_{t-1}^2)^{1/2}} = \frac{\sum_{t=1}^n m_1^{-(n-t+1)} e_t}{(m_1^2 - 1)^{-1/2}} \text{sgn } U + O_p(n|m_1|^{-n}).$$

Now,

$$\sum_{t=1}^n Y_{t-1} Z_{t-j} = \sum_{t=1}^n m_1^{t-1} U Z_{t-j} + \sum_{t=1}^n m_1^{t-1} (U_{t-1} - U) Z_{t-j}.$$

We have

$$E\{|U_t - U| | Z_{t-j}\} = O(|m_1|^{-t})$$

and

$$E \left\{ \sum_{t=1}^n m_1^{t-1} (U_{t-1} - U) Z_{t-j} \right\} = O(n|m_1|^{-n}).$$

Therefore,

$$m_1^{-n} \sum_{t=1}^n Y_{t-1} Z_{t-j} = U \sum_{t=1}^n m_1^{-n+t-1} Z_{t-j} + O_p(n|m_1|^{-n})$$

and

$$\left( n \sum_{t=1}^n Y_{t-1}^2 \right)^{-1/2} \sum_{t=1}^n Y_{t-1} Z_{t-j} = O_p(n^{-1/2}).$$

It follows that

$$\mathbf{D}_n^{-1/2} \sum_{t=1}^n \mathbf{L}'_t \mathbf{L}_t \mathbf{D}_n^{-1/2} \longrightarrow \mathbf{A}^*,$$

where the row in  $\mathbf{A}^*$  associated with  $\hat{\theta}_1$  is  $(0, 1, 0, \dots, 0)$ . The matrix obtained by deleting the second row and column of  $\mathbf{A}^*$  is  $\mathbf{A}$ . By the arguments of Theorem 8.2.1,

$$n^{-1/2} \left[ \sum_{t=1}^n e_t, \sum_{t=1}^n Z_{t-1} e_t, \dots, \sum_{t=1}^n Z_{t-p} e_t \right]$$

converges to a normal vector random variable and we have the limiting result for  $\xi_n$ . The remaining results follow from arguments used in the proof of Theorem 10.2.1. ▲

In Theorem 10.2.2 we obtained the limiting distribution of estimators where some of the explanatory variables were a function of the unknown true maximum root. These results can be used to construct confidence intervals for the largest root. This operation is illustrated in Example 10.2.1. In Corollary 10.2.2 we give the limiting distribution of the ordinary regression pivotal for the largest root.

**Corollary 10.2.2.** Let  $(\tilde{\theta}_0, \tilde{m}_1, \tilde{\theta}_2, \dots, \tilde{\theta}_p)$  be the nonlinear least squares estimator of the parameters of (10.2.19), where  $|\tilde{m}_1|$  is at least as large as the largest absolute value of the roots of the estimated equation (10.2.20). Let the assumptions of Theorem 10.2.2 hold with  $e_t \sim NI(0, \sigma^2)$ . Let

$$\tilde{\mathbf{V}} = \left[ \sum_{t=1}^n \tilde{\mathbf{L}}'_t \tilde{\mathbf{L}}_t \right]^{-1} \hat{\sigma}^2,$$

where  $\tilde{\mathbf{L}}_t = (1, Y_{t-1} - \sum_{j=2}^p \tilde{\theta}_j Y_{t-j}, Y_{t-1} - \tilde{m}_1 Y_{t-2}, \dots, Y_{t-p+1} - \tilde{m}_1 Y_{t-p})$ , and

$$\hat{\sigma}^2 = (n - p - 1)^{-1} \sum_{t=1}^n \left[ Y_t - \tilde{\theta}_0 - \tilde{m}_1 Y_{t-1} - \sum_{j=2}^p \tilde{\theta}_j (Y_{t-j+1} - \tilde{m}_1 Y_{t-j}) \right]^2.$$

Then

$$\tilde{v} = \tilde{v}_{11}^{-1/2} (\tilde{m}_1 - m_1^0) \xrightarrow{\mathcal{L}} N(0, 1),$$

where  $\tilde{v}_{11}$  is the second diagonal element of  $\tilde{V}$ , the element associated with  $\tilde{m}_1$ .

**Proof.** The nonlinear least squares estimator for  $m_1$ , based on the model (10.2.19), differs from the root calculated with the characteristic equation of the linear least squares fit of (10.2.17) only in that the  $\tilde{m}_1$  from the nonlinear fit is always a real number. By Theorem 10.2.2, the linear least squares estimators are consistent. Because the true largest root is real, the probability approaches one that the two estimation procedures agree. The estimated characteristic equation, based on the estimator of (10.2.21), is

$$m^p - (\hat{\theta}_1 + \hat{\theta}_2)m^{p-1} - \sum_{j=2}^p (\hat{\theta}_{j+1} - m_1^0 \hat{\theta}_j)m^{p-j} + m_1^0 \hat{\theta}_p = 0. \quad (10.2.23)$$

The coefficients of  $Y_{t-j}$ ,  $j = 1, 2, \dots, p$ , obtained from the estimator (10.2.21) are the same as those obtained by the least squares fit of the model (10.2.17). Therefore, the roots of (10.2.23) are the same as the roots of the estimated characteristic equation associated with the least squares estimator of (10.2.17).

By a Taylor expansion, we have that the largest root of (10.2.23), denoted by  $\hat{m}_1$ , satisfies

$$\begin{aligned} \hat{m}_1 - m_1^0 &= \left[ (m_1^0)^{p-1} - \sum_{j=2}^p \theta_j^0 (m_1^0)^{p-j} \right]^{-1} (m_1^0)^{p-1} (\hat{\theta}_1 - m_1^0) \\ &\quad + o_p(|m_1^0|^{-n}) \end{aligned} \quad (10.2.24)$$

and  $\tilde{m}_1 - m_1^0 = O_p(|m_1^0|^{-n})$ . It follows that

$$\sum_{t=1}^n (Y_{t-j+1} - \tilde{m}_1 Y_{t-j})^2 = \sum_{t=1}^n (Y_{t-j+1} - m_1^0 Y_{t-j})^2 + O_p(1).$$

Also,

$$\sum_{t=1}^n \left( Y_{t-1} - \sum_{j=2}^p \tilde{\theta}_j Y_{t-j} \right)^2 = \sum_{t=1}^n \left( Y_{t-1} - \sum_{j=2}^p \theta_j^0 Y_{t-j} \right)^2 + o_p(|m_1^0|^n),$$

where

$$\sum_{t=1}^n \left( Y_{t-1} - \sum_{j=2}^p \theta_j^0 Y_{t-j} \right)^2 = U^2 \sum_{t=1}^n \left[ (m_1^0)^{t-1} \left( 1 - \sum_{j=2}^p \theta_j^0 (m_1^0)^{-j+1} \right) \right]^2 + O_p(|m_1^0|^n) \quad (10.2.25)$$

and  $U$  is defined in the proof of Theorem 10.2.2. By the arguments used in the proof of Theorem 10.2.2,

$$\left[ n \sum_{t=1}^n \left( Y_{t-1} - \sum_{j=2}^p \theta_j^0 Y_{t-j} \right)^2 \right]^{-1/2} \sum_{t=1}^n \left( Y_{t-1} - \sum_{j=2}^p \theta_j^0 Y_{t-j} \right) Z_{t-j} = O_p(n^{-1/2}).$$

Therefore,

$$\operatorname{plim}_{n \rightarrow \infty} \sigma^{-2} \sum_{t=1}^n \left( Y_{t-1} - \sum_{j=2}^p \theta_j^0 Y_{t-j} \right)^2 \tilde{v}_{11} = 1.$$

Because the multiplier for  $\hat{\theta}_1 - m_1^0$  in (10.2.24) is the square root of the multiplier for  $(m_1^0)^{2n}[(m_1^0)^2 - 1]U^2$  in (10.2.25), the result follows from the limiting distribution of  $d_{n1}^{1/2}(\hat{\theta}_1 - \theta_1^0)$  given in Theorem 10.2.2.  $\blacktriangle$

**Example 10.2.1.** Engle and Kraft (1981) analyzed the logarithm of the implicit price deflator for the gross national product as an autoregressive time series. The analysis we present follows Fuller (1984). We simplify the model of Engle and Kraft and use data for the period 1955 first quarter through 1980 third quarter. If we use least squares to estimate an autoregressive equation of the third order, 100 observations are used in the regression. The estimated characteristic equation associated with the third order process is

$$m^3 - 1.429m^2 + 0.133m + 0.290 = 0,$$

and the largest root of this equation is 1.0178. Because the largest root is greater than one, the estimated model is explosive. Before accepting an explosive model, it is reasonable to test the hypothesis that the largest root is one. To test the hypothesis of a unit root, we regress the first difference on  $Y_{t-1}$  and the lagged first differences. The estimated equation is

$$\begin{aligned} \hat{Y}_t - Y_{t-1} &= -0.0211 + 0.0054 Y_{t-1} + 0.423 (Y_{t-1} - Y_{t-2}) \\ &\quad (0.0082) \quad (0.0020) \quad (0.098) \\ &\quad + 0.290 (Y_{t-2} - Y_{t-3}), \\ &\quad (0.099) \end{aligned}$$

where the numbers in parentheses are the estimated standard errors obtained from the ordinary least squares regression program. By Theorem 10.1.4, the statistic

$$\hat{\tau}_\mu = (0.0020)^{-1}(0.0054) = 2.72$$

has the distribution tabulated in the second part of Table 10.A.2 when the largest root is one. According to that table,  $\hat{\tau}_A$  will exceed 0.63 about 1% of the time when the true root is one. Therefore, the hypothesis of a unit root is easily rejected.

To set confidence limits for the largest root, we use Corollary 10.2.2. Let the coefficient of  $Y_{t-1}$  in the regression of  $Y_t - m_1 Y_{t-1}$  on  $1, Y_{t-1}, Y_{t-1} - m_1 Y_{t-2}$ , and  $Y_{t-2} - m_1 Y_{t-3}$  be denoted by  $\hat{b}$ . If  $m_1 > 1$  is the largest root of the characteristic equation and if all other roots are less than one in absolute value, then the limiting distribution of the statistic  $\hat{t} = (\text{s.e. } \hat{b})^{-1} \hat{b}$ , where s.e.  $\hat{b}$  is the ordinary least squares standard error, is that of a  $N(0, 1)$  random variable. Therefore, we define a confidence set for  $m_1$  to be those  $m_1$  such that the absolute value of the calculated  $\hat{t}$  is less than the tabular value of Student's  $t$  for the desired confidence level. The confidence set can be constructed by fitting models for a range of  $m_1$ -values and choosing the values for which the test statistic is equal to the tabular value. For our data

$$\begin{aligned}\hat{Y}_t - 1.0091 Y_{t-1} = & -0.0211 + 0.00274 Y_{t-1} + 0.417 (Y_{t-1} - 1.0091 Y_{t-2}) \\ & (0.0082) \quad (0.00139) \quad (0.098) \\ & + 0.288 (Y_{t-2} - 1.0091 Y_{t-3}) \\ & (0.099)\end{aligned}$$

and

$$\begin{aligned}\hat{Y}_t - 1.0256 Y_{t-1} = & -0.0211 - 0.00254 Y_{t-1} + 0.406 (Y_{t-1} - 1.0256 Y_{t-2}) \\ & (0.0082) \quad (0.00128) \quad (0.098) \\ & + 0.283 (Y_{t-2} - 1.0256 Y_{t-3}). \\ & (0.097)\end{aligned}$$

It follows that a 95% confidence interval for  $m_1$  based on the large sample theory is (1.0091, 1.0256).

By Corollary 10.2.2, we can construct tests for the stationary part of the model. For example, the ordinary regression  $t$ -statistic for  $Y_{t-3} - m_1 Y_{t-4}$  in the regression of  $Y_t - m_1 Y_{t-1}$  on  $(1, Y_{t-1}, Y_{t-1} - m_1 Y_{t-2}, Y_{t-2} - m_1 Y_{t-3}, Y_{t-3} - m_1 Y_{t-4})$  has a  $N(0, 1)$  distribution in the limit. Because the  $t$ -statistic for the hypothesis that the coefficient for  $Y_{t-3} - m_1 Y_{t-4}$  is zero is identical (for any  $m_1 \neq 0$ ) to the  $t$ -statistic for the coefficient of  $Y_{t-4}$  in the regression of  $Y_t$  on  $(1, Y_{t-1}, Y_{t-2}, Y_{t-3}, Y_{t-4})$ , we have a test of the hypothesis that the process is third order against the hypothesis that it is fourth order. The fitted fourth order model, using the estimated value of  $m_1$  is

$$\begin{aligned}\hat{Y}_t - 1.0178 Y_{t-1} = & -0.0207 - 0.00000 Y_{t-1} + 0.409 (Y_{t-1} - 1.0178 Y_{t-2}) \\ & (0.0086) \quad (0.00114) \quad (0.104) \\ & + 0.280 (Y_{t-2} - 1.0178 Y_{t-3}) + 0.012 (Y_{t-3} - 1.0178 Y_{t-4}). \\ & (0.108) \quad (0.104)\end{aligned}$$

Because the regression pivotal for  $Y_{t-4}$  is  $t = (0.104)^{-1} 0.012 = 0.12$ , we easily

accept the hypothesis that the process is third order. The argument extends to the use of an  $F$  to test that the vector of coefficients for a set of variables is zero. ▲▲

We now consider the problem of predicting  $Y_{n+s}$  for the explosive autoregressive model. We derive the results for the first order model

$$Y_t = \theta_1 Y_{t-1} + e_t,$$

where the  $e_t$  are  $N(0, \sigma^2)$  random variables. Given  $(Y_1, Y_2, \dots, Y_n)$ , we take as our prediction of  $Y_{n+s}$ ,

$$\hat{Y}_{n+s} = \hat{\theta}_1^n Y_n, \quad (10.2.26)$$

where  $s > 0$ . The error in our prediction is

$$\hat{Y}_{n+s} - Y_{n+s} = (\hat{\theta}_1^s - \theta_1^s) Y_n - \sum_{j=1}^s \theta_1^{s-j} e_{n+j},$$

where

$$\hat{\theta}_1^s - \theta_1^s = s\theta_1^{s-1}(\hat{\theta}_1 - \theta_1) + O_p(|\theta_1|^{-2n}).$$

Also,

$$\hat{\theta}_1 - \theta_1 = \theta_1^{-n}(\theta_1^2 - 1)X^{-1}Z_n + O_p(n|\theta_1|^{-2n}),$$

where  $Z_n = \sum_{t=1}^n \theta_1^{t-n-1} e_t$  and  $X = y_0 + \sum_{j=1}^{\infty} \theta_1^{-j} e_j$ . Therefore,

$$\hat{Y}_{n+s} - Y_{n+s} = s\theta_1^{s-1}(\theta_1^2 - 1)Z_n - \sum_{j=1}^s \theta_1^{s-j} e_{n+j} + O_p(|\theta_1|^{-n}),$$

where we have used  $\theta_1^{-n} Y_n = X + O_p(\theta_1^{-n})$ . Now  $Z_n$  and  $\sum_{j=1}^s \theta_1^{s-j} e_{n+j}$  are independent,

$$\text{Var}(Z_n) = \sigma^2(\theta_1^2 - 1)^{-1}(1 - \theta_1^{-2n}),$$

and

$$\text{Var}\left(\sum_{j=1}^s \theta_1^{s-j} e_{n+j}\right) = \sigma^2(\theta_1^{2s} - 1)(\theta_1^2 - 1)^{-1}.$$

Therefore, the variance of the leading term in the prediction error is

$$\text{Var}(\epsilon_n) = \sigma^2[s^2\theta_1^{2(s-1)}(\theta_1^2 - 1) + (\theta_1^2 - 1)^{-1}(\theta_1^{2s} - 1)] + O(|\theta_1|^{-2n}), \quad (10.2.27)$$

where

$$\epsilon_n = s\theta_1^{s-1}(\theta_1^2 - 1)Z_n - \sum_{j=1}^s \theta_1^{s-j} e_{n+j}.$$

The variance of the prediction error may be estimated by

$$\hat{V}(\hat{Y}_{n+s} - Y_{n+s}) = \hat{\sigma}^2 \left[ \frac{\hat{\theta}_1^{2s} - 1}{\hat{\theta}_1^2 - 1} + \frac{(s\hat{\theta}_1^{s-1})^2 Y_n^2}{\sum_{t=1}^n Y_{t-1}^2} \right]. \quad (10.2.28)$$

For  $s = 1$ , this estimated variance reduces to the familiar regression formula for the variance of a prediction.

In the explosive case, unlike the case with  $|\theta_1| \leq 1$ , the error in  $\hat{\theta}_1$  contributes to the leading term of the variance of the prediction error. It follows that the usual approximations, such as (8.5.13), are not formally correct when the process has a root larger than one. However, the approximations based upon regression formulas remain appropriate for predictions with  $s > 1$  and for higher order processes with a root greater than one in absolute value.

**Example 10.2.2.** In this example, we construct predictions using the model of Example 10.2.1 and the method of indicator variables described in Example 8.5.1. To obtain predictions for three periods, we write

$$Y_t = \theta_0 + \theta_1 X_{t1} + \theta_2 X_{t2} + \theta_3 X_{t3} + \theta_5 X_{t5} \\ + (\theta_6 - \theta_1 \theta_5) X_{t6} + (\theta_7 - \theta_1 \theta_6 - \theta_2 \theta_5) X_{t7} + e_t,$$

where the regression variables are defined in Table 8.5.1 of Example 8.5.1 and the estimator of  $(\theta_5, \theta_6, \theta_7)$  is the predictor of  $(Y_{n+1}, Y_{n+2}, Y_{n+3})$ . Using a nonlinear least squares program and the data for 1955–1 through 1980–4, we obtain

$$(\hat{Y}_{81-1}, \hat{Y}_{81-2}, \hat{Y}_{81-3}) = (4.5144, \quad 4.5410, \quad 4.5672), \\ (0.0044) \quad (0.0074) \quad (0.0107)$$

where the numbers in parentheses are the standard errors output by the nonlinear least squares program. There are 117 observations used in the regression, and the residual mean square is  $\hat{\sigma}^2 = 0.0000179$ . Using this estimate of  $\sigma^2$ , the approximation (8.5.13) gives 0.0042 as the standard error for the one-period prediction. Thus, for a root close to one the difference between the appropriate estimator and the approximation based on the stationary model is not large for a one-period prediction.  $\blacktriangle \blacktriangle$

## 10.3. MULTIVARIATE AUTOREGRESSIVE PROCESSES WITH UNIT ROOTS

### 10.3.1. Multivariate Random Walk

Before studying estimators for the general autoregressive model, we give the distribution of the least squares estimator for a multivariate version of the first order unit root process. Let

$$\mathbf{Y}_t = \begin{cases} \mathbf{Y}_{t-1} + \mathbf{e}_t, & t = 1, 2, \dots, \\ \mathbf{Y}_0, & t = 0, \end{cases} \quad (10.3.1)$$

where  $\mathbf{Y}_t$  is a  $k$ -dimensional column vector and the  $\mathbf{e}_t$  are independent identically distributed  $(\mathbf{0}, \Sigma_{ee})$  random variables. It is always possible to transform the vector  $\mathbf{Y}_t$  into a new vector  $\mathbf{X}_t$  that satisfies an equation of the form (10.3.1) with  $\Sigma_{ee} = \mathbf{I}$ . One transformation is

$$\mathbf{X}_t = \mathbf{T}\mathbf{Y}_t,$$

where  $\mathbf{T} = \Sigma_{ee}^{-1/2} = \mathbf{Q}\mathbf{D}_e^{-1/2}\mathbf{Q}'$ ,  $\mathbf{Q}'$  is the orthogonal matrix such that

$$\mathbf{Q}'\Sigma_{ee}\mathbf{Q} = \mathbf{D}_e,$$

$\mathbf{Q}'\mathbf{Q} = \mathbf{I}$ , and  $\mathbf{D}_e$  is diagonal.

We write the general first order autoregressive process as

$$\mathbf{Y}_t = \mathbf{H}_1 \mathbf{Y}_{t-1} + \mathbf{e}_t, \quad (10.3.2)$$

and consider the least squares estimator of  $\mathbf{H}_1$ ,

$$\hat{\mathbf{H}}_1' = \left( \sum_{t=2}^n \mathbf{Y}_{t-1} \mathbf{Y}_{t-1}' \right)^{-1} \sum_{t=2}^n \mathbf{Y}_{t-1} \mathbf{Y}_t'. \quad (10.3.3)$$

If  $\mathbf{H}_1 = \mathbf{I}$ , the normalized error in the least squares estimator is

$$n(\hat{\mathbf{H}}_1' - \mathbf{I}) = \left( n^{-2} \sum_{t=2}^n \mathbf{Y}_{t-1} \mathbf{Y}_{t-1}' \right)^{-1} n^{-1} \sum_{t=2}^n \mathbf{Y}_{t-1} \mathbf{e}_t'.$$

The quantities in the expression converge in distribution to random variables that are analogous to the limit random variables of Theorem 10.1.1.

The first order vector autoregressive model with an intercept is

$$\mathbf{Y}_t = \mathbf{H}_0 + \mathbf{H}_1 \mathbf{Y}_{t-1} + \mathbf{e}_t, \quad (10.3.4)$$

where  $\mathbf{H}_0$  is a  $k$ -dimensional column vector. The ordinary least squares estimator of  $\mathbf{H}_1$  for the model (10.3.4) is

$$\begin{aligned} \hat{\mathbf{H}}_{1,\mu}' &= \left[ \sum_{t=2}^n (\mathbf{Y}_{t-1} - \bar{\mathbf{Y}}_{(-1)})(\mathbf{Y}_{t-1} - \bar{\mathbf{Y}}_{(-1)})' \right]^{-1} \\ &\quad \times \sum_{t=2}^n (\mathbf{Y}_{t-1} - \bar{\mathbf{Y}}_{(-1)})(\mathbf{Y}_t - \bar{\mathbf{Y}}_{(0)})', \end{aligned} \quad (10.3.5)$$

where  $(\bar{\mathbf{Y}}_{(-1)}, \bar{\mathbf{Y}}_{(0)}) = (n-1)^{-1} \sum_{t=2}^n (\mathbf{Y}_{t-1}, \mathbf{Y}_t)$ .

**Lemma 10.3.1.** Let the model (10.3.1) hold. Let  $\mathbf{Y}_0$  be a vector random variable with finite covariance matrix, perhaps the zero matrix, independent of  $\{\mathbf{e}_t\}$ . Let  $\{\mathbf{e}_t\}_{t=1}^\infty$  be a sequence of random vectors, and let  $\mathcal{A}_{t-1}$  be the sigma-field generated by  $\{\mathbf{e}_1, \mathbf{e}_2, \dots, \mathbf{e}_{t-1}\}$ . Assume

$$E\{(\mathbf{e}_t, \mathbf{e}_t \mathbf{e}_t') | \mathcal{A}_{t-1}\} = (\mathbf{0}, \Sigma_{ee}) \quad \text{a.s.}$$

and

$$E\{|\mathbf{e}_t|^{2+\delta} \mid \mathcal{A}_{t-1}\} < M < \infty \quad \text{a.s.}$$

for some  $\delta > 0$ , or assume the  $\mathbf{e}_t$  are iid( $\mathbf{0}, \Sigma_{ee}$ ), where  $\Sigma_{ee}$  is positive definite. Let  $\hat{\mathbf{H}}'_1$  be defined by (10.3.3), and  $\hat{\mathbf{H}}'_{1,\mu}$  by (10.3.5). Then

$$n\Sigma_{ee}^{1/2}(\hat{\mathbf{H}}'_1 - \mathbf{I})\Sigma_{ee}^{-1/2} \xrightarrow{\mathcal{L}} \mathbf{G}^{-1}\mathbf{Y},$$

and

$$n\Sigma_{ee}^{1/2}(\hat{\mathbf{H}}'_{1,\mu} - \mathbf{I})\Sigma_{ee}^{-1/2} \xrightarrow{\mathcal{L}} (\mathbf{G} - \zeta\zeta')^{-1}(\mathbf{Y} - \zeta\eta'),$$

where

$$\begin{aligned} \mathbf{G} &= \sum_{i=1}^{\infty} \gamma_i^2 \mathbf{U}_i \mathbf{U}'_i = \int_0^1 \mathbf{W}(r) \mathbf{W}'(r) dr, \\ \mathbf{Y} &= \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} 2(\gamma_i + \gamma_j)^{-1} \gamma_i^2 \gamma_j \mathbf{U}_i \mathbf{U}'_j - 0.5\mathbf{I} = \int_0^1 \mathbf{W}(r) d\mathbf{W}'(r), \\ \zeta &= \sum_{i=1}^{\infty} 2^{1/2} \gamma_i^2 \mathbf{U}_i = \int_0^1 \mathbf{W}(r) dr, \\ \eta &= \sum_{i=1}^{\infty} 2^{1/2} \gamma_i \mathbf{U}_i = \mathbf{W}(1), \\ \gamma &= (-1)^{l+1} 2[(2l-1)\pi]^{-1}, \end{aligned}$$

$\mathbf{U}_l = (U_{1l}, U_{2l}, \dots, U_{kl})'$ ,  $l = 1, 2, \dots$ , is a sequence of NI( $\mathbf{0}, \mathbf{I}$ ) vectors, and  $\mathbf{W}(r)$  is the standard vector Wiener process.

**Proof.** The result follows immediately from Theorem 5.3.7. See Phillips and Durlauf (1986) and Nagaraj (1990).  $\blacktriangle$

The extension of Theorem 5.3.7 to a moving average process will be used repeatedly in this section.

**Lemma 10.3.2.** Let  $\{\mathbf{e}_t\}$  satisfy the assumptions of Theorem 5.3.7. Let

$$\mathbf{Y}_t = \sum_{i=1}^t \sum_{j=0}^{\infty} \mathbf{K}_j \mathbf{e}_{i-j},$$

where

$$\sum_{j=0}^{\infty} \|\mathbf{K}_j\| < \infty, \quad \sum_{j=0}^{\infty} \mathbf{K}_j = \mathbf{C},$$

$\|\mathbf{K}_j\|$  is the maximum of the absolute values of the elements of  $\mathbf{K}_j$ , and  $\mathbf{C}$  is a nonsingular real matrix. Then

$$\begin{aligned}
n^{-3/2} \sum_{t=1}^n \mathbf{Y}_t &\Rightarrow \mathbf{C} \boldsymbol{\Sigma}_{ee}^{1/2} \int_0^1 \mathbf{W}(u) du = \mathbf{C} \boldsymbol{\Sigma}_{ee}^{1/2} \boldsymbol{\zeta}, \\
n^{-2} \sum_{t=1}^n \mathbf{Y}_t \mathbf{Y}'_t &\Rightarrow \mathbf{C} \boldsymbol{\Sigma}_{ee}^{1/2} \int_0^1 \mathbf{W}(u) \mathbf{W}'(u) du \boldsymbol{\Sigma}_{ee}^{1/2} \mathbf{C}' = \mathbf{C} \boldsymbol{\Sigma}_{ee}^{1/2} \mathbf{G} \boldsymbol{\Sigma}_{ee}^{1/2} \mathbf{C}', \\
n^{-2} \sum_{t=1}^n (\mathbf{Y}_t - \bar{\mathbf{y}}_n)(\mathbf{Y}_t - \bar{\mathbf{y}}_n)' &\Rightarrow \mathbf{C} \boldsymbol{\Sigma}_{ee}^{1/2} (\mathbf{G} - \boldsymbol{\zeta} \boldsymbol{\zeta}') \boldsymbol{\Sigma}_{ee}^{1/2} \mathbf{C}', \\
n^{-1} \sum_{t=1}^n \mathbf{Y}_{t-1} \mathbf{e}'_t &\Rightarrow \mathbf{C} \boldsymbol{\Sigma}_{ee}^{1/2} \int_0^1 \mathbf{W}(u) d\mathbf{W}'(u) \boldsymbol{\Sigma}_{ee}^{1/2} = \mathbf{C} \boldsymbol{\Sigma}_{ee}^{1/2} \mathbf{Y} \boldsymbol{\Sigma}_{ee}^{1/2}, \\
n^{-1} \sum_{t=1}^n (\mathbf{Y}_{t-1} - \bar{\mathbf{y}}_n) \mathbf{e}'_t &\Rightarrow \mathbf{C} \boldsymbol{\Sigma}_{ee}^{1/2} [\mathbf{Y} - \boldsymbol{\zeta} \mathbf{W}'(1)] \boldsymbol{\Sigma}_{ee}^{1/2},
\end{aligned}$$

where  $\mathbf{G}$ ,  $\mathbf{Y}$ ,  $\boldsymbol{\zeta}$ , and  $\mathbf{W}(u)$  are defined in Lemma 10.3.1.

**Proof.** Using the arguments in the proof of Theorem 10.1.2, the sums of squares and products of the  $\mathbf{Y}_t$  can be approximated by the sums of squares and products of  $\mathbf{C} \boldsymbol{\Sigma}_{j=1}^t \mathbf{e}_j$ . The results then follow from Theorem 5.3.7.  $\blacktriangle$

### 10.3.2. Vector Process with a Single Unit Root

In this section we examine estimation of the parameters of a vector autoregressive process whose characteristic equation has a unit root. We begin by considering least squares estimation of the parameters for two special equations. The first is

$$Y_t = \theta_1 Y_{t-1} + \boldsymbol{\beta}' \mathbf{X}_{t-1} + e_t, \quad t = 1, 2, \dots, \quad (10.3.6)$$

where the  $e_t$  are independent  $(0, \boldsymbol{\Sigma}_{ee})$  random variables with bounded  $2 + \delta$  ( $\delta > 0$ ) moments and  $Y_t$  is a scalar. Assume  $E\{e_t | \mathcal{A}_{t-1}\} = 0$ , where  $\mathcal{A}_{t-1}$  is the sigma-field generated by  $[(Y_{t-1}, \mathbf{X}_{t-1}, e_{t-1}), (Y_{t-2}, \mathbf{X}_{t-2}, e_{t-2}), \dots, (Y_0, \mathbf{X}_0, e_0)]$ . Assume that  $\mathbf{X}_t$  is a zero mean stationary  $k$ -dimensional  $p$ th order autoregressive process satisfying

$$\mathbf{X}_t + \sum_{j=1}^p \mathbf{A}_j \mathbf{X}_{t-j} = \boldsymbol{\epsilon}_t \quad (10.3.7)$$

where the  $\boldsymbol{\epsilon}_t$  are independent  $(0, \boldsymbol{\Sigma}_{ee})$  random variables with bounded  $2 + \delta$  ( $\delta > 0$ ) moments.

If  $\theta_1 = 1$ , then from (10.3.6),

$$Y_t = Y_0 + \sum_{j=1}^t (\boldsymbol{\beta}' \mathbf{X}_{j-1} + e_j). \quad (10.3.8)$$

Also, from (10.3.7) and Theorem 2.8.1,  $\boldsymbol{\beta}' \mathbf{X}_{t-1}$  has a representation as an infinite moving average

$$\boldsymbol{\beta}' \mathbf{X}_{t-1} = \boldsymbol{\beta}' \sum_{j=0}^{\infty} \mathbf{K}_j \boldsymbol{\epsilon}_{t-1-j}. \quad (10.3.9)$$

Hence, by the proof of Theorem 10.1.2,

$$Y_t = Y_0 + \sum_{j=1}^t u_j, \quad (10.3.10)$$

where  $u_t = \mathbf{c}' \boldsymbol{\epsilon}_{t-1} + e_t$ , and  $\mathbf{c}' = \boldsymbol{\beta}' \sum_{j=0}^{\infty} \mathbf{K}_j$ . These properties of the processes can be used to obtain the limiting distribution of the least squares estimator of  $(\theta_1, \boldsymbol{\beta}')$ .

**Theorem 10.3.1.** Let the model (10.3.6)–(10.3.7) hold with  $\theta_1 = 1$ , where  $(e_t, \boldsymbol{\epsilon}_t')$  is a sequence of independent zero mean vectors with common covariance matrix and bounded  $2 + \delta$  ( $\delta > 0$ ) moments. Let  $\hat{\boldsymbol{\theta}}$  be the least squares estimator of  $(\theta_1, \boldsymbol{\beta}')$ , where

$$\hat{\boldsymbol{\theta}} = (\hat{\theta}_1, \hat{\boldsymbol{\beta}}')' = \left( \sum_{t=2}^n \mathbf{F}_t' \mathbf{F}_t \right)^{-1} \sum_{t=2}^n \mathbf{F}_t' Y_t, \quad (10.3.11)$$

and  $\mathbf{F}_t = (Y_{t-1}, \mathbf{X}_{t-1}')$ . Let

$$t_1 = [\hat{V}\{\hat{\theta}_1\}]^{-1/2}(\hat{\theta}_1 - 1),$$

where  $\hat{V}\{\hat{\theta}_1\}$  is the least squares estimator of the variance of  $\hat{\theta}_1$ . Then

$$\hat{\theta}_1 - 1 = O_p(n^{-1}),$$

$$t_1 \xrightarrow{\mathcal{L}} \rho_{ue} \hat{\tau}_1 + (1 - \rho_{ue}^2)^{1/2} d,$$

and

$$n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \Gamma_{XX}^{-1} \sigma_{ee}),$$

where  $\Gamma_{XX} = E\{\mathbf{X}_t \mathbf{X}_t'\}$ ,  $\hat{\tau} \xrightarrow{\mathcal{L}} \hat{\tau}_1$ ,  $\hat{\tau}$  is defined in Corollary 10.1.1.2,  $\hat{\tau}_1$  has the limiting distribution of Corollary 10.1.1.2,  $u_t$  is defined in (10.3.10),  $\rho_{ue}$  is the correlation between  $u_t$  and  $e_t$ ,  $d$  is a  $N(0, 1)$  random variable, and  $\hat{\tau}_1$  is independent of  $d$ . If  $(\theta_1, \boldsymbol{\beta}') = (1, \mathbf{0})$ , then

$$n(\hat{\theta}_1 - 1) \xrightarrow{\mathcal{L}} (2G)^{-1}(T^2 - 1),$$

where  $(G, T)$  is defined in Theorem 10.1.1.

**Proof.** From representation (10.3.10) and Theorem 10.1.1, we have

$$n^{-2} \sum_{t=2}^n Y_{t-1}^2 \xrightarrow{\mathcal{L}} \sigma_{uu} G.$$

Also

$$\left( \sum_{t=2}^n Y_{t-1}^2 \right)^{-1} \sum_{t=2}^n Y_{t-1} e_t = O_p(n^{-1}),$$

because

$$V \left\{ \sum_{t=2}^n Y_{t-1} e_t \right\} = \sum_{t=2}^n E\{Y_{t-1}^2\} \sigma_{ee} = O(n^2).$$

Following the arguments used in the proof of Theorem 10.1.2, we have

$$\begin{aligned} n^{-3/2} \sum_{t=2}^n Y_{t-1} \mathbf{X}'_{t-1} &= O_p(n^{-1/2}), \\ \left( n^{-1} \sum_{t=2}^n Y_{t-1} e_t, n^{-1/2} \sum_{t=2}^n \mathbf{X}'_{t-1} e_t \right) &= O_p(1). \end{aligned} \quad (10.3.12)$$

Thus,

$$\begin{aligned} \mathbf{D}_n(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) &= \left( \mathbf{D}_n^{-1} \sum_{t=2}^n \mathbf{F}' \mathbf{F}_t \mathbf{D}_n^{-1} \right)^{-1} \mathbf{D}_n^{-1} \sum_{t=2}^n \mathbf{F}'_t e_t \\ &= \mathbf{B}_n^{-1} \mathbf{D}_n^{-1} \sum_{t=2}^n \mathbf{F}'_t e_t + O_p(n^{-1/2}), \end{aligned}$$

where

$$\mathbf{B}_n = \text{block diag} \left( n^{-2} \sum_{t=2}^n Y_{t-1}^2, \Gamma_{XX} \right)$$

and  $\mathbf{D}_n = \text{diag}(n, n^{1/2}, n^{1/2}, \dots, n^{1/2})$ . Therefore,

$$n(\hat{\theta}_1 - 1) = \left( n^{-2} \sum_{t=2}^n Y_{t-1}^2 \right)^{-1} n^{-1} \sum_{t=2}^n Y_{t-1} e_t + O_p(n^{-1/2}) = O_p(1).$$

If  $\boldsymbol{\beta} = \mathbf{0}$ , then  $Y_t = \sum_{j=1}^t e_j$ ,

$$n(\hat{\theta}_1 - 1) = \left[ \sum_{t=2}^n \left( \sum_{j=1}^{t-1} e_j \right)^2 \right]^{-1} \sum_{t=2}^n \left( \sum_{j=1}^{t-1} e_j \right) e_t + O_p(n^{-1/2}),$$

and

$$n(\hat{\theta}_1 - 1) \xrightarrow{\mathcal{L}} (2G)^{-1}(T^2 - 1).$$

The distributional result for  $n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$  follows because  $n^{-1/2} \sum_{t=2}^n \mathbf{X}'_{t-1} e_t$  satisfies the conditions of Theorem 5.3.3 with  $Z_{tn} = n^{-1/2} \mathbf{X}'_{t-1} e_t$ .

To establish the limiting result for  $t_1$ , we see that

$$\begin{aligned} t_1 &= \left[ s_e^2 \left( \sum_{t=2}^n Y_{t-1}^2 \right)^{-1} \right]^{-1/2} \left( \sum_{t=2}^n Y_{t-1}^2 \right)^{-1} \sum_{t=2}^n Y_{t-1} e_t + O_p(n^{-1/2}) \\ &= \sigma_{ee}^{-1/2} \left[ \sum_{t=2}^n \left( \sum_{j=1}^{t-1} u_j \right)^2 \right]^{-1/2} \sum_{t=2}^n \left( \sum_{j=1}^{t-1} u_j \right) e_t + o_p(1). \end{aligned}$$

Now

$$\sum_{t=2}^n \left( \sum_{j=1}^{t-1} u_j \right) e_t = \sum_{t=2}^n \left( \sum_{j=1}^{t-1} u_j \right) (b_{eu} u_t + d_t),$$

where  $b_{eu} = \sigma_{uu}^{-1} \sigma_{ue}$ , and  $d_t = e_t - b_{eu} u_t$  is uncorrelated with  $u_t$ . By Corollary 10.1.1.1,

$$\sigma_{ee}^{-1/2} b_{eu} \left[ \sum_{t=2}^n \left( \sum_{j=1}^{t-1} u_j \right)^2 \right]^{-1/2} \sum_{t=2}^n \left( \sum_{j=1}^{t-1} u_j \right) u_t \xrightarrow{\mathcal{L}} \rho_{ue} \hat{\tau}_t.$$

By Lemma 10.3.1,

$$n^{-1} \sigma_{uu}^{-1/2} \sigma_{ee}^{-1/2} \sum_{t=2}^n Y_{t-1} d_t \xrightarrow{\mathcal{L}} \sum_{i=1}^{\infty} \sum_{j=1}^{\infty} 2(\gamma_i + \gamma_j)^{-1} \gamma_i^2 \gamma_j U_{ui} U_{dj} = Y_{ud},$$

where the limit normal random variables  $(U_{ui}, U_{dj})$  are defined for the process  $(u_t, d_t)$  by the expression following (10.1.12). Because  $u_t$  and  $d_t$  are uncorrelated,  $U_{ui}$  and  $U_{dj}$  are independent. Therefore, the conditional distribution of  $Y_{ud}$  conditional on  $\{U_{u1}, U_{u2}, \dots\}$  is normal with mean zero and variance

$$\sum_{j=1}^{\infty} \left[ \sum_{i=1}^{\infty} 2(\gamma_i + \gamma_j)^{-1} \gamma_i^2 \gamma_j U_{ui} \right]^2 = \sum_{i=1}^{\infty} \gamma_i^2 U_{ui}^2.$$

Also,

$$\sigma_{dd}^{-1/2} \left( \sum_{i=1}^{\infty} \gamma_i^2 U_{ui}^2 \right)^{-1/2} Y_{ud} \sim N(0, 1)$$

for almost every  $\{U_{u1}, U_{u2}, \dots\}$ . It follows that  $(\sum_{i=1}^{\infty} \gamma_i^2 U_{ui}^2)^{-1/2} Y_{ud}$  is independent of  $\hat{\tau}_t$ , and

$$\left( \sum_{t=2}^n Y_{t-1}^2 \right)^{-1/2} \sum_{t=2}^n Y_{t-1} d_t \xrightarrow{\mathcal{L}} \sigma_{dd}^{1/2} \xi, \quad (10.3.13)$$

where  $\xi$  is a  $N(0, 1)$  random variable independent of  $\hat{\tau}_t$ . ▲

The  $\mathbf{X}_{t-1}$  of the model (10.3.6) is a rather general vector. The vector could contain current values of an explanatory variable and might contain variables of the form  $\Delta Y_{t-i} = Y_{t-i} - Y_{t-i-1}$  for  $i \geq 1$ . If  $\mathbf{X}_{t-1}$  contains only  $\Delta Y_{t-i}$ ,  $i =$

$1, 2, \dots, p - 1$ , then we obtain Theorem 10.1.2 as a special case of Theorem 10.3.1.

In Theorem 10.3.1, the model contains no intercept. If the model contains a constant term, then the pivotal statistic for the coefficient of lagged  $Y$  is a linear combination of  $\hat{\tau}_\mu$  of Table 10.A.2 and a  $N(0, 1)$  random variable.

Note that the usual regression statistics are appropriate, in large samples, for testing and confidence statements concerning the  $\beta$ -parameters. The usual pivotal statistic for the coefficient for  $Y_{t-1}$  has the distribution of Table 10.A.2 if  $\rho_{ue} = 0$ . This can happen if the only elements of  $\mathbf{X}_{t-1}$  with nonzero coefficients are elements of the form  $\Delta Y_{t-j}$ ,  $j \geq 1$ . If one is interested in testing  $\rho = 1$  in a model such as (10.3.6), the statistic  $\hat{\lambda}_k$  of Theorem 10.3.4 can be computed for the vector process  $(Y_t, \mathbf{X}'_t)$ .

A special case of the model (10.3.6) is that in which  $Y_t$  is a function of a vector, say  $\mathbf{X}_{1,t-1}$ , and an error that is autoregressive with parameter  $\theta_1$ . The model can be written

$$\begin{aligned} Y_t &= \boldsymbol{\beta}'_1 \mathbf{X}_{1,t-1} + a_t, \\ a_t &= \theta_1 a_{t-1} + e_t, \end{aligned}$$

or as

$$Y_t = \boldsymbol{\beta}'_1 \mathbf{X}_{1,t-1} - \theta_1 \boldsymbol{\beta}'_1 \mathbf{X}_{1,t-2} + \theta_1 Y_{t-1} + e_t.$$

Under rather general conditions for  $\mathbf{X}_{1,t}$  and  $\theta_1 = 1$ , the nonlinear least squares estimator  $(\hat{\boldsymbol{\beta}}'_1, \hat{\theta}_1)$  satisfies

$$\begin{aligned} n^{1/2}(\hat{\boldsymbol{\beta}}_1 - \boldsymbol{\beta}_1) &\xrightarrow{\mathcal{D}} N(\mathbf{0}, \mathbf{A}_{XX}^{-1}), \\ [\hat{V}\{\hat{\theta}_1\}]^{-1/2}(\hat{\theta}_1 - 1) &\xrightarrow{\mathcal{D}} \hat{\tau}_1, \end{aligned}$$

where  $\hat{V}\{\hat{\theta}_1\}$  is the usual nonlinear least squares estimator of the variance of  $\hat{\theta}_1$ , and

$$\mathbf{A}_{XX} = E\{(\mathbf{X}_{t-1} - \theta_1 \mathbf{X}_{t-2})(\mathbf{X}_{t-1} - \theta_1 \mathbf{X}_{t-2})'\}.$$

We next consider an equation in which one of the explanatory variables satisfies an autoregressive equation with a unit root.

**Theorem 10.3.2.** Let

$$\mathbf{X}_t = \boldsymbol{\beta}_1 Y_{t-1} + \boldsymbol{\beta}'_2 \mathbf{Z}_{t-1} + e_t, \quad (10.3.14)$$

$$Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} = u_t, \quad (10.3.15)$$

$$\mathbf{Z}_t + \sum_{j=1}^r \mathbf{B}_j \mathbf{Z}_{t-j} = \mathbf{v}_t, \quad (10.3.16)$$

where  $(e_t, v'_t, u_t)$  is a sequence of independent zero mean random vectors with covariance matrix  $\Sigma$  and bounded  $2 + \delta$  ( $\delta > 0$ ) moments. Let one of the roots of

$$m^p + \sum_{j=1}^p \alpha_j m^{p-j} = 0$$

be one, and let the remaining  $p - 1$  roots be less than one in absolute value. Let  $\mathbf{Z}_t$  be a stationary process where the roots of

$$\left| Im' + \sum_{j=1}^r \mathbf{B}_j m'^{-j} \right| = 0$$

are less than one in absolute value. Let  $\beta = (\beta_1, \beta_2')'$ , and let

$$\hat{\beta} = \left( \sum_{t=2}^n \mathbf{F}'_{t-1} \mathbf{F}_{t-1} \right)^{-1} \sum_{t=2}^n \mathbf{F}'_{t-1} \mathbf{X}_t,$$

where  $\mathbf{F}_{t-1} = (Y_{t-1}, \mathbf{Z}'_{t-1})$ . Let

$$t_{\beta_1} = [\hat{V}\{\hat{\beta}_1\}]^{-1/2}(\hat{\beta}_1 - \beta_1),$$

where  $\hat{V}\{\hat{\beta}_1\}$  is the least squares estimator of the variance of  $\hat{\beta}_1$ . Then

$$\begin{aligned} \hat{\beta}_1 - \beta_1 &= O_p(n^{-1}), \\ n^{1/2}(\hat{\beta}_2 - \beta_2) &\xrightarrow{\mathcal{L}} N(\mathbf{0}, \Gamma_{ZZ}^{-1} \sigma_{ee}), \\ t_{\beta_1} &\xrightarrow{\mathcal{L}} \rho_{ue} \hat{\tau}_i + (1 - \rho_{ue}^2)^{1/2} d, \end{aligned} \quad (10.3.17)$$

where  $\Gamma_{ZZ} = E\{\mathbf{Z}\mathbf{Z}'\}$ ,  $\hat{\tau} \xrightarrow{\mathcal{L}} \hat{\tau}_i$ ,  $\hat{\tau}$  is defined in Corollary 10.1.1.2.,  $\rho_{ue}$  is the correlation between  $u_t$  and  $e_t$ ,  $u_t$  is defined in (10.3.15),  $d$  is a  $N(0, 1)$  random variable, and  $\hat{\tau}_i$  is independent of  $d$ .

**Proof.** From the proof of Theorem 10.1.2,

$$n^{-2} \sum_{t=2}^n Y_{t-1}^2 \xrightarrow{\mathcal{L}} c^2 \sigma_{uu} G,$$

where  $c = [1 + \sum_{i=1}^p (i-1)\alpha_i]^{-1}$ ,  $G$  is defined in Theorem 10.1.1, and

$$V\left\{ n^{-1} \sum_{t=2}^n Y_{t-1} e_t \right\} = n^{-2} \sum_{t=2}^n E\{Y_{t-1}^2\} \sigma_{ee} = O(1).$$

The error in the least squares estimator is

$$\begin{aligned}\hat{\beta} - \beta &= \left( \sum_{t=2}^n \mathbf{F}'_{t-1} \mathbf{F}_{t-1} \right)^{-1} \sum_{t=2}^n \mathbf{F}'_t e_t \\ &= \mathbf{B}_n^{-1} \mathbf{D}_n^{-1} \sum_{t=2}^n \mathbf{F}'_{t-1} e_t + O_p(n^{-1/2}),\end{aligned}$$

where  $\mathbf{D}_n = \text{diag}(n, n^{1/2}, n^{1/2}, \dots, n^{1/2})$ ,

$$\mathbf{B}_n = \text{block diag} \left( n^{-2} \sum_{t=2}^n Y_{t-1}^2, \Gamma_{zz} \right),$$

and we have used

$$\begin{aligned}\left( n^{-1} \sum_{t=2}^n Y_{t-1} \mathbf{Z}_{t-1}, \sum_{t=2}^n \mathbf{D}_n^{-1} \mathbf{F}'_{t-1} e_t \right) &= O_p(1), \\ n^{-1} \sum_{t=2}^n \mathbf{Z}'_{t-1} \mathbf{Z}_{t-1} - \Gamma_{zz} &= O_p(n^{-1/2}).\end{aligned}$$

Thus,

$$n(\hat{\beta}_1 - \beta_1) = \left( n^{-2} \sum_{t=2}^n Y_{t-1}^2 \right)^{-1} n^{-1} \sum_{t=2}^n Y_{t-1} e_t + O_p(n^{-1/2}) = O_p(1).$$

The limiting distribution of  $n^{-1/2} \sum_{t=2}^n \mathbf{Z}'_{t-1} e_t$  follows from Theorem 5.3.3.

The derivation of the limiting distribution of  $t_{\beta_1}$  follows that for  $t_1$  of Theorem 10.3.1.  $\blacktriangle$

Theorem 10.3.2 identifies the difficulties in making inferential statements about parameters of equations such as (10.3.15) where the explanatory variable is a time series. In Chapter 9 we assumed the explanatory variables to be fixed sequences. An operationally equivalent assumption is that the explanatory variables form a random sequence independent of the error sequence. In such cases, the  $t$ -statistics are  $N(0, 1)$  random variables in the limit. If an explanatory variable is an autoregressive process, the distribution of the  $t$ -statistics depends on the roots of the process and on the correlation properties of the error processes. If the autoregressive process defining the explanatory variable has a unit root and if the error in the autoregressive process is correlated with the error in the equation, then by Theorem 10.3.2 the usual  $t$ -statistics have nonstandard limiting distributions. If the explanatory process is stationary, then the  $t$ -statistics for the parameters of the model (10.3.6) are  $N(0, 1)$  random variables in the limit.

If it is known that the explanatory process has a unit root, then imposing that condition on the system (10.3.14), (10.3.15) produces efficient estimates of the parameters of interest, and the  $t$ -statistics converge in distribution to  $N(0, 1)$  random variables. See Phillips (1991) and Exercise 10.12. However, in many practical situations it is not known that the  $Y_t$ -process of equation (10.3.15) has a

unit root. In such situations, the investigator must recognize the potential effect of a unit root in the  $Y_t$ -process on inferences about  $\beta_1$ .

**Example 10.3.1.** In this example, we consider estimation for the model

$$\begin{aligned} X_t &= \beta_0 + \beta_1 Y_{t-1} + e_t, \\ Y_t &= \theta_0 + \theta_1 Y_{t-1} + \theta_2 \Delta Y_{t-1} + u_t, \end{aligned}$$

where  $(e_t, u_t)' \sim N(0, \Sigma)$ . Our situation is assumed to be that described in Section 10.1.3 in that we specify  $\theta_1 \in (-1, 1]$  and  $\theta_0 = 0$  if  $\theta_1 = 1$ . The data of Table 10.B.4 were artificially generated to satisfy the model. The ordinary least squares estimates of the two equations are

$$\begin{aligned} \hat{X}_t &= 12.992 + 0.7783 y_{t-1}, \\ &\quad (0.099) \quad (0.0075) \\ \hat{y}_t &= 0.9798 y_{t-1} + 0.7869 \Delta Y_{t-1}, \\ &\quad (0.0076) \quad (0.0597) \end{aligned}$$

where  $y_t = Y_t - \bar{y}_n$  and the ordinary least squares standard errors are given in parentheses. The estimated covariance matrix of  $(e_t, u_t)$  is

$$\hat{V}\{(e_t, u_t)'\} = \begin{pmatrix} 0.9618 & 0.7291 \\ 0.7291 & 0.9795 \end{pmatrix},$$

and the estimated correlation between  $e_t$  and  $u_t$  is  $\hat{\rho}_{12} = 0.751$ . The weighted least squares estimate of  $\theta_1$  is  $\hat{\theta}_{w1} = 0.9850$ , and the test for a unit root is  $\hat{t}_{w1} = -1.94$ . The estimator of  $\theta_1$  constructed from (10.1.55) is  $\tilde{\theta}_{w1} = 0.9921$ .

To estimate the parameters of the  $X$ -equation, we use a system method of estimation such as full information maximum likelihood or three-stage least squares, available in SAS/ETS®. The two equations are estimated simultaneously, subject to the restriction that  $\theta_1 = 0.9921$ . The standard errors computed by the program should not be used as the standard errors of the estimators. They are only appropriate if it is known that  $\theta_1 = 0.9921$ . We suggest that the standard errors from the initial ordinary least squares regression be modified as in Example 10.1.3. Thus, the standard error for  $y_{t-1}$  in the regression of  $y_t$  on  $y_{t-1}$  is increased by 15%, and the standard error for  $y_{t-1}$  in the regression of  $X_t$  on  $(1, y_{t-1})$  is multiplied by  $1 + 0.15\hat{\rho}_{12}$ . Our estimate of the model is

$$\begin{aligned} \hat{X}_t &= 12.937 + 0.7876 y_{t-1}, \\ &\quad (0.099) \quad (0.0083) \\ \hat{y}_t &= 0.9921 y_{t-1} + 0.7308 \Delta Y_{t-1}, \\ &\quad (0.0087) \quad (0.0597) \end{aligned}$$

where the numbers in parentheses are the modified ordinary least squares standard errors.  $\blacktriangle \blacktriangle$

We now investigate estimation for the vector autoregressive process. Let  $\mathbf{Y}_t$  be a  $p$ th order autoregressive process of dimension  $k$ . Let

$$\mathbf{Y}_t + \sum_{i=1}^p \mathbf{A}_i \mathbf{Y}_{t-i} = \mathbf{e}_t, \quad t = 1, 2, \dots, \quad (10.3.18)$$

where the  $\mathbf{e}_t$  are independent and identically distributed  $(\mathbf{0}, \Sigma_{ee})$  random vectors, and the  $\mathbf{A}_i$  are  $k \times k$  parameter matrices. The representation (10.3.18) was introduced in Section 2.8. Assume that the characteristic equation

$$\left| \mathbf{I}m^p + \sum_{i=1}^p \mathbf{A}_i m^{p-i} \right| = 0 \quad (10.3.19)$$

has  $g$  roots equal to one and that all remaining roots are less than one in absolute value. By analogy to (10.1.25), we can write

$$\mathbf{Y}_t = \mathbf{H}_1 \mathbf{Y}_{t-1} + \sum_{i=2}^p \mathbf{H}_i \Delta \mathbf{Y}_{t-i+1} + \mathbf{e}_t, \quad (10.3.20)$$

where  $\Delta \mathbf{Y}_t = \mathbf{Y}_t - \mathbf{Y}_{t-1}$ ,  $\mathbf{H}_1 = -\sum_{j=1}^p \mathbf{A}_j$ ,  $\mathbf{H}_p = \mathbf{A}_p$ , and

$$\mathbf{H}_{p-i} = \sum_{j=p-i}^p \mathbf{A}_j \quad (10.3.21)$$

for  $i = 2, 3, \dots, p-1$ . We will study the process in which the portion of the Jordan canonical form of  $\mathbf{H}_1$  associated with the unit roots is a  $g$ -dimensional identity matrix.

By Theorem 2.4.2, there exists a  $\mathbf{Q}$  such that

$$\mathbf{Q}^{-1} \mathbf{H}_1 \mathbf{Q} = \Lambda, \quad (10.3.22)$$

where the elements of  $\Lambda$  are determined by the roots of  $\mathbf{H}_1$ . If the roots of  $\mathbf{H}_1$  are distinct,  $\Lambda$  is a diagonal matrix with the roots on the diagonal. If  $\mathbf{H}_1$  has repeated roots, the matrix  $\Lambda$  is of the form described in Theorem 2.4.2. Assume that  $\Lambda$  is block diagonal with the  $g \times g$  identity matrix as the first block and  $\Lambda_{22}$  as the second block, where  $\Lambda_{22} - \mathbf{I}$  is nonsingular.

Let  $\mathbf{T}$  be a nonsingular matrix whose first  $g$  rows are the first  $g$  rows of  $\mathbf{Q}^{-1}$ , the rows associated with the roots of one. The remaining rows of  $\mathbf{T}$  are linear combinations of the rows of  $\mathbf{Q}^{-1}$  chosen such that all elements of  $\mathbf{T}$  are real. The matrix  $\mathbf{T}$  can be chosen so that the upper left  $g \times g$  matrix of  $\mathbf{T} \Sigma_{ee} \mathbf{T}'$  is an identity matrix. Let  $\mathbf{X}_t = \mathbf{T} \mathbf{Y}_t$ . Then

$$\mathbf{X}_t = \Phi_1 \mathbf{X}_{t-1} + \sum_{i=2}^p \Phi_i \Delta \mathbf{X}_{t-i+1} + \mathbf{a}_t, \quad (10.3.23)$$

where  $\Phi_i = T H_i T^{-1}$ ,  $i = 1, 2, \dots, p$ , and  $\mathbf{a}_t = T \mathbf{e}_t$ . The matrix  $\Phi_1$  has a  $g \times g$  identity matrix as the upper left block. The remaining elements of the first  $g$  rows and first  $g$  columns of  $\Phi_1$  are zero. It follows that the vector process

$$\hat{\mathbf{X}}_t = (X_{1,t} - X_{1,t-1}, X_{2,t} - X_{2,t-1}, \dots, X_{g,t} - X_{g,t-1}, X_{g+1,t}, X_{g+2,t}, \dots, X_{k,t})'$$

is a vector autoregressive process with all roots less than one in absolute value.

A vector process  $\mathbf{Y}_t$  for which there exists a transformation  $T$  such that  $k-g$  components of the transformed vector are stationary and  $g$  components are unit root processes is called *cointegrated of order  $k-g$*  in the econometrics literature. See Granger (1981) and Engle and Granger (1987). Granger's original definition of cointegration required every element of  $\mathbf{Y}_t$  to contain a unit root. The last  $k-g$  rows of  $T$  that define stationary processes are called *cointegrating vectors*.

A canonical form for the  $p$ th order process of interest is

$$\mathbf{Y}_t = \mathbf{H}_1 \mathbf{Y}_{t-1} + \sum_{i=2}^p \mathbf{H}_i \Delta \mathbf{Y}_{t-i+1} + \mathbf{e}_t, \quad (10.3.24)$$

where  $(\mathbf{Y}'_{1,t}, \mathbf{Y}'_{2,t}) = \mathbf{Y}'_t$ ,  $\mathbf{Y}'_{1,t}$  is the vector composed of the first  $g$  elements of the  $\mathbf{Y}_t$ , the upper left  $g \times g$  portion of  $\Sigma_{ee}$  is  $I_g$ ,  $\mathbf{H}_1 = \text{block diag}\{I_g, \mathbf{H}_{1,22}\}$ , and  $\mathbf{H}_{1,22} - I_{k-g}$  is nonsingular. The conformable partition of  $\mathbf{e}_t$  of (10.3.24) is  $(\mathbf{e}'_{1,t}, \mathbf{e}'_{2,t})'$ , where  $\mathbf{e}'_{1,t}$  is a  $g$ -dimensional vector. We can write

$$\mathbf{e}'_{2,t} = \mathbf{e}'_{1,t} \Sigma_{ee12} + \mathbf{a}_{2,t}, \quad (10.3.25)$$

where  $\Sigma_{ee12}$  is the  $g \times (k-g)$  upper right part of  $\Sigma_{ee}$ ,  $\mathbf{a}_{2,t}$  is uncorrelated with  $\mathbf{e}'_{1,t}$  and

$$E\{\mathbf{a}_{2,t} \mathbf{a}'_{2,t}\} = \Sigma_{aa22}. \quad (10.3.26)$$

The vector  $(\Delta \mathbf{Y}'_{1,t}, \mathbf{Y}'_{2,t})$  is a stationary process and can be expressed as

$$(\Delta \mathbf{Y}'_{1,t}, \mathbf{Y}'_{2,t})' = \sum_{j=0}^{\infty} \mathbf{K}_j \mathbf{e}_{t-j}, \quad (10.3.27)$$

by Theorem 2.8.1. Also,

$$\mathbf{Y}_{1,t} = \sum_{i=1}^t \sum_{j=2}^p \mathbf{H}_{j,11} \Delta \mathbf{Y}_{1,t-j+1} + \sum_{i=1}^t \sum_{j=2}^p \mathbf{H}_{j,12} \Delta \mathbf{Y}_{2,t-j+1} + \sum_{i=1}^t \mathbf{e}'_{1,i},$$

where  $\mathbf{H}_{j,11}$  is the upper left  $g \times g$  block of  $\mathbf{H}_j$  and  $\mathbf{H}_{j,12}$  is the upper right  $g \times (k-g)$  block of  $\mathbf{H}_j$ . Because  $\sum_{i=1}^t \Delta \mathbf{Y}_{2,i} = \mathbf{Y}_{2,t} - \mathbf{Y}_{2,1}$  and  $\Delta \mathbf{Y}_{1,t}$  is stationary, we have

$$\mathbf{Y}_{1t} = \sum_{j=2}^p \mathbf{H}_{j,11} \mathbf{Y}_{1t} + \sum_{i=1}^t \mathbf{e}_{1i} + O_p(1)$$

and

$$\begin{aligned} \mathbf{Y}_{1t} &= \left( \mathbf{I} - \sum_{j=2}^p \mathbf{H}_{j,11} \right)^{-1} \sum_{i=1}^t \mathbf{e}_{1i} + O_p(1), \\ &\stackrel{\text{def}}{=} \mathbf{C}_g \sum_{i=1}^t \mathbf{e}_{1i} + O_p(1), \end{aligned} \quad (10.3.28)$$

where the  $O_p(1)$  remainder is uncorrelated with  $\mathbf{e}_s$  for  $s > t$ .

Let the ordinary least squares estimators of the parameters of the model (10.3.24) be

$$\hat{\mathbf{H}}' = \left( \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{L}_{t-1} \right)^{-1} \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{Y}'_t, \quad (10.3.29)$$

and

$$\hat{\Sigma}_{ee} = (n - p - pk)^{-1} \sum_{t=p+1}^n \hat{\mathbf{e}}'_t \hat{\mathbf{e}}'_t, \quad (10.3.30)$$

where

$$\begin{aligned} \mathbf{L}_{t-1} &= (\mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-1}, \dots, \Delta \mathbf{Y}'_{t-p+1}), \\ \mathbf{H}' &= (\mathbf{H}_1, \mathbf{H}_2, \dots, \mathbf{H}_p)', \end{aligned}$$

and  $\hat{\mathbf{e}}'_t = \mathbf{Y}'_t - \hat{\mathbf{L}}_{t-1}' \hat{\mathbf{H}}'$ . The ordinary least squares estimator of the "covariance matrix" of  $\text{vec } \hat{\mathbf{H}}'$  is

$$\hat{\Sigma}_{ee} \otimes \left( \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{L}_{t-1} \right)^{-1}. \quad (10.3.31)$$

This matrix will normalize the estimators, but it is not an estimated covariance matrix in the traditional sense because some of its elements cannot be standardized to converge to constants. This is analogous to the results for the univariate case discussed in Section 10.1. The matrix  $\mathbf{V} = \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{L}_{t-1}$  is  $kp \times kp$ , and we denote the  $ij$ th  $k \times k$  block of the matrix by  $\mathbf{V}_{ij}$ .

From the derivation of the canonical form (10.3.24), we know that the least squares estimator is invariant to linear transformations in the sense that if  $\mathbf{T}$  is nonsingular, then the least squares estimators of the coefficients for the transformed model based on  $\mathbf{T}\mathbf{Y}$ , are  $\mathbf{T}\hat{\mathbf{H}}_i \mathbf{T}^{-1}$ ,  $i = 1, 2, \dots, p$ , where  $\hat{\mathbf{H}}_i$ ,  $i = 1, 2, \dots, p$ , are the least squares estimators for the original model. Also, the estimator of the error covariance matrix for the transformed problem is  $\mathbf{T}\hat{\Sigma}_{ee}\mathbf{T}^{-1}$ , where  $\hat{\Sigma}_{ee}$  is the regression residual mean square for the original problem.

We now give the limiting distribution of the estimators for the canonical model.

**Theorem 10.3.3.** Let  $\mathbf{Y}_t$  be a  $k$ -dimensional  $p$ th order autoregressive process written in the canonical form (10.3.24). Assume the  $\mathbf{e}_t$  are iid( $0, \Sigma_{ee}$ ) random variables or that the  $\mathbf{e}_t$  are martingale differences satisfying the conditions of Lemma 10.3.1. Partition  $\mathbf{Y}'_t$  as  $(\mathbf{Y}'_{1,t}, \mathbf{Y}'_{2,t})$ , where  $\mathbf{Y}_{1,t}$  is the  $g$ -dimensional vector with  $g$  unit roots. Assume that  $(\Delta\mathbf{Y}'_{1,t}, \mathbf{Y}'_{2,t})$  is a stationary process with  $E\{\Delta\mathbf{Y}_{1,t}\} = 0$  and  $\mathbf{Y}_{1,0} = 0$ . Let the ordinary least squares estimators be defined by (10.3.29) and (10.3.30). Then

$$\mathbf{D}_n(\hat{\mathbf{H}}' - \mathbf{H}') \xrightarrow{\mathcal{L}} \begin{pmatrix} \mathbf{C}_g^{-1} \mathbf{G}_{gg}^{-1} (\mathbf{Y}_{11}, \mathbf{Y}_{11} \Sigma_{ee12} + \mathbf{Y}_{12} \Sigma_{aa22}^{1/2}) \\ \boldsymbol{\Psi}_2 \end{pmatrix},$$

where  $\mathbf{C}_g = (\mathbf{I} - \sum_{j=2}^p \mathbf{H}_{j,11})^{-1}$  is defined in (10.3.28),  $\mathbf{a}_{2t}$  is defined in (10.3.25),

$$\mathbf{D}_n = \text{diag}(n, n, \dots, n, n^{1/2}, n^{1/2}, \dots, n^{1/2}),$$

$(\mathbf{G}, \mathbf{Y})$  is defined in Lemma 10.3.1,  $\mathbf{G}_{gg}$  is the upper left  $g \times g$  portion of  $\mathbf{G}$ ,  $\mathbf{Y}_{1.} = (\mathbf{Y}_{11}, \mathbf{Y}_{12})$  is the matrix composed of the first  $g$  rows of  $\mathbf{Y}$ , the elements of  $\boldsymbol{\Psi}_2$  are zero mean normal random variables with

$$\mathbf{V}\{\text{vec } \boldsymbol{\Psi}_2\} = \Sigma_{ee} \otimes \Sigma_{LL22}^{-1},$$

and

$$\Sigma_{LL22} = E\{(\mathbf{Y}'_{2,t-1}, \Delta\mathbf{Y}'_{t-1}, \dots, \Delta\mathbf{Y}'_{t-p+1})' (\mathbf{Y}'_{2,t-1}, \Delta\mathbf{Y}'_{t-1}, \dots, \Delta\mathbf{Y}'_{t-p+1})\}.$$

Also,  $\hat{\Sigma}_{ee}$  converges in probability to  $\Sigma_{ee}$ .

**Proof.** By (10.3.28) and by the arguments used in the proof of Theorem 10.1.2,

$$\mathbf{D}_n^{-1} \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{L}_{t-1} \mathbf{D}_n^{-1} \xrightarrow{\mathcal{L}} \begin{pmatrix} \mathbf{C}_g \mathbf{G}_{gg} \mathbf{C}'_g & \mathbf{0} \\ \mathbf{0} & \Sigma_{LL22} \end{pmatrix}.$$

Likewise, by Theorem 5.3.7 and Theorem 8.2.3,

$$\mathbf{D}_n^{-1} \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{e}'_t \xrightarrow{\mathcal{L}} \begin{pmatrix} \mathbf{C}_g (\mathbf{Y}_{11}, \mathbf{Y}_{11} \Sigma_{ee12} + \mathbf{Y}_{12} \Sigma_{aa22}^{1/2}) \\ \boldsymbol{\Phi}_2 \end{pmatrix},$$

where  $\boldsymbol{\Phi}_2$  is a matrix of zero mean normal random variables with covariance matrix

$$\mathbf{V}\{\text{vec } \boldsymbol{\Phi}_2\} = \Sigma_{ee} \otimes \Sigma_{LL22}.$$

The covariance between  $n^{-1/2} \sum \Delta\mathbf{Y}_{t-j} \mathbf{e}'_j$ ,  $j \geq 1$ , and  $n^{-1} \sum \mathbf{Y}_{t-1} \mathbf{e}'_t$  goes to zero as  $n$  increases and the limit random variables are independent. See Lai and Wei (1985a) and Chan and Wei (1988). Using

$$\mathbf{D}_n(\hat{\mathbf{H}}' - \mathbf{H}') \approx \left( \mathbf{D}_n^{-1} \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{L}_{t-1} \mathbf{D}_n^{-1} \right)^{-1} \mathbf{D}_n^{-1} \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{e}'_t,$$

we obtain the limiting distribution.

The estimated covariance matrix  $\hat{\Sigma}_{ee}$  converges to  $\Sigma_{ee}$  because

$$\begin{aligned} [n - (p + 1)k]^{-1} \sum_{t=p+1}^n (\mathbf{Y}'_t - \mathbf{L}_{t-1} \hat{\mathbf{H}}')' (\mathbf{Y}'_t - \mathbf{L}_{t-1} \hat{\mathbf{H}}') \\ = [n - (p + 1)k]^{-1} \sum_{t=p+1}^n \mathbf{e}_t \mathbf{e}'_t + O_p(n^{-1}). \end{aligned} \quad \blacktriangle$$

Theorem 10.3.3 gives the distribution for the parameters of the canonical form (10.3.24). The coefficient matrices for any other representation are linear combinations of the elements of the canonical coefficient matrices. Hence, the limiting distributions are defined in terms of those of Theorem 10.3.3. As in the univariate case, the limiting distribution of coefficients associated with unit roots differ from the limiting distribution of coefficients associated with the stationary portion of the model.

The following corollary to Theorem 10.3.3 was derived for  $g = 1$  by Fountis and Dickey (1989).

**Corollary 10.3.3.** Let  $\mathbf{Y}_t$  be a  $k$ -dimensional  $p$ th order autoregressive process with  $g$  unit roots and all other roots less than one in absolute value. Assume the process can be transformed to the canonical form (10.3.24) with  $g$  unit roots. Let the assumptions on  $\mathbf{e}_t$  of Theorem 10.3.3 hold. Let  $\hat{m}_1 \geq \hat{m}_2 \geq \dots \geq \hat{m}_g$  be the  $g$  roots of

$$|\mathbf{I}m^p + \hat{\mathbf{A}}_1 m^{p-1} + \dots + \hat{\mathbf{A}}_p| = 0,$$

with largest absolute values, where the  $\hat{\mathbf{A}}_i$ ,  $i = 1, 2, \dots, p$ , are the least squares estimators. Then the limiting distribution of  $n(\hat{m}_i - 1)$ ,  $i = 1, 2, \dots, g$ , is the distribution of the  $g$  roots of

$$|\mathbf{I}m - \mathbf{Y}'_1 \mathbf{G}^{-1}| = 0,$$

where  $(\mathbf{G}, \mathbf{Y})$  is the  $g \times 2g$  matrix defined in Lemma 10.3.1.

**Proof.** The roots of the characteristic equation are unchanged by nonsingular transformations and we assume, without loss of generality, that the equation is in the canonical form (10.3.24).

By Theorem 10.3.3, the least squares coefficients are converging to the true values. Therefore, by the results of Section 5.8, the roots of the characteristic equation are converging to the true roots. We expand the largest roots about one and write  $\hat{m}' = 1 + r\delta + o_p(\delta)$ . Then the local approximation to the determinantal equation becomes

$$|\mathbf{I}(1 + p\delta) + \hat{\mathbf{A}}_1[1 + (p - 1)\delta] + \hat{\mathbf{A}}_2[1 + (p - 2)\delta] + \cdots + \hat{\mathbf{A}}_p| = 0.$$

Using the definitions of  $\hat{\mathbf{H}}_i$  in terms of the  $\hat{\mathbf{A}}_i$ , we have

$$\left| [1 + (p - 1)\delta](\mathbf{I} - \hat{\mathbf{H}}_1) + \left( \mathbf{I} - \sum_{j=2}^p \hat{\mathbf{H}}_j \right) \delta \right| = 0. \quad (10.3.32)$$

The lower right portion of  $\mathbf{I} - \hat{\mathbf{H}}_1$  is converging to a nonsingular matrix. The matrix  $\mathbf{I} - \sum_{j=2}^p \hat{\mathbf{H}}_j$  is converging to a matrix, say  $\mathbf{M}$ , with  $\mathbf{C}_g^{-1}$  as the upper left  $g \times g$  block. Therefore, we multiply the matrix in (10.3.32) on the left by the nonsingular matrix

$$\begin{pmatrix} \mathbf{C}_g & \mathbf{0} \\ -\mathbf{M}_{21}\mathbf{C}_g & \mathbf{I} \end{pmatrix},$$

where  $\mathbf{M}_{21}$  is the lower left  $(k - g) \times g$  portion of  $\mathbf{M}$ . The values of  $\delta$  that satisfy the resulting determinantal equation are equal to the values that satisfy (10.3.32). Using the fact that the first  $g$  columns of  $\hat{\mathbf{H}}_1 - \mathbf{I}$  are  $O_p(n^{-1})$ , the  $g$  smallest roots of (10.3.32) are converging to the  $g$  smallest roots of

$$|\mathbf{C}_g(\hat{\mathbf{H}}_{1,11} - \mathbf{I}_g) - \mathbf{I}_g\delta| = 0.$$

The result follows from the limiting distribution of  $\hat{\mathbf{H}}_{1,11} - \mathbf{I}_g$  given in Theorem 10.3.3.  $\blacktriangle$

We now turn to the problem of estimating the parameters of (10.3.18) subject to the restriction that one of the roots of (10.3.19) is equal to one. If we write the model in the form (10.3.20), the determinant of  $\mathbf{H}_1 - \mathbf{I}$  is zero under the null model and there exists a vector  $\boldsymbol{\kappa}_1$ , with  $\boldsymbol{\kappa}'_1 \boldsymbol{\kappa}_1 \neq 0$ , such that

$$\boldsymbol{\kappa}'_1(\mathbf{H}_1 - \mathbf{I}) = \mathbf{0}. \quad (10.3.33)$$

Therefore, if the first observations are treated as fixed, the maximum likelihood estimator of  $(\mathbf{H}_1, \dots, \mathbf{H}_p)$  is obtained by imposing the restriction (10.3.33) on the least squares estimator. Assume the first element of  $\boldsymbol{\kappa}_1$ , denoted by  $\kappa_{11}$ , is not zero, and let  $\boldsymbol{\kappa}_{11}^{-1} \boldsymbol{\kappa}'_1 = (1, -\mathbf{B}_{0,12})$ . If we multiply (10.3.20) by  $\boldsymbol{\kappa}_{11}^{-1} \boldsymbol{\kappa}'_1$ , we have

$$\Delta Y_{1t} = \mathbf{B}_{0,12} \Delta Y_{2t} + \sum_{i=1}^{p-1} \mathbf{B}_{i+1,11} \Delta Y_{t-i} + v_{1t}, \quad (10.3.34)$$

where  $\mathbf{B}_{i+1,11} = \kappa_{11}^{-1} \boldsymbol{\kappa}'_1 \mathbf{H}_{i+1}$ ,  $\mathbf{Y}'_t = (Y_{1t}, Y'_{2t})$ , and  $v_{1t} = \kappa_{11}^{-1} \boldsymbol{\kappa}'_1 \mathbf{e}_t$ .

Let  $\boldsymbol{\theta} = (\mathbf{B}_{0,12}, \mathbf{B}_{2,1}, \dots, \mathbf{B}_{p,1})$ . Let  $\hat{\boldsymbol{\pi}} = (\hat{\pi}_1, \hat{\pi}_2, \hat{\pi}_3)$  be the matrix of regression coefficients obtained in the regression of  $[\Delta Y_{1t}, \Delta Y'_{2t}, (\Delta Y'_{t-1}, \dots, \Delta Y'_{t-p+1})]$  on  $(\mathbf{Y}'_{t-1}, \Delta Y'_{t-2}, \dots, \Delta Y'_{t-p+1})$ . Of course, the elements of  $\hat{\boldsymbol{\pi}}_3$  are zeros and ones. Let  $\hat{\lambda}_k$  be the smallest root of

$$\left| \hat{\pi}' \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{L}_{t-1} \hat{\pi} - \lambda \hat{\Sigma}_{\epsilon\epsilon} \right| = 0, \quad (10.3.35)$$

where

$$\hat{\Sigma}_{\epsilon\epsilon} = \text{block diag}(\hat{\Sigma}_{\epsilon\epsilon}, \mathbf{0}_{k(p-1), k(p-1)}),$$

$\mathbf{0}_{k(p-1), k(p-1)}$  is a  $k(p-1) \times k(p-1)$  matrix of zeros,  $\hat{\Sigma}_{\epsilon\epsilon}$  is defined in (10.3.30), and  $\mathbf{L}_{t-1} = (\mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-1}, \dots, \Delta \mathbf{Y}'_{t-p+1})$ . Then, the maximum likelihood estimator of  $\boldsymbol{\theta}$  is

$$\hat{\boldsymbol{\theta}} = [(\hat{\pi}_2, \hat{\pi}_3)' \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{L}_{t-1} (\hat{\pi}_2, \hat{\pi}_3) - \hat{\lambda}_k \hat{\Sigma}_{\epsilon\epsilon 22}]^{-1} \\ \times [(\hat{\pi}_2, \hat{\pi}_3)' \sum_{t=p+1}^n \mathbf{L}'_{t-1} \mathbf{L}_{t-1} \hat{\pi}_1 - \hat{\lambda}_k \hat{\Sigma}_{\epsilon\epsilon 21}], \quad (10.3.36)$$

where  $\hat{\Sigma}_{\epsilon\epsilon 22}$  is the lower right  $(kp-1) \times (kp-1)$  submatrix of  $\hat{\Sigma}_{\epsilon\epsilon}$ , and  $\hat{\Sigma}_{\epsilon\epsilon 21}$  is the lower left  $(kp-1) \times 1$  submatrix of  $\hat{\Sigma}_{\epsilon\epsilon}$ .

The vector  $\hat{\boldsymbol{\theta}}$  is called the limited information maximum likelihood estimator in the econometric literature. Hence,  $\hat{\boldsymbol{\theta}}$  can be calculated by any of the many econometric computer packages. The limiting distributions of the test statistic and the limiting distribution of the coefficients associated with the conditional likelihood were obtained by Johansen (1988) and Ahn and Reinsel (1990). We give a proof that differs somewhat from the proofs of those authors.

**Theorem 10.3.4.** Suppose the model (10.3.18) holds with exactly one root of (10.3.19) equal to one and exactly one root of  $\mathbf{H}_1$  equal to one. Assume the  $\mathbf{e}_t$  are iid( $\mathbf{0}, \Sigma_{\epsilon\epsilon}$ ) random variables or that the  $\mathbf{e}_t$  are martingale differences satisfying the conditions of Lemma 10.3.1. Assume that  $\mathbf{Y}_t$  is such that  $\kappa_t$  can be normalized to give equation (10.3.34). Let  $\hat{\boldsymbol{\theta}}$  of (10.3.36) be the estimator of  $\boldsymbol{\theta}$ . Let

$$\hat{\mathbf{V}}_{\theta\theta} = \left( \sum_{t=p+1}^n \hat{\psi}'_t \hat{\psi}_t \right)^{-1} (1, -\hat{\mathbf{B}}_{0,12}) \hat{\Sigma}_{\epsilon\epsilon} (1, -\hat{\mathbf{B}}_{0,12})',$$

where  $\hat{\Sigma}_{\epsilon\epsilon}$  is defined in (10.3.30),

$$\hat{\psi}_t = (\Delta Y_{2t}, \Delta Y_{3t}, \dots, \Delta Y_{kt}, \Delta \mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-2}, \dots, \Delta \mathbf{Y}'_{t-p+1}),$$

and  $\hat{\psi}_t$  is the predicted value of  $\psi_t$  from the regression of  $\psi_t$  on  $(\mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-1}, \dots, \Delta \mathbf{Y}'_{t-p+1})$ . Then

$$\hat{\mathbf{V}}_{\theta\theta}^{-1/2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I})$$

and

$$\hat{\lambda}_k \xrightarrow{\mathcal{L}} \hat{\tau}_I^2,$$

where  $\hat{\tau} \xrightarrow{d} \hat{\tau}_l$ ,  $\hat{\tau}$  is defined in Corollary 10.1.1.2, the distribution of  $\hat{\tau}_l$  is given in Corollary 10.1.1.2, and  $\hat{\lambda}_k$  is defined in (10.3.35).

**Proof.** Let  $\hat{\mathbf{Y}}_{t-1}'$  be the deviation from fit for the regression of  $\mathbf{Y}_{t-1}'$  on  $(\Delta\mathbf{Y}_{t-1}', \dots, \Delta\mathbf{Y}_{t-p+1}')$ . Then  $\hat{\lambda}_k$  is also the smallest root of

$$\left| (\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12})' \sum_{t=p+1}^n \hat{\mathbf{Y}}_{t-1} \hat{\mathbf{Y}}_{t-1}' (\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12}) - \lambda \hat{\Sigma}_{ee} \right| = 0,$$

where  $(\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12})$  is the upper  $k \times k$  portion of  $(\hat{\boldsymbol{\pi}}_1, \hat{\boldsymbol{\pi}}_2)$ . Note that  $(\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12}) = \hat{\mathbf{H}}_1' - \mathbf{I}$ .

Let  $\mathbf{T}_{11}$  be a transformation matrix such that the unit root is associated with the first element of  $\mathbf{X}_t = \mathbf{T}_{11} \mathbf{Y}_t$ . That is,  $\mathbf{X}_t$  satisfies the equation

$$\mathbf{X}_t = \mathbf{H}_{x1} \mathbf{X}_{t-1} + \sum_{i=2}^p \mathbf{H}_{xi} \Delta \mathbf{X}_{t-i+1} + \mathbf{T}_{11} \mathbf{e}_t,$$

where  $\mathbf{H}_{xi} = \mathbf{T}_{11} \mathbf{H}_i \mathbf{T}_{11}^{-1}$ ,  $\mathbf{H}_i$  is defined in (10.3.21), and the first row and first column of  $\mathbf{H}_{x1} - \mathbf{I}$  are vectors of zeros. Then

$$\begin{aligned} & (\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12})' \sum_{t=p+1}^n \hat{\mathbf{Y}}_{t-1} \hat{\mathbf{Y}}_{t-1}' (\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12}) \\ &= (\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12})' \mathbf{T}_{11}^{-1} \sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1} \hat{\mathbf{X}}_{t-1}' \mathbf{T}_{11}^{-1} (\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12}) \\ &= \hat{\boldsymbol{\xi}}' \sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1} \hat{\mathbf{X}}_{t-1}' \hat{\boldsymbol{\xi}}, \end{aligned}$$

where  $\hat{\boldsymbol{\xi}} = (\hat{\boldsymbol{\xi}}_1, \hat{\boldsymbol{\xi}}_2) = \mathbf{T}_{11}^{-1} (\hat{\boldsymbol{\pi}}_{11}, \hat{\boldsymbol{\pi}}_{12})$ , and  $\hat{\mathbf{X}}_{t-1}$  is the deviation from fit for the regression of  $\mathbf{X}_{t-1}$  on  $(\Delta\mathbf{Y}_{t-1}', \dots, \Delta\mathbf{Y}_{t-p+1}')$ . Also,

$$\left( \hat{\boldsymbol{\xi}}' \sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1} \hat{\mathbf{X}}_{t-1}' \hat{\boldsymbol{\xi}} - \hat{\lambda}_k \hat{\Sigma}_{ee} \right) (1, -\hat{\mathbf{B}}_{0,12})' = \mathbf{0}, \quad (10.3.37)$$

where  $\hat{\lambda}_k$  is the smallest root of (10.3.35) and the roots of (10.3.35) are the roots of

$$\left| \hat{\boldsymbol{\xi}}' \sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1} \hat{\mathbf{X}}_{t-1}' \hat{\boldsymbol{\xi}} - \hat{\lambda}_k \hat{\Sigma}_{ee} \right| = 0.$$

Now,  $(\sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1}^2)'^{-1} = O_p(n^{-2})$ ,  $\hat{\boldsymbol{\xi}}_1 = O_p(n^{-1})$ , and

$$\sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1} \Delta \mathbf{Y}_{jt} \left( \sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1}^2 \right)^{-1} \sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1} \Delta \mathbf{Y}_{it} = O_p(1)$$

for  $i = 1, 2, \dots, k$  and  $j = 1, 2, \dots, k$ , by the arguments used in the proof of Theorem 10.1.2. Therefore, the elements of the first row and column of  $\hat{\xi}' \Sigma \hat{X}_{t-1} \hat{X}'_{t-1} \hat{\xi}$  are  $O_p(1)$ . Also, because  $(X_2, \dots, X_k)$  is stationary with positive definite covariance matrix, the matrix of mean squares and products of  $(\hat{X}_2, \dots, \hat{X}_k)$  is an estimator of a nonsingular matrix. Also, by the assumption that  $H_1 - I$  has one zero root,  $\hat{\xi}_2$  is of rank  $k - 1$  and

$$(n - p)^{-1}(\hat{\pi}_2, \hat{\pi}_3)' \sum_{t=p+1}^n L'_{t-1} L_{t-1} (\hat{\pi}_2, \hat{\pi}_3) \xrightarrow{P} \bar{M},$$

where  $\bar{M}$  is a nonsingular matrix. It follows that  $\hat{\lambda}_k = O_p(1)$ . Also see Theorem 2.4.1 of Fuller (1987, p. 151).

From the definition of  $\hat{\theta}$ ,

$$\begin{aligned} \hat{\theta} - \theta &= \bar{M}^{-1}(n - p)^{-1}(\pi_2, \pi_3)' \sum_{t=p+1}^n L'_{t-1} L_{t-1} [\hat{\pi}_1 - (\hat{\pi}_2, \hat{\pi}_3)\theta] + o_p(n^{-1/2}) \\ &= \bar{M}^{-1}(n - p)^{-1}(\pi_2, \pi_3)' \sum_{t=p+1}^n L'_{t-1} e_t B'_{0,1} + o_p(n^{-1/2}), \end{aligned}$$

where  $B_{0,1} = (1, -B_{0,12})$  and  $\pi$  is the probability limit of  $\hat{\pi}$ . Because  $(\pi_2, \pi_3)' L_{t-1}$  is a stationary process, the asymptotic normality of  $n^{1/2}(\hat{\theta} - \theta)$  follows from Theorem 5.3.4. See Theorem 8.2.1.

Note that  $L_{t-1}(\hat{\pi}_2, \hat{\pi}_3)$  is the predicted value for  $\psi$  obtained in the regression of  $\psi$  on  $L_{t-1}$ . Because  $\hat{\pi}$  is converging to  $\pi$ ,  $\hat{B}_{0,12}$  is converging to  $B_{0,12}$ , and  $\hat{\Sigma}_{ee}$  is converging to  $\Sigma_{ee}$ , we have

$$(n - p)\hat{V}_{\theta\theta} \xrightarrow{P} \bar{M}^{-1} B_{0,1} \Sigma_{ee} B'_{0,1},$$

and the distribution result for  $\hat{V}_{\theta\theta}^{-1/2}(\hat{\theta} - \theta)$  is established.

We now obtain the limiting distribution of  $\hat{\lambda}_k$ . Because the roots are unchanged by linear transformation, we assume, without loss of generality, that the model is in the canonical form. In the canonical form,  $B_{0,1} = (1, 0, \dots, 0)$ . From (10.3.37),

$$\hat{\lambda}_k = (\hat{B}_{0,1} \hat{\Sigma}_{ee} \hat{B}'_{0,1})^{-1} \hat{B}_{0,1} \hat{\xi}' \hat{S}_{XX} \hat{\xi} \hat{B}'_{0,1},$$

where  $\hat{S}_{XX} = \sum_{t=p+1}^n \hat{X}_{t-1} \hat{X}'_{t-1}$ .

Using  $(\xi_1, \xi_2) B'_{0,1} = 0$ ,  $B_{0,12} = 0$ ,  $(\xi_1, \xi_2) = p\lim(\hat{\xi}_1, \hat{\xi}_2)$ , and a Taylor expansion, we have

$$\begin{aligned} \hat{\xi}_2 \hat{B}'_{0,12} &= \hat{\xi}_2 B'_{0,12} + \hat{\xi}_2 (\hat{B}'_{0,12} - B'_{0,12}) + O_p(n^{-1}) \\ &= \hat{\xi}_2 \hat{B}'_{0,12} + O_p(n^{-1}), \end{aligned}$$

where

$$\hat{\mathbf{B}}_{0,12}' = (\xi_2' \hat{\mathbf{S}}_{XX} \xi_2)^{-1} \xi_2' \sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1} e_{1t} + o_p(n^{-1/2})$$

and the first row of  $\xi_2$  is a zero vector. It follows that

$$\begin{aligned}\hat{\lambda}_k &= \sigma_{ee11}^{-1} (\hat{\xi}_1 - \xi_2 \hat{\mathbf{B}}_{0,12}') \hat{\mathbf{S}}_{XX} (\hat{\xi}_1 - \xi_2 \hat{\mathbf{B}}_{0,12}') + o_p(1) \\ &= \sigma_{ee11}^{-1} \left( \sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1} e_{1t} \right)' [\hat{\mathbf{S}}_{XX}^{-1} - \xi_2 (\xi_2' \hat{\mathbf{S}}_{XX} \xi_2)^{-1} \xi_2'] \sum_{t=p+1}^n \hat{\mathbf{X}}_{t-1} e_{1t} + o_p(1).\end{aligned}$$

The standardized matrix

$$\hat{\mathbf{D}}_{XX}^{-1/2} [\hat{\mathbf{S}}_{XX}^{-1} - \xi_2 (\xi_2' \hat{\mathbf{S}}_{XX} \xi_2)^{-1} \xi_2'] \hat{\mathbf{D}}_{XX}^{-1/2},$$

where  $\hat{\mathbf{D}}_{XX}^{-1} = \text{diag } \hat{\mathbf{S}}_{XX}$ , is converging to a matrix with  $(1, 0, \dots, 0)$  as the first row. Therefore,

$$\hat{\lambda}_k = \sigma_{ee11}^{-1} \left( \sum_{t=p+1}^n \hat{\mathbf{X}}_{1,t-1}^2 \right)^{-1} \left( \sum_{t=p+1}^n \hat{\mathbf{X}}_{1,t-1} e_{1t} \right)^2 + o_p(1)$$

and  $\hat{\lambda}_k \xrightarrow{\mathcal{L}} \hat{\tau}_{\mu,l}^2$ . See the proof of Theorem 10.1.2.  $\blacktriangle$

The results extend to the model with an intercept.

**Corollary 10.3.4.** Let the model

$$\mathbf{Y}_t = \mathbf{H}_0 + \mathbf{H}_1 \mathbf{Y}_{t-1} + \sum_{i=2}^p \mathbf{H}_i \Delta \mathbf{Y}_{t-i+1} + e_t, \quad t = 2, 3, \dots, n,$$

be estimated by ordinary least squares where  $\mathbf{H}_0$  is a  $k$ -dimensional column vector. Let  $\boldsymbol{\kappa}_1^0, \mathbf{H}_0^0 = \mathbf{0}$ , where  $\boldsymbol{\kappa}_1^0(\mathbf{H}_1^0 - \mathbf{I}) = \mathbf{0}$ , and  $\boldsymbol{\kappa}_1^0$  and  $\mathbf{H}_1^0$  are the true parameters of the process. Let all statistics be defined as in Theorem 10.3.4 except that the regression equations contain an intercept. Let the error assumptions of Theorem 10.3.4 hold. Let  $\hat{\boldsymbol{\theta}} = (\hat{\mathbf{B}}_{0,12}, \hat{\xi}_1, \hat{\mathbf{B}}_{2,1}, \dots, \hat{\mathbf{B}}_{p,1})$ , where  $\hat{\xi}_1$  is the estimated intercept for the model (10.3.4) expanded to include an intercept. Let  $\hat{\mathbf{V}}_{\theta\theta}$  be defined as in Theorem 10.3.4 with

$$\boldsymbol{\psi}_t = (\Delta Y_{2t}, \Delta Y_{3t}, \dots, \Delta Y_{kt}, 1, \Delta \mathbf{Y}_{t-1}', \Delta \mathbf{Y}_{t-2}', \dots, \Delta \mathbf{Y}_{t-p}').$$

Then

$$\hat{\mathbf{V}}_{\theta\theta}^{-1/2} (\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{I})$$

and

$$\hat{\lambda}_k \xrightarrow{\mathcal{L}} \hat{\tau}_{\mu,l}^2,$$

where  $\hat{\tau}_\mu \xrightarrow{d} \hat{\tau}_{\mu_1}$ , the limit random variable  $\hat{\tau}_{\mu_1}$  has the distribution given in Theorem 10.1.3, and  $\hat{\lambda}_k$  is defined by (10.3.35) with  $\mathbf{L}_{t-1}$  expanded to include a one.

**Proof.** Omitted. ▲

One method of estimating the entire system (10.3.20) subject to the restriction (10.3.33) is to use a nonlinear estimation program. The computations are illustrated in Example 10.3.1 of the next subsection.

### 10.3.3. Vector Process with Several Unit Roots

We now extend our discussion to the estimation of the multivariate process (10.3.18) subject to the restriction that exactly  $g$  roots of (10.3.19) are equal to one, under the assumption that the part of the Jordan canonical form of  $\mathbf{H}_1$  associated with the unit roots is the  $g$ -dimensional identity matrix. This means that there are  $g$  vectors  $\kappa_i$  such that

$$\kappa'_i(\mathbf{H}_1 - \mathbf{I}) = \mathbf{0}, \quad (10.3.38)$$

where the  $\kappa_i$  define a  $g$ -dimensional subspace.

Let

$$\hat{\mathbf{S}}_{hh}^* = (\hat{\mathbf{H}}_1 - \mathbf{I})\mathbf{V}_{11}^{-1}(\hat{\mathbf{H}}_1 - \mathbf{I})', \quad (10.3.39)$$

where  $\mathbf{V}_{11}$  is the portion of the inverse of

$$\sum_{t=p+1}^n (\mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-1}, \dots, \Delta \mathbf{Y}'_{t-p+1})' (\mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-1}, \dots, \Delta \mathbf{Y}'_{t-p+1})$$

associated with  $\mathbf{Y}_{t-1}$ . Let  $\hat{\kappa}_i$  be the characteristic vectors of  $\hat{\mathbf{S}}_{hh}$  in the metric  $\hat{\Sigma}_{ee}$ , and let

$$\hat{\mathbf{K}} = (\hat{\kappa}_1, \hat{\kappa}_2, \dots, \hat{\kappa}_k),$$

where  $\hat{\Sigma}_{ee}$  is defined in (10.3.30). That is,

$$(\hat{\mathbf{S}}_{hh}^* - \hat{\lambda}_i \hat{\Sigma}_{ee}) \hat{\kappa}_i = \mathbf{0}, \quad i = 1, 2, \dots, k, \quad (10.3.40)$$

where  $\hat{\lambda}_i$ ,  $i = 1, 2, \dots, k$ , are the roots of

$$|\hat{\mathbf{S}}_{hh}^* - \lambda \hat{\Sigma}_{ee}| = 0 \quad (10.3.41)$$

and  $\hat{\mathbf{K}}' \hat{\Sigma}_{ee} \hat{\mathbf{K}} = \mathbf{I}$ .

For normal  $e_t$ , the method of maximum likelihood defines the estimated  $g$ -dimensional subspace with the  $g$  vectors associated with the  $g$  smallest roots of (10.3.41). The vectors  $\hat{\kappa}_i$ ,  $i = k-g+1, k-g+2, \dots, k$ , define a  $g$ -dimensional

subspace, but any set of  $g$  vectors of rank  $g$  formed as linear combinations of the  $\hat{\kappa}_i$  can be used to define the subspace.

We have solved the problem of estimating those linear combinations of the original variables that are unit root processes. One could ask for those linear combinations of the original variables that define stationary processes, the cointegrating vectors. The cointegrating vectors are obtained in Example 10.3.2 using a maximum likelihood computer program. The cointegrating vectors can be computed directly using a determinantal equation different than (10.3.41). Define  $S_{00}$ ,  $S_{10}$ , and  $S_{01}$  by  $S_{01} = S'_{10}$ ,

$$H'_1 - I = V_{11} S_{10}, \quad (10.3.42)$$

and

$$d_f \hat{\Sigma}_{ee} = S_{00} - S_{01} V_{11} S_{10}, \quad (10.3.43)$$

where  $d_f$  is the degrees of freedom for  $\hat{\Sigma}_{ee}$ . For the first order model,  $V_{11}^{-1} = \sum_{t=2}^n Y_{t-1} Y'_{t-1}$  and

$$(S_{00}, S_{01}) = \left( \sum_{t=2}^n \Delta Y_t \Delta Y'_t, \sum_{t=2}^n \Delta Y_t Y'_{t-1} \right).$$

Then, the  $k-g$  cointegrating vectors are the vectors associated with the  $k-g$  largest roots of

$$|S_{10} S_{00}^{-1} S_{01} - \nu V_{11}^{-1}| = 0. \quad (10.3.44)$$

The roots  $\hat{\nu}$  of (10.3.44) are related to the roots of (10.3.41) by

$$\hat{\nu} = (1 + d_f^{-1} \hat{\lambda}) d_f^{-1} \hat{\lambda} \quad (10.3.45)$$

and  $\hat{\lambda} = (1 - \hat{\nu})^{-1} \hat{\nu} d_f$ .

Given that (10.3.38) holds, there is some arrangement of the elements of  $Y$ , such that

$$(I, -B_{0,12})(H_1 - I) = 0,$$

where  $(I, -B_{0,12})$  is a  $g \times k$  matrix obtained as a nonsingular transformation of  $(\kappa_1, \kappa_2, \dots, \kappa_g)'$ . Thus, if we let

$$B_0 = \begin{pmatrix} I, & -B_{0,12} \\ 0, & I \end{pmatrix},$$

we can multiply the model (10.3.20) by  $B_0$  to obtain

$$\mathbf{B}_0 \Delta \mathbf{Y}_t = \mathbf{B}_1 (\mathbf{Y}'_{1,t-1}, \mathbf{Y}'_{2,t-1})' + \sum_{j=1}^{p-1} \mathbf{B}_{j+1} \Delta \mathbf{Y}_{t-j} + \mathbf{v}_t, \quad (10.3.46)$$

where  $\mathbf{v}_t = \mathbf{B}_0 \mathbf{e}_t$ , and the first  $g$  rows of  $\mathbf{B}_1$  are composed of zeros. Then the maximum likelihood estimator of  $\mathbf{B}_{0,12}$  is

$$\hat{\mathbf{B}}_{0,12} = \hat{\Sigma}_{\pi\pi 22}^{-1} \hat{\Sigma}_{\pi\pi 21}, \quad (10.3.47)$$

where

$$\hat{\Sigma}_{\pi\pi} = \hat{\Sigma}_{ee} \hat{\mathbf{K}} [\text{Diag}(\hat{\lambda}_1 - 1, \hat{\lambda}_2 - 1, \dots, \hat{\lambda}_{k-g} - 1, 0, 0, \dots, 0)] \hat{\mathbf{K}}' \hat{\Sigma}_{ee},$$

$\hat{\Sigma}_{\pi\pi 22}$  is the lower right  $(k-g) \times (k-g)$  submatrix of  $\hat{\Sigma}_{\pi\pi}$ ,  $\hat{\mathbf{K}}$  is the matrix of vectors  $\hat{\kappa}_i$ , and  $\hat{\kappa}_i$ ,  $i = 1, 2, \dots, k$ , are the characteristic vectors in the metric  $\hat{\Sigma}_{ee}$  associated with equation (10.3.40). A derivation of this estimator in a different context is given in Fuller (1987, Theorems 4.1.1 and 4.1.2). The estimators of  $\mathbf{B}_{j,1}$ ,  $j = 2, 3, \dots, p$ , are given by the regression of  $\hat{\mathbf{B}}_0 \Delta \mathbf{Y}_t$  on  $(\Delta \mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-2}, \dots, \Delta \mathbf{Y}'_{t-p+1})$ .

We now give the limiting distributions for some of the statistics associated with maximum likelihood estimation.

**Theorem 10.3.5.** Assume that the model (10.3.19), (10.3.46) holds, where  $\mathbf{e}_t$  are iid  $(0, \Sigma_{ee})$  random vectors,  $g$  of the roots of  $\mathbf{H}_1$  are one, and  $k-g$  of the roots are less than one in absolute value. Assume that the part of the Jordan canonical form associated with the unit roots is diagonal. Let  $\hat{\mathbf{H}}_1$  be the least squares estimator of  $\mathbf{H}_1$  obtained in the least squares fit of equation (10.3.20). Let  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_k$  be the roots of the equation

$$|(\hat{\mathbf{H}}_1 - \mathbf{I}) \mathbf{V}_{11}^{-1} (\hat{\mathbf{H}}_1 - \mathbf{I})' - \lambda \hat{\Sigma}_{ee}| = 0,$$

where  $\hat{\Sigma}_{ee}$  is the least squares estimator of  $\Sigma_{ee}$ , and  $\mathbf{V}_{11}$  is defined with (10.3.39). Then the distribution of  $(\hat{\lambda}_{k-g+1}, \hat{\lambda}_{k-g+2}, \dots, \hat{\lambda}_k)$  converges to the distribution of the roots  $(\hat{\nu}_1, \hat{\nu}_2, \dots, \hat{\nu}_g)$  of

$$|\mathbf{Y}'_{gg} \mathbf{G}_{gg}^{-1} \mathbf{Y}_{gg} - \nu \mathbf{I}_g| = 0,$$

where  $\mathbf{G}_{gg}$  and  $\mathbf{Y}_{gg}$  are  $g \times g$  matrices defined in Lemma 10.3.1.

If the statistics are computed using deviations from the mean, the distribution of  $(\hat{\lambda}_{k-g+1}, \hat{\lambda}_{k-g+2}, \dots, \hat{\lambda}_k)$  converges to the distribution of the roots  $(\hat{\nu}_1, \hat{\nu}_2, \dots, \hat{\nu}_g)$  of

$$|(\mathbf{Y}_{gg} - \zeta \boldsymbol{\eta}') (\mathbf{G}_{gg} - \zeta \boldsymbol{\zeta}')^{-1} (\mathbf{Y}_{gg} - \zeta \boldsymbol{\eta}') - \nu \mathbf{I}_g| = 0,$$

where  $\zeta$  and  $\boldsymbol{\eta}$  are  $g$ -dimensional vectors defined in Lemma 10.3.1.

Let  $\hat{\mathbf{B}}_{0,12}$  be defined by (10.3.47), let  $\mathbf{B}_{j,1}$  be the  $g \times k$  matrix composed of the

first  $g$  rows of  $\mathbf{B}_j$ ,  $j = 2, 3, \dots, p$ , of (10.3.46), and let  $\hat{\mathbf{B}}_{j,1}$ ,  $j = 2, 3, \dots, p$ , be the associated estimators of  $\mathbf{B}_{j,1}$ ,  $j = 2, 3, \dots, p$ . Then

$$n^{1/2} \text{vec}\{(\hat{\mathbf{B}} - \mathbf{B})'\} \xrightarrow{\mathcal{L}} N(\mathbf{0}, \mathbf{V}_{BB}),$$

where  $\mathbf{V}_{BB} = \Sigma_{vv} \otimes \bar{\mathbf{M}}^{-1}$ ,  $\bar{\mathbf{M}} = E\{\bar{\psi}'_t \bar{\psi}_t\}$ ,

$$\Sigma_{vv} = (\mathbf{I}, -\mathbf{B}_{0,12}) \Sigma_{ee} (\mathbf{I}, -\mathbf{B}_{0,12})'$$

$$\bar{\psi}_t = (\Delta \bar{Y}'_{2,t}, \Delta \mathbf{Y}'_{t-1}, \dots, \Delta \mathbf{Y}'_{t-p+1}),$$

$$\Delta \bar{Y}_t = (\mathbf{H}_1 - \mathbf{I}) \mathbf{Y}_{t-1} + \sum_{j=1}^{p-1} \mathbf{H}_{j+1} \Delta \mathbf{Y}_{t-j},$$

and

$$\hat{\mathbf{B}} - \mathbf{B} = (\hat{\mathbf{B}}_{0,12} - \mathbf{B}_{0,12}, \hat{\mathbf{B}}_{2,1} - \mathbf{B}_{2,1}, \dots, \hat{\mathbf{B}}_{p,1} - \mathbf{B}_{p,1}).$$

**Proof.** We outline the proof for the no-intercept model. The model can be transformed to the canonical form (10.3.24) and the roots of (10.3.41) computed with the canonical form are the same as those computed with the original variables.

Let  $\hat{\mathbf{Y}}_{t-1}$  be the column vector of residuals obtained in the regression of  $\mathbf{Y}_{t-1}$  on  $(\Delta \mathbf{Y}'_{t-1}, \dots, \Delta \mathbf{Y}'_{t-p+1})$ , and let  $s_{ii}$  be the regression residual mean square for the regression of  $\Delta Y_{it}$  on  $(\mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-1}, \dots, \Delta \mathbf{Y}'_{t-p+1})$ . Then, in the set of ordinary regression statistics, the coefficients for  $\mathbf{Y}_{t-1}$  in the regression of  $\Delta Y_{it}$  on  $(\mathbf{Y}'_{t-1}, \Delta \mathbf{Y}'_{t-1}, \dots, \Delta \mathbf{Y}'_{t-p+1})$  are the regression coefficients for the regression of  $\Delta Y_{it}$  on  $\hat{\mathbf{Y}}_{t-1}$  and

$$\mathbf{V}_{11} s_{ii} = \left( \sum_{i=p+1}^n \hat{\mathbf{Y}}_{t-1} \hat{\mathbf{Y}}'_{t-1} \right)^{-1} s_{ii}$$

is the estimated covariance matrix for that vector of coefficients. For the model (10.3.26) in canonical form,

$$\mathbf{D}_n^{-1} \sum_{i=p+1}^n \hat{\mathbf{Y}}_{t-1} \hat{\mathbf{Y}}'_{t-1} \mathbf{D}_n^{-1} \Rightarrow \begin{pmatrix} \mathbf{C}_g \mathbf{G}_{gg} \mathbf{C}'_g & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_{22} \end{pmatrix},$$

where  $\mathbf{C}_g \mathbf{G}_{gg} \mathbf{C}'_g$  is  $g \times g$ ,  $\mathbf{M}_{22}$  is  $(k-g) \times (k-g)$ ,

$$\mathbf{M}_{22} = E\{(\hat{Y}_{g+1,t}, \hat{Y}_{g+2,t}, \dots, \hat{Y}_{kt})' (\hat{Y}_{g+1,t}, \hat{Y}_{g+2,t}, \dots, \hat{Y}_{kt})\},$$

and  $\mathbf{D}_n = \text{diag}(n, n, \dots, n, n^{1/2}, n^{1/2}, \dots, n^{1/2})$ . Also,

$$\begin{aligned} \mathbf{V}_{11}^{1/2} \sum_{t=p+1}^n \hat{\mathbf{Y}}_{t-1} \mathbf{e}'_t &\stackrel{\text{def}}{=} \mathbf{R}_n = \begin{pmatrix} \mathbf{R}_{n11} & \mathbf{R}_{n12} \\ \mathbf{R}_{n21} & \mathbf{R}_{n22} \end{pmatrix} \\ &\Rightarrow \begin{pmatrix} (\mathbf{C}_g \mathbf{G}_{gg} \mathbf{C}'_g)^{-1/2} \mathbf{C}_g \hat{\mathbf{Y}}_{gg}^* & (\mathbf{C}_g \mathbf{G}_{gg} \mathbf{C}'_g)^{-1/2} \mathbf{C}_g \hat{\mathbf{Y}}_{12}^* \\ \Psi_{21} & \Psi_{22} \end{pmatrix}, \end{aligned}$$

where  $\hat{\mathbf{Y}}$  is the limit random variable for  $n^{-1} \sum_{t=1}^n (\Sigma_{j=1}^{t-1} \mathbf{e}_j) \mathbf{e}'_t$ ,  $\hat{\mathbf{Y}}_{gg}^*$  is the upper left  $g \times g$  submatrix of  $\hat{\mathbf{Y}}$ ,  $\hat{\mathbf{Y}}_{12}^*$  is the upper right  $g \times (k-g)$  submatrix of  $\hat{\mathbf{Y}}$ , and the elements of  $\Psi_{ij}$  are normal random variables. See the proof of Theorem 10.3.1. Then

$$(\hat{\mathbf{H}}_1 - \mathbf{I}) \sum_{t=p+1}^n \hat{\mathbf{Y}}_{t-1} \hat{\mathbf{Y}}'_{t-1} (\hat{\mathbf{H}}_1 - \mathbf{I})' = (\mathbf{R}_n + \hat{\mathbf{M}}_n)' (\mathbf{R}_n + \hat{\mathbf{M}}_n),$$

where

$$\hat{\mathbf{M}}_n = n^{1/2} \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{M}_{22}^{-1/2} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ 0 & \mathbf{H}'_{1,22} - \mathbf{I} \end{pmatrix}.$$

The nonsingular matrix  $n^{1/2} \mathbf{M}_{22} (\mathbf{H}'_{1,22} - \mathbf{I})$  dominates  $\mathbf{R}_{n22}$ , which is  $O_p(1)$ . Thus,

$$\begin{aligned} (\mathbf{R}_n + \hat{\mathbf{M}}_n)' (\mathbf{R}_n + \hat{\mathbf{M}}_n) &= \begin{pmatrix} \mathbf{R}'_{n11} \mathbf{R}_{n11} & 0 \\ 0 & 0 \end{pmatrix} + \begin{pmatrix} 0 & \mathbf{R}'_{n21} \\ 0 & \hat{\mathbf{M}}'_{n22} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \mathbf{R}_{n21} & \hat{\mathbf{M}}_{n22} \end{pmatrix} \\ &\quad + \text{remainder}, \end{aligned}$$

where the remainder terms are of small order relative to the included terms. Now  $\hat{\mathbf{M}}'_{n22} \hat{\mathbf{M}}_{n22}$  is a nonsingular matrix multiplied by  $n$ ,  $\mathbf{R}_{n11}$  is  $O_p(1)$ ,  $\mathbf{R}_{n21}$  is  $O_p(1)$ , and the upper left  $g \times g$  portion of  $\hat{\Sigma}_{ee}$  converges to  $\mathbf{I}_g$ . Therefore, the  $g$  smallest roots of (10.3.41) converge to the roots of

$$|\mathbf{Y}_{gg} (\mathbf{C}_g \mathbf{C}'_g \mathbf{G}_{gg} \mathbf{C}'_g)^{-1} \mathbf{C}_g \mathbf{Y}_{gg} - \nu \mathbf{I}_g| = 0.$$

Because the elements of  $\hat{\Sigma}_{ee}$  are small relative to  $\hat{\mathbf{M}}_n$  and the  $g$  smallest roots are  $O_p(1)$ , the limiting behavior of  $\hat{\Sigma}_{\pi\pi}$  of (10.3.47) is that of

$$\begin{pmatrix} 0 & \mathbf{R}'_{n21} \\ 0 & \hat{\mathbf{M}}'_{n22} \end{pmatrix} \begin{pmatrix} 0 & 0 \\ \mathbf{R}_{n21} & \hat{\mathbf{M}}_{n22} \end{pmatrix}.$$

Therefore, the limiting behavior of  $n^{1/2} (\hat{\mathbf{B}}_{0,12} - \mathbf{B}_{0,12})$  is the limiting behavior of

$$\begin{aligned} n^{1/2} (\hat{\mathbf{M}}'_{n22} \hat{\mathbf{M}}_{n22})^{-1} \hat{\mathbf{M}}'_{n22} \mathbf{R}_{n21} &= n^{1/2} \hat{\mathbf{M}}_{n22}^{-1} \mathbf{R}_{n21} \\ &= (\mathbf{H}_{1,22} - \mathbf{I})^{-1} \mathbf{M}_{22}^{-1/2} \Psi_{21}. \end{aligned}$$

In the canonical form,  $\hat{\mathbf{B}}_{0,12}$  is estimating the zero matrix with an error that is

converging to a matrix of normal random variables. Because the maximum likelihood estimators are invariant to linear transformations, the distribution for the estimators of the parameters of the original model are the appropriate transformation of the canonical  $\hat{\mathbf{B}}_{0,12}$ . The limiting distribution of the vector of other elements of  $\hat{\mathbf{B}}$  is obtained by the arguments used in the proof of Theorem 10.3.4. ▲

Table 10.A.6 of Appendix 10.A contains percentiles to be used in testing for unit roots in the characteristic equation (10.3.19). The entries in the table are the percentiles associated with

$$\hat{\lambda}_i^* = (1 + d_f^{-1} \hat{\lambda}_i)^{-1} \hat{\lambda}_i, \quad (10.3.48)$$

where  $\hat{\lambda}_i$ ,  $i = 1, 2, \dots, k$ , are the roots of the determinantal equation (10.3.41), and  $d_f$  is the degrees of freedom for  $\hat{\Sigma}_{ee}$ . Monte Carlo studies conducted by Heon Jin Park (personal communication) indicate that in small samples with  $g > 1$  that statistics based upon  $\hat{\lambda}_i^*$  have power superior to the corresponding test based on  $\hat{\lambda}_i$ .

The first row of Table 10.A.6 is the distribution of the test statistic for testing the hypothesis of a single unit root against the alternative of no unit roots, the second row is for the hypothesis of two unit roots against the alternative of no unit roots, the third row is for the hypothesis of two unit roots against the alternative of one unit root, the fourth row is for the hypothesis of three unit roots against the alternative of no unit roots, etc. The test statistic for row 1 is  $\hat{\lambda}_k$ , the test statistic for row 2 is  $\hat{\lambda}_{k-1} + \hat{\lambda}_k$ , the test statistic for row 3 is  $\hat{\lambda}_k$ , the test statistic for row 4 is  $\hat{\lambda}_{k-2} + \hat{\lambda}_{k-1} + \hat{\lambda}_k$ , the test statistic for row 5 is  $\hat{\lambda}_{k-1} + \hat{\lambda}_k$ , etc. The numbers at the left are the degrees of freedom associated with  $\hat{\Sigma}_{ee}$ .

Table 10.A.7 contains the percentiles of analogous statistics for the model (10.3.20) estimated with an intercept. Table 10.A.8 contains the percentiles of analogous statistics for the model estimated with a time trend.

The tables were constructed by Heon Jin Park using the Monte Carlo method. The limiting percentiles were calculated by simulation using the infinite series representations of Lemma 10.3.1. The percentiles were smoothed using a function of  $d_f$ , where  $d_f$  is the degrees of freedom. The standard errors of the entries in the tables are a function of the size of the entries and range from 0.001 for the smallest entries to 0.25 for the largest entries.

We now discuss the estimation of  $\mathbf{H}_1 - \mathbf{I}$  of equation (10.3.20) subject to the linear constraints associated with the presence of unit roots. Assume that the least squares statistics have been computed, and let  $\hat{\mathbf{H}}_1'$  be the least squares estimator of  $\mathbf{H}_1$ . Let  $\mathbf{V}_{11}$  be the part of the inverse of the matrix of sums of squares and products associated with  $\mathbf{Y}_{t-1}$ . Each column of  $\hat{\mathbf{H}}_1'$  is a vector of regression coefficients and the ordinary least squares estimator of the "covariance matrix" of a column is  $\mathbf{V}_{11}$  multiplied by the error variance. Thus, arguing in a nonrigorous manner, the rows of

$$\mathbf{V}_{11}^{-1/2}(\hat{\mathbf{H}}_1' - \mathbf{H}_1')$$

have the property that they are estimated to be uncorrelated.

Under the hypothesis of  $g$  unit roots in the original system, equivalent to  $g$  zero roots for  $\mathbf{H}_1 - \mathbf{I}$ , there is a  $\mathbf{B}_{0,1}$ , such that

$$\mathbf{V}_{11}^{-1/2}(\mathbf{H}'_1 - \mathbf{I})\mathbf{B}'_{0,1} = \mathbf{0}. \quad (10.3.49)$$

The estimator of  $\mathbf{B}_{0,1}$ , up to a linear transformation, is given by the  $g$  characteristic vectors of (10.3.40) associated with the  $g$  smallest roots. Let

$$\hat{\mathbf{R}}'_1 = \mathbf{V}_{11}^{-1/2}(\hat{\mathbf{H}}'_1 - \mathbf{I})$$

and

$$\mathbf{r}'_1 = \mathbf{V}_{11}^{-1/2}(\mathbf{H}'_1 - \mathbf{I}).$$

Now each column of  $\mathbf{v} = \mathbf{B}_{0,1}\hat{\mathbf{R}}_1$  contains  $g$  linear contrasts that are estimating zero and  $\hat{\mathbf{B}}_{0,1}\hat{\mathbf{R}}_1$  is an estimator of those contrasts. Therefore, an improved estimator of a column of  $\mathbf{r}_1$  is obtained by subtracting from the column of  $\hat{\mathbf{R}}_1$  an estimator of the error in  $\hat{\mathbf{R}}_1$  constructed with the estimator of the  $g$  linear contrasts. Thus, the estimator is

$$\begin{aligned} \tilde{\mathbf{r}}'_{1,i} &= \hat{\mathbf{R}}'_{1,i} - \hat{\mathbf{v}}'_i \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{ve} \\ &= \hat{\mathbf{R}}'_{1,i} (\mathbf{I} - \hat{\mathbf{B}}'_{0,1} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{ve}), \quad i = 1, 2, \dots, k, \end{aligned} \quad (10.3.50)$$

where  $\hat{\mathbf{v}}'_i = \hat{\mathbf{R}}'_{1,i} \hat{\mathbf{B}}'_{0,1}$ ,  $\hat{\Sigma}_{vv} = \hat{\mathbf{B}}_{0,1} \hat{\Sigma}_{ee} \hat{\mathbf{B}}'_{0,1}$ , and  $\hat{\Sigma}_{ve} = \hat{\mathbf{B}}_{0,1} \hat{\Sigma}_{ee}$ . It follows that the estimator of  $\mathbf{H}'_1 - \mathbf{I}$  is

$$\hat{\mathbf{H}}'_1 - \mathbf{I} = \mathbf{V}_{11}^{1/2} \tilde{\mathbf{r}}'_1 = (\hat{\mathbf{H}}'_1 - \mathbf{I})(\mathbf{I} - \hat{\mathbf{B}}'_{0,1} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{ve}). \quad (10.3.51)$$

The estimated covariance matrix of the estimators  $\tilde{\mathbf{r}}_{1,i}$ ,  $i = 1, 2, \dots, k$ , is

$$\hat{\Sigma}_{rr} = \hat{\Sigma}_{ee} - \hat{\Sigma}_{ev} \hat{\Sigma}_{vv}^{-1} \hat{\Sigma}_{ve}.$$

Therefore, the estimated covariance matrix of  $\text{vec}(\hat{\mathbf{H}}_1 - \mathbf{I})$  is

$$\hat{\mathbf{V}}\{\text{vec}(\hat{\mathbf{H}}_1 - \mathbf{I})\} = (\mathbf{V}_{11}^{1/2} \otimes \mathbf{I})(\mathbf{I} \otimes \hat{\Sigma}_{rr})(\mathbf{V}_{11}^{1/2} \otimes \mathbf{I})'.$$

The estimator of  $\mathbf{H}_1 - \mathbf{I}$  is seen to be the least squares estimator adjusted by the estimators of zero, where the coefficient matrix is the matrix of regression coefficients for the regression of the error in  $\hat{\mathbf{H}}_1 - \mathbf{I}$  on the estimators of zero. This formulation can be used to construct the maximum likelihood estimators of the entire model. Let

$$\mathbf{d} = \text{vec}\{(\hat{\mathbf{H}}_1 - \mathbf{I})'(\mathbf{I}, -\hat{\mathbf{B}}_{0,12})'\}$$

be the  $kg$ -dimensional vector of estimators of zero. Then

$$\text{vec}((\hat{\mathbf{H}}_1 - \mathbf{I}, \hat{\mathbf{H}}_2, \dots, \hat{\mathbf{H}}_p)') = \text{vec}((\hat{\mathbf{H}}_1 - \mathbf{I}, \hat{\mathbf{H}}_2, \dots, \hat{\mathbf{H}}_p)' - \mathbf{d}' \hat{\Sigma}_{dd}^{-1} \hat{\Sigma}_{d,\text{vec}\mathbf{H}},$$

where  $\hat{\Sigma}_{dd}$  is the estimated covariance matrix of  $\mathbf{d}$  and  $\hat{\Sigma}_{d,\text{vec}\mathbf{H}}$  is the  $gk \times (p \times k)$  estimated covariance matrix of  $\mathbf{d}$  and  $\text{vec } \hat{\mathbf{H}}'$ . The estimated covariance matrix of the estimators,

$$\hat{\mathbf{V}}\{\text{vec } \hat{\mathbf{H}}'\} = \hat{\Sigma}_{\text{vec } \mathbf{H}, \text{vec } \mathbf{H}} - \hat{\Sigma}_{\text{vec } \mathbf{H}, d} \hat{\Sigma}_{dd}^{-1} \hat{\Sigma}_{d,\text{vec } \mathbf{H}}, \quad (10.3.52)$$

follows from regression theory.

If the original vector autoregressive process is

$$\mathbf{Y}_t = \mathbf{H}_0 + \mathbf{H}_1 \mathbf{Y}_{t-1} + \mathbf{H}_2 \Delta \mathbf{Y}_{t-1} + \dots + \mathbf{H}_p \Delta \mathbf{Y}_{t-p+1} + \mathbf{e}_t, \quad (10.3.53)$$

we may wish to estimate the vector of intercepts  $\mathbf{H}_0$  with the restriction that

$$\mathbf{B}_{0,1} \mathbf{H}_0 = \mathbf{0}. \quad (10.3.54)$$

This is the restriction that the linear combinations of the original variables that are unit root processes are processes with no drift. The estimator of  $\mathbf{B}_{0,1}$  is given by the eigenvectors associated with the smallest roots of

$$|(\hat{\mathbf{H}}_0, \hat{\mathbf{H}}_1 - \mathbf{I}) \mathbf{V}_{**}^{-1} (\hat{\mathbf{H}}_0, \hat{\mathbf{H}}_1 - \mathbf{I})' - \lambda \hat{\Sigma}_{ee}| = 0, \quad (10.3.55)$$

where  $\mathbf{V}_{**}$  is the portion of the estimated covariance matrix of the least squares estimators associated with the vector of regressor variables  $(1, \mathbf{Y}'_{t-1})$ .

In Example 10.3.2 we demonstrate how existing software for simultaneous equation systems can be used to construct the estimators and estimated standard errors of the estimators.

**Example 10.3.2.** To illustrate the testing and estimation methods for multivariate autoregressive time series, we use the data on U.S. interest rates studied by Stock and Watson (1988). The three series are the federal funds rate, denoted by  $Y_{1,t}$ ; the 90-day treasury bill rate, denoted by  $Y_{2,t}$ ; and the one-year treasury bill rate, denoted by  $Y_{3,t}$ . The data are 236 monthly observations for the period, January 1960 through August 1979. The one-year treasury bill interest rate was studied as a univariate series in Examples 10.1.1 and 10.1.2.

Following the procedures of Example 10.1.1, a second order process fit to the differences yields

$$\begin{aligned} \Delta^2 \hat{Y}_{1,t} &= 0.017 - 0.550 \Delta Y_{1,t-1} - 0.161 \Delta^2 Y_{1,t-1}, \\ &\quad (0.025) \quad (0.075) \quad (0.065) \\ \Delta^2 \hat{Y}_{2,t} &= 0.021 - 0.782 \Delta Y_{2,t-1} - 0.025 \Delta^2 Y_{2,t-1}, \\ &\quad (0.021) \quad (0.082) \quad (0.065) \\ \Delta^2 \hat{Y}_{3,t} &= 0.018 - 0.768 \Delta Y_{3,t-1} + 0.105 \Delta^2 Y_{3,t-1}. \\ &\quad (0.019) \quad (0.076) \quad (0.065) \end{aligned} \quad (10.3.56)$$

The  $\hat{\tau}_\mu$  statistics for the hypothesis of a second unit root under the maintained hypothesis of a single unit root are  $-7.37$ ,  $-9.52$ , and  $-10.13$ , respectively. Thus, in all cases, the hypothesis of a second unit root is rejected at the one percent level.

Third order autoregressive models fit by ordinary least squares to each series gives

$$\begin{aligned}\Delta \hat{Y}_{1,t} &= 0.108 - 0.017 Y_{1,t-1} + 0.297 \Delta Y_{1,t-1} + 0.177 \Delta Y_{1,t-2}, \\ &\quad (0.062) \quad (0.010) \quad (0.065) \quad (0.066) \\ \Delta \hat{Y}_{2,t} &= 0.078 - 0.011 Y_{2,t-1} + 0.200 \Delta Y_{2,t-1} + N(0.035 \Delta Y_{2,t-2}), \\ &\quad (0.062) \quad (0.012) \quad (0.066) \quad (0.066) \quad (10.3.57) \\ \Delta \hat{Y}_{3,t} &= 0.082 - 0.012 Y_{3,t-1} + 0.343 \Delta Y_{3,t-1} - 0.094 \Delta Y_{3,t-2}.\end{aligned}$$

$$\quad (0.062) \quad (0.011) \quad (0.065) \quad (0.066)$$

The three residual mean squares are  $0.143$ ,  $0.098$ , and  $0.083$  for the federal funds rate, 90-day treasury rate, and one-year treasury rate, respectively. There are 233 observations in the regression, and the residual mean square has 229 degrees of freedom. The  $\hat{\tau}_\mu$  statistics for the hypothesis of a unit root in the univariate processes are  $-1.62$ ,  $-0.98$ , and  $-1.09$ . The 10% point for  $\hat{\tau}_\mu$  for a sample with 229 degrees of freedom is about  $-2.57$ . Hence, the null hypothesis of a unit root is not rejected for each series. Models were also fit with time as an additional explanatory variable. In all cases, the hypothesis of a unit root was accepted at the 10% level. On the basis of these results, we proceed as if each series were an autoregressive process with a single unit root and other roots less than one in absolute value.

We write the multivariate process as

$$\mathbf{Y}_t = \mathbf{H}_0 + \mathbf{H}_1 \mathbf{Y}_{t-1} + \mathbf{H}_2 \Delta \mathbf{Y}_{t-1} + \mathbf{H}_3 \Delta \mathbf{Y}_{t-2} + \mathbf{e}_t, \quad (10.3.58)$$

or as

$$\Delta \mathbf{Y}_t = \mathbf{H}_0 + (\mathbf{H}_1 - \mathbf{I}) \mathbf{Y}_{t-1} + \mathbf{H}_2 \Delta \mathbf{Y}_{t-1} + \mathbf{H}_3 \Delta \mathbf{Y}_{t-2} + \mathbf{e}_t, \quad (10.3.59)$$

where  $\mathbf{Y}_t = (Y_{1,t}, Y_{2,t}, Y_{3,t})'$  is the column vector of interest rates,  $\mathbf{H}_0$  is a three-dimensional column vector, and  $\mathbf{H}_i$ ,  $i = 1, 2, 3$ , are  $3 \times 3$  matrices of parameters. If the multivariate process has a single unit root, the rank of  $\mathbf{H}_1 - \mathbf{I}$  is 2. The rank of  $\mathbf{H}_1 - \mathbf{I}$  is one if the multivariate process has two unit roots.

To investigate the number of unit roots associated with the multivariate process, we compute the roots of

$$|(\hat{\mathbf{H}}_1 - \mathbf{I}) \mathbf{V}_{11}^{-1} (\hat{\mathbf{H}}_1 - \mathbf{I})' - \lambda \hat{\Sigma}_{ee}| = 0, \quad (10.3.60)$$

where  $\hat{\mathbf{H}}_1$  is the least squares estimate of  $\mathbf{H}_1$ ,  $\hat{\Sigma}_{ee}$  is the least squares estimate of  $\Sigma_{ee}$ , and  $\mathbf{V}_{11} \hat{\sigma}_{eelli}$  is the least squares estimated covariance matrix of the  $i$ th row of  $\hat{\mathbf{H}}_1$ . The matrix  $(\hat{\mathbf{H}}_1 - \mathbf{I}) \mathbf{V}_{11}^{-1} (\hat{\mathbf{H}}_1 - \mathbf{I})'$  can be computed by subtracting the residual

sum of squares and products matrix for the full model from that for the reduced model

$$\Delta \mathbf{Y}_t = \mathbf{H}_0 + \mathbf{H}_2 \Delta \mathbf{Y}_{t-1} + \mathbf{H}_3 \Delta \mathbf{Y}_{t-2} + \mathbf{e}_t. \quad (10.3.61)$$

The full model has nine explanatory variables ( $\mathbf{Y}'_{t-1}$ ,  $\Delta \mathbf{Y}'_{t-1}$ ,  $\Delta \mathbf{Y}'_{t-2}$ ) plus the intercept and is fitted to 233 observations. Hence, the estimator of  $\Sigma_{ee}$  has 223 degrees of freedom. The reduced model contains ( $\Delta \mathbf{Y}_{t-1}$ ,  $\Delta \mathbf{Y}_{t-2}$ ) and the intercept. The matrices are

$$\hat{\mathbf{H}}_1' - \mathbf{I} = \begin{pmatrix} -0.167 & 0.043 & 0.020 \\ (0.043) & (0.034) & (0.033) \\ 0.314 & -0.245 & -0.058 \\ (0.121) & (0.098) & (0.094) \\ -0.109 & 0.189 & 0.019 \\ (0.098) & (0.080) & (0.077) \end{pmatrix},$$

$$\hat{\Sigma}_{ee} = \begin{pmatrix} 0.1304 & 0.0500 & 0.0516 \\ 0.0500 & 0.0856 & 0.0712 \\ 0.0516 & 0.0712 & 0.0788 \end{pmatrix}, \quad (10.3.62)$$

$$(\hat{\mathbf{H}}_1 - \mathbf{I}) \mathbf{V}_{11}^{-1} (\hat{\mathbf{H}}_1 - \mathbf{I})' = \begin{pmatrix} 2.0636 & -0.4693 & -0.1571 \\ -0.4693 & 0.5912 & 0.1752 \\ -0.1571 & 0.1752 & 0.1574 \end{pmatrix}. \quad (10.3.63)$$

The standard errors as output by an ordinary regression program are given in parentheses below the estimated elements of  $(\hat{\mathbf{H}}_1 - \mathbf{I})'$ . The estimates of  $\mathbf{H}_1 - \mathbf{I}$  are not normally distributed in the limit when the  $\mathbf{Y}_t$ -process contains unit roots.

The three roots of (10.3.60) are 1.60, 12.67, and 32.30. The test statistic tabled by Park is

$$\sum_{i=k-g+1}^k (1 + d_f^{-1} \hat{\lambda}_i)^{-1} \hat{\lambda}_i = \sum_{i=k-g+1}^k \hat{\lambda}_i^*,$$

where  $d_f$  is the degrees of freedom for  $\hat{\Sigma}_{ee}$ ,  $\hat{\lambda}_1 \geq \hat{\lambda}_2 \geq \dots \geq \hat{\lambda}_k$ , and  $g$  is the hypothesized number of unit roots in the process. The test statistic for the hypothesis of a single unit root is  $\hat{\lambda}_3^* = 1.59$ . Because the median of the statistic for a single root is 2.43, the hypothesis of one unit root is easily accepted. The test for two unit roots against the alternative of one unit root is based on the second smallest root  $\hat{\lambda}_2^* = 11.99$ . On the basis of this statistic, the hypothesis of two unit roots is accepted at the 10% level because the 10% point for the largest of two  $\hat{\lambda}$  given two unit roots is about 12.5. The hypothesis of three unit roots is rejected because  $\hat{\lambda}_1^* = 28.21$ , exceeds the 1% tabular value for the largest of three  $\hat{\lambda}$  given three unit roots. For illustrative purposes, we estimate the process as a multivariate process with two unit roots and diagonal Jordan canonical form.

The eigenvectors associated with the determinantal equation (10.3.60) are

$$\begin{aligned}\hat{\kappa}_1 &= (0.4374, 0.1042, 3.1472)', \\ \hat{\kappa}_2 &= (1.8061, 5.3600, -6.3725)', \\ \hat{\kappa}_3 &= (2.6368, -4.3171, 1.8279').\end{aligned}\quad (10.3.64)$$

If the vector process contains two unit roots, the two time series defined by

$$(\mathbf{W}_1, \mathbf{W}_2) = (\hat{\kappa}_1' \mathbf{Y}_t, \hat{\kappa}_2' \mathbf{Y}_t), \quad (10.3.65)$$

are estimated to be the unit root processes. Furthermore, the errors in these two processes are estimated to be uncorrelated. Thus, it is estimated that

$$(\mathbf{H}_1 - \mathbf{I})(\hat{\kappa}_1, \hat{\kappa}_2) = (\mathbf{0}, \mathbf{0}) \quad (10.3.66)$$

and that

$$(\hat{\kappa}_1, \hat{\kappa}_2)' \Sigma_{ee} (\hat{\kappa}_1, \hat{\kappa}_2) = \mathbf{I}.$$

The determinantal equation (10.3.44) that can be used to define the cointegrating vectors is

$$\left| \begin{pmatrix} 15.00 & 6.47 & 7.18 \\ 6.47 & 5.26 & 5.50 \\ 7.18 & 5.50 & 6.69 \end{pmatrix} - \nu \begin{pmatrix} 1261.6 & 897.0 & 834.9 \\ 897.0 & 679.4 & 639.1 \\ 834.9 & 639.1 & 615.8 \end{pmatrix} \right| = 0.$$

The roots are 0.1264, 0.0537, and 0.0071, and the associated vectors are

$$\begin{aligned}&(-0.104, 0.313, -0.188)', \\ &(-0.056, -0.106, 0.195)', \\ &(0.004, 0.056, -0.025').\end{aligned}\quad (10.3.67)$$

If we assume that there are two unit roots, then the first vector of (10.3.67) defines the single cointegrating vector.

We note that if (10.3.66) holds,

$$(\mathbf{H}_1 - \mathbf{I})(\hat{\kappa}_1, \hat{\kappa}_2) \mathbf{K}_{11}^{-1} = (\mathbf{0}, \mathbf{0}) \quad (10.3.68)$$

for any nonsingular  $2 \times 2$  matrix  $\mathbf{K}_{11}$ . In particular, we could let

$$\mathbf{K}_{11} = \begin{pmatrix} \hat{\kappa}_{12} & \hat{\kappa}_{22} \\ \hat{\kappa}_{13} & \hat{\kappa}_{23} \end{pmatrix}. \quad (10.3.69)$$

The  $\mathbf{K}_{11}$  of (10.3.69) is nonsingular if  $Y_2$ , and  $Y_3$ , are both unit root processes and are not both functions of a single common unit root process. Therefore, in practice, care should be exercised in the choice of normalization. For our example, we assume that

$$(\mathbf{H}_1 - \mathbf{I})'\boldsymbol{\beta} = \mathbf{0},$$

where

$$\boldsymbol{\beta}' = \begin{pmatrix} -\beta_{21} & 1 & 0 \\ -\beta_{31} & 0 & 1 \end{pmatrix}. \quad (10.3.70)$$

We create a square nonsingular matrix by augmenting  $\boldsymbol{\beta}$  with a column  $(0, 0, 1)'$  to obtain

$$\mathbf{B}' = \begin{pmatrix} 1 & 0 & 0 \\ -\beta_{21} & 1 & 0 \\ -\beta_{31} & 0 & 1 \end{pmatrix}.$$

If we multiply (10.3.59) by  $\mathbf{B}'$ , we obtain the restricted model

$$\begin{aligned} \Delta Y_{1t} &= \theta_{11} Y_{1,t-1} + \theta_{12} Y_{2,t-1} + \theta_{13} Y_{3,t-1} + \mathbf{W}_t \boldsymbol{\theta}_{1w} + u_{1t}, \\ \Delta Y_{2t} &= \beta_{21} \Delta Y_{1t} + \mathbf{W}_t \boldsymbol{\theta}_{2w} + u_{2t}, \\ \Delta Y_{3t} &= \beta_{31} \Delta Y_{1t} + \mathbf{W}_t \boldsymbol{\theta}_{3w} + u_{3t}, \end{aligned} \quad (10.3.71)$$

where  $\mathbf{W}_t = (1, \Delta Y'_{t-1}, \Delta Y'_{t-2})$ . This model is in the form of a set of simultaneous equations. A number of programs are available to estimate the parameters of such models. We use the full information maximum likelihood option of SAS/ETS® to estimate the parameters. The estimates are

$$\begin{aligned} \hat{\Delta Y}_{1t} &= -0.130 Y_{1,t-1} + 0.396 Y_{2,t-1} - 0.238 Y_{3,t-1} + \mathbf{W}_t \hat{\boldsymbol{\theta}}_{1w}, \\ &\quad (0.039) \quad (0.110) \quad (0.080) \\ \hat{\Delta Y}_{2t} &= -0.487 \Delta Y_{1t} + \mathbf{W}_t \hat{\boldsymbol{\theta}}_{2w}, \\ &\quad (0.319) \\ \hat{\Delta Y}_{3t} &= -0.122 \Delta Y_{1t} + \mathbf{W}_t \hat{\boldsymbol{\theta}}_{3w}, \\ &\quad (0.238) \end{aligned} \quad (10.3.72)$$

where we have omitted the numerical values of  $(\hat{\boldsymbol{\theta}}_{1w}, \hat{\boldsymbol{\theta}}_{2w}, \hat{\boldsymbol{\theta}}_{3w})$  to simplify the display. The vector  $(-0.130, 0.396, -0.238)$  applied to  $\mathbf{Y}_t$  defines a process that is estimated to be stationary. The vector is a multiple of the cointegrating vector given as the first vector of (10.3.67).

The estimator of  $\mathbf{H}'_1 - \mathbf{I}$  estimated subject to the restriction that the rank of  $\mathbf{H}_1 - \mathbf{I}$  is one is

$$\hat{\mathbf{H}}'_1 - \mathbf{I} = \begin{pmatrix} -0.130 & 0.063 & 0.016 \\ (0.039) & (0.030) & (0.027) \\ 0.396 & -0.192 & -0.048 \\ (0.110) & (0.091) & (0.087) \\ -0.238 & 0.116 & 0.029 \\ (0.080) & (0.058) & (0.052) \end{pmatrix}. \quad (10.3.73)$$

The standard errors of the estimates are given in parentheses below the estimates. The estimates are not all normally distributed. However, the difference between the estimate and the true value divided by the standard error converges to a  $N(0, 1)$  random variable for a correctly specified model. Therefore, the usual regression testing and confidence interval procedures can be applied to the restricted estimates under the assumption that the process contains two unit roots. The tests are subject to the usual preliminary test bias if one estimates the number of unit roots on the basis of hypothesis tests.

The two time series defined by

$$Y_{2t} + 0.487Y_{1t}$$

and

$$Y_{3t} + 0.122Y_{1t}$$

are processes estimated to have unit roots. These two processes are linear combinations of the two processes in (10.3.65). That is, the two processes are obtained by transforming the first two vectors of (10.3.64),

$$\begin{pmatrix} 0.437 & 1.806 \\ 0.104 & 5.360 \\ 3.147 & -6.372 \end{pmatrix} \begin{pmatrix} 0.104 & 5.360 \\ 3.147 & -6.372 \end{pmatrix}^{-1} = \begin{pmatrix} 0.487 & 0.122 \\ 1 & 0 \\ 0 & 1 \end{pmatrix}. \quad \blacktriangleleft$$

#### 10.4. TESTING FOR A UNIT ROOT IN A MOVING AVERAGE MODEL

The distributional results developed for the autoregressive process with a unit root can be used to test the hypothesis that a moving average process has a unit root. Let the model be

$$Y_t = e_t + \beta e_{t-1}, \quad t = 1, 2, \dots, \quad (10.4.1)$$

where the  $e_t$  are iid( $0, \sigma^2$ ) random variables. We assume  $e_0$  is unknown and to be estimated. We have

$$\begin{aligned} Y_t &= \beta e_0 + e_t, \\ Y_t &= f_t(\mathbf{Y}; \beta, e_0) + e_t, \quad t = 2, 3, \dots, \end{aligned} \quad (10.4.2)$$

where

$$f_t(\mathbf{Y}; \beta, e_0) = -\sum_{j=1}^{t-1} (-\beta)^j Y_{t-j} - (-\beta)^t e_0.$$

On the basis of equation (10.4.2) we define the function  $e_t = e_t(\mathbf{Y}; \beta, e_0)$ , where

$$e_t(\mathbf{Y}; \beta, e_0) = \sum_{j=0}^{t-1} (-\beta)^j Y_{t-j} + (-\beta)^t e_0.$$

Then

$$\frac{\partial}{\partial \beta} e_t(\mathbf{Y}; \beta, e_0) = - \sum_{j=0}^{t-1} j(-\beta)^{j-1} Y_{t-j} - t(-\beta)^{t-1} e_0$$

and

$$\frac{\partial e_t(\mathbf{Y}; \beta, e_0)}{\partial e_0} = (-\beta)^t.$$

If  $\beta = -1$ , we have from (10.4.1)

$$\sum_{j=1}^t Y_j = -e_0 + e_t. \quad (10.4.3)$$

Therefore, treating  $e_0$  as a fixed parameter to be estimated,

$$\hat{e}_0 = -n^{-1} \sum_{t=1}^n \sum_{j=1}^t Y_j \quad (10.4.4)$$

is the best linear unbiased estimator of  $e_0$  when  $\beta = -1$ , and the error in  $\hat{e}_0$  is  $-\bar{e}_n = -n^{-1} \sum_{t=1}^n e_t$ .

We shall construct a test of the hypothesis that  $\beta = -1$  by considering one step of a Gauss–Newton iteration for  $(\beta, e_0)$  with  $(-1, \hat{e}_0)$  as the initial value for  $(\beta, e_0)$ . The dependent variable in such an iteration is

$$e_t(\mathbf{Y}; -1, \hat{e}_0) = \sum_{j=1}^t Y_j + \hat{e}_0. \quad (10.4.5)$$

Under the null hypothesis,

$$e_t(\mathbf{Y}; -1, \hat{e}_0) = e_t - \bar{e}_n.$$

Let  $W_t(\mathbf{Y}; \beta, e_0)$  denote the partial derivative of  $f(\mathbf{Y}; \beta, e_0)$  with respect to  $\beta$ , and note that  $W_t(\mathbf{Y}; \hat{\beta}, \hat{e}_0)$  is the negative of the partial derivative of  $e_t(\mathbf{Y}; \beta, e_0)$  with respect to  $\beta$  evaluated at  $(\hat{\beta}, \hat{e}_0)$ . Then, under the null,

$$W_t(\mathbf{Y}; -1, \hat{e}_0) = -t\bar{e}_n + \sum_{j=0}^{t-1} e_j, \quad t = 2, 3, \dots, n. \quad (10.4.6)$$

The  $W_t(\mathbf{Y}; -1, \hat{e}_0)$  satisfy the recursive relationship

$$W_1(\mathbf{Y}; -1, \hat{e}_0) = \hat{e}_0,$$

$$W_2(\mathbf{Y}; -1, \hat{e}_0) = Y_1 + 2\hat{e}_0,$$

$$W_t(\mathbf{Y}; -1, \hat{e}_0) = Y_{t-1} + 2W_{t-1}(\mathbf{Y}; -1, \hat{e}_0) - W_{t-2}(\mathbf{Y}; -1, \hat{e}_0),$$

$$t = 3, 4, \dots,$$

and the  $e_t(\mathbf{Y}; -1, \hat{e}_0)$  satisfy the recursive relationship

$$\begin{aligned} e_1(\mathbf{Y}; -1, \hat{\epsilon}_0) &= Y_1 + \hat{\epsilon}_0, \\ e_t(\mathbf{Y}; -1, \hat{\epsilon}_0) &= Y_t + e_{t-1}(\mathbf{Y}; -1, \hat{\epsilon}_0), \quad t = 2, 3, \dots. \end{aligned}$$

The Gauss–Newton iteration consists of regressing  $e_t(\mathbf{Y}; -1, \hat{\epsilon}_0)$  on  $W_t(\mathbf{Y}; -1, \hat{\epsilon}_0)$  and a second variable that is identically equal to negative one. The regression equation is

$$e_t(\mathbf{Y}; -1, \hat{\epsilon}_0) = W_t(\mathbf{Y}; -1, \hat{\epsilon}_0) \Delta\beta - \Delta\epsilon_0 + e_t, \quad (10.4.7)$$

and the estimator of  $\Delta\beta$  is

$$\Delta\hat{\beta} = \left[ \sum_{t=1}^n (\hat{W}_t - \bar{w}_n)^2 \right]^{-1} \sum_{t=1}^n (\hat{W}_t - \bar{w}_n) \hat{e}_t,$$

where  $\hat{W}_t = W_t(\mathbf{Y}; -1, \hat{\epsilon}_0)$ ,  $\bar{w}_n = n^{-1} \sum_{t=1}^n \hat{W}_t$ , and  $\hat{e}_t = e_t(\mathbf{Y}; -1, \hat{\epsilon}_0)$ . The pivotal statistic associated with (10.4.7) is

$$\tau^* = \left\{ \left[ \sum_{t=1}^n (\hat{W}_t - \bar{w}_n)^2 \right]^{-1} s^2 \right\}^{-1/2} \Delta\hat{\beta}, \quad (10.4.8)$$

where

$$s^2 = (n-2)^{-1} \sum_{t=1}^n (\hat{e}_t + \Delta\hat{\epsilon}_0 - \hat{W}_t \Delta\hat{\beta})^2$$

and  $\Delta\hat{\epsilon}_0$  is the estimated regression coefficient for the regression equation (10.4.7). The limiting distribution of  $n \Delta\hat{\beta}$  follows from the results of Section 10.1.1.

**Lemma 10.4.1.** Let  $Y_t$  satisfy the model (10.4.1) with  $\beta^0 = -1$ . Then

$$\begin{aligned} n \Delta\hat{\beta} &\xrightarrow{\mathcal{L}} -0.5[G - H^2 - TK + (12)^{-1}T^2]^{-1}, \\ \tau^* &\xrightarrow{\mathcal{L}} -0.5[G - H^2 - TK + (12)^{-1}T^2]^{-1/2}, \end{aligned}$$

where  $\Delta\hat{\beta}$  is defined in (10.4.7),  $G$ ,  $T$ , and  $H$  are defined in Theorem 10.1.3, and  $K$  is defined in Theorem 10.1.6.

**Proof.** Under the null model, the regression coefficient estimating the change in  $\beta$  is

$$\Delta\hat{\beta} = \frac{\sum_{t=1}^n [X_{t-1} - \bar{x}_{(-1)} - (t-\bar{t})\bar{e}_n] e_t}{\sum_{t=1}^n [X_{t-1} - \bar{x}_{(-1)} - (t-\bar{t})\bar{e}_n]^2}, \quad (10.4.9)$$

where  $X_t = \sum_{j=0}^t e_j$  and  $\bar{x}_{(-1)} = n^{-1} \sum_{t=1}^n X_{t-1}$ . Now

$$\sum_{t=1}^n \left( \sum_{j=0}^{t-1} e_j \right) e_t = 0.5 \left( X_n^2 - \sum_{t=0}^n e_t^2 \right),$$

$$-\bar{e}_n \sum_{t=1}^n (t - \bar{t}) e_t = -0.5 n^{-1} (n-1) X_n^2 + n \bar{x}_{(-1)} \bar{e}_n,$$

and

$$\sum_{t=1}^n [X_{t-1} - (t - \bar{t}) \bar{e}_n - \bar{x}_{(-1)}] e_t = -0.5 \sum_{t=1}^n (e_t - \bar{e}_n)^2.$$

Hence,

$$n \Delta \hat{\beta} = \frac{-0.5 n^{-1} \sum_{t=1}^n (e_t - \bar{e}_n)^2}{G_n - H_n^2 - T_n K_n + (12)^{-1} T_n^2}, \quad (10.4.10)$$

where

$$(G_n, T_n, H_n) = \left( n^{-2} \sum_{t=2}^n X_{t-1}^2, n^{-1/2} X_n, n^{-3/2} \sum_{t=2}^n Y_{t-1} \right)$$

and

$$K_n = n^{-5/2} \sum_{j=1}^{n-1} (n-j)(j-1) e_j.$$

The result for  $n \Delta \hat{\beta}$  follows by the arguments of Theorem 10.1.1. Because  $\Delta \hat{\beta}_* = O_p(n^{-1})$  and  $\Delta \hat{e}_0 = O_p(n^{-1/2})$ ,  $s^2$  converges to  $\sigma^2$  and we have the result for  $\tau$ .  $\blacktriangle$

Percentiles for the test statistic  $\tau^*$  are given in Table 10.A.12 of Appendix 10.A. If the true  $\beta^0$  is greater than  $-1$ , the Gauss–Newton procedure should produce a positive  $\Delta \hat{\beta}$  to move the estimate away from  $-1$ . However, under the null model all values of  $\Delta \hat{\beta}$  are negative. Therefore, the null hypothesis is rejected for the least negative of the values, the right tail of Table 10.A.12.

Consider the testing situation in which the null model is

$$Y_t = e_t - e_{t-1}, \quad t = 1, 2, \dots, \quad (10.4.11)$$

and the alternative model is

$$Y_t = \mu + \beta e_{t-1} + e_t, \quad t = 1, 2, \dots, \quad (10.4.12)$$

where  $|\beta| < 1$ . Under the alternative model

$$Y_t = f_t(\mathbf{Y}; \mu, \beta, e_0) + e_t, \quad (10.4.13)$$

where

$$f_t(\mathbf{Y}; \mu, \beta, e_0) = \mu \sum_{j=0}^{t-1} (-\beta)^j - \sum_{j=1}^{t-1} (-\beta)^j Y_{t-j} - (-\beta)^t e_0$$

for  $t = 2, 3, \dots$ . The partial derivatives are such that one step of a Gauss–Newton iteration for  $(\mu, \beta, e_0)$  with  $(0, -1, \hat{e}_0)$  as the initial value is given by the regression of  $e_t(\mathbf{Y}; 0, -1, \hat{e}_0)$  defined in (10.4.5) on  $(t, W_t, -1)$ , where  $\hat{e}_0$  is defined in (10.4.4) and  $W_t = W_t(\mathbf{Y}; 0, -1, \hat{e}_0)$  is defined in (10.4.6). Under the null, the distribution of the regression pivotal for  $W_t(\mathbf{Y}; 0, -1, \hat{e}_0)$  is that tabulated as  $\hat{\tau}_r$  in Table 10.A.2.

**Lemma 10.4.2.** Let the model (10.4.1) hold with  $\beta = -1$ . Let  $\Delta\hat{\beta}$  be the coefficient of  $W_t$  in the regression of  $e_t(\mathbf{Y}; 0, -1, \hat{e}_0)$  on  $(t, W_t, -1)$ , and let  $\hat{\tau}_r$  be the corresponding regression pivotal. Then

$$\hat{\tau}_r \xrightarrow{\mathcal{L}} 0.5(G - H^2 - 3K^2)^{-1/2}[(T - 2H)(T - 6K) - 1],$$

where  $G$ ,  $T$ ,  $H$ , and  $K$  are defined in Theorem 10.1.6.

**Proof.** Under the null model

$$\Delta\hat{\beta} = \frac{\sum_{t=1}^n [X_{t-1} - \bar{x}_{(-1)} - \hat{b}(t - \bar{t})]e_t}{\sum_{t=1}^n [X_{t-1} - \bar{x}_{(-1)} - \hat{b}(t - \bar{t})]^2},$$

where

$$\hat{b} = \left[ \sum_{t=1}^n (t - \bar{t})^2 \right]^{-1} \sum_{t=1}^n (t - \bar{t}) X_{t-1}$$

and  $X_t = \sum_{j=0}^t e_j$ . This is the same as the expression for  $\hat{\theta}_1$  in the first order model (10.1.34), and the result follows from Theorem 10.1.6.  $\blacktriangle$

A testing situation in which the null model contains a constant term is also of interest. The null model is

$$Y_t = \mu + e_t - e_{t-1}, \quad t = 1, 2, \dots, \quad (10.4.14)$$

and the alternative model is (10.4.12) with  $|\beta| < 1$ . The test statistic is constructed in a similar manner to that for the null model (10.4.11). The initial values for  $\mu$  and  $e_0$  are

$$\tilde{\mu} = \left[ \sum_{t=1}^n (t - \bar{t})^2 \right]^{-1} \sum_{t=1}^n (t - \bar{t}) \sum_{j=1}^t Y_j,$$

and

$$\tilde{e}_0 = -n^{-1} \sum_{t=1}^n \sum_{j=1}^t Y_j + 0.5\bar{\mu}(n+1).$$

The value for  $\tilde{e}_1$  is  $Y_1 + \tilde{e}_0 - \bar{\mu}$ , we have  $\tilde{e}_t = \sum_{j=1}^t Y_j + \tilde{e}_0 - t\bar{\mu}$  for  $t = 2, 3, \dots, n$ , and

$$W_t(Y; \bar{\mu}, -1, \tilde{e}_0) = -0.5t(t-1)\bar{\mu} + \sum_{j=1}^{t-1} jY_{t-j} + t\tilde{e}_0.$$

The test statistic  $\hat{\tau}_{\mu}$  is the regression pivotal for  $\Delta\beta$  associated with the regression equation

$$e_t(Y; \bar{\mu}, -1, \tilde{e}_0) = (t-1)\Delta\mu + W_t(Y; \bar{\mu}, -1, \tilde{e}_0) \Delta\beta - \Delta e_0 + e_t. \quad (10.4.15)$$

The limiting distribution of the test statistic is given in Lemma 10.4.3. The distribution of the test statistic is tabled in the second part of Table 10.A.12.

**Lemma 10.4.3.** Let the model (10.4.12) hold with  $\beta = -1$ . Let  $\Delta\hat{\beta}$  be the regression coefficient for  $W_t(Y; \bar{\mu}, -1, \tilde{e}_0)$  in the regression (10.4.15), and let  $\hat{\tau}_{\mu}$  be the corresponding regression pivotal. Then

$$(n \Delta\hat{\beta}, \hat{\tau}_{\mu}) \xrightarrow{\mathcal{L}} -0.5(D_{\tau}^{-1}, D_{\tau}^{-1/2}),$$

where  $D_{\tau} = G - H^2 - 3K^2 + 0.5(T-2H)^2 - (T-2H)(12L-2H-3K)$ ,

$$L = \int_0^1 \int_0^t \int_0^s W(r) dr ds dt,$$

$G$ ,  $H$ , and  $K$  are defined in Theorem 10.1.6, and  $W(r)$  is the standard Wiener process.

**Proof.** Omitted. See Arellano and Pantula (1995). ▲

The testing procedure can be extended to the autoregressive moving average process. Let the model be

$$Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} = e_t + \sum_{i=1}^q \beta_i e_{t-i} \quad (10.4.16)$$

or

$$Y_t + \sum_{j=1}^p \alpha_j Y_{t-j} = e_t + \zeta_1 e_{t-1} + \sum_{i=2}^q \zeta_i (e_{t-i+1} - e_{t-i}),$$

where the roots of

$$m^p + \sum_{j=1}^p \alpha_j m^{p-j} = 0$$

are less than one in absolute value,  $\zeta_1 = \sum_{i=1}^q \beta_i$ , and  $\zeta_j = -\sum_{i=j}^p \beta_i$ ,  $j = 2, 3, \dots, q$ . Assume it is desired to test  $\zeta_1 = -1$ . The derivatives of the function with respect to the autoregressive moving average parameters are given in (8.4.9) of Section 8.4. If we set  $\zeta_1 = -1$  in the original model, we have

$$X_t + \sum_{j=1}^p \alpha_j X_{t-j} = \kappa + \sum_{i=2}^q \zeta_i e_{t-i+1} + e,$$

for  $t = p + 1, p + 2, \dots$ , where  $X_t = \sum_{j=1}^q Y_j$ , and

$$\kappa = -e_0 - \sum_{i=2}^q \zeta_i e_{1-i}.$$

Therefore, treating  $\kappa$  as fixed,  $X_t$  is a stationary process (or is converging to a stationary process) with mean  $\kappa$ . If the original model contains a mean  $\mu$ , then the model for  $X_t$  contains the mean function  $\kappa_0 + \kappa_1 t$ . The testing procedure consists of the following steps:

1. Create the variable  $X_t = \sum_{j=1}^q Y_j$ . If the null model is of the form (10.4.16), estimate  $\theta = (\alpha_1, \alpha_2, \dots, \alpha_p, \zeta_2, \dots, \zeta_q)$  and  $\kappa$  by any of the procedures of Section 8.4. If the null model contains an unknown mean, estimate  $\theta$  and  $(\kappa_0, \kappa_1)$  by any of the procedures of Section 8.4. Let  $\tilde{e}_t = e_t(Y_t; \tilde{\theta}', \tilde{\kappa}, \tilde{\zeta})$  be the residuals from this fitting operation for (10.4.16).
2. For the model (10.4.16), evaluate the derivatives at  $(\theta', \kappa, \zeta_1) = (\tilde{\theta}', \tilde{\kappa}, -1)$  and carry out one step of the Gauss–Newton procedure with  $\tilde{e}_t$  as the dependent variable. The test of  $\zeta_1 = -1$  is a test for a unit root of the moving average part of the process. The critical values for the test are the  $\tau^*$  of Table 10.A.9. If the null model contains an unknown mean,  $(\tilde{\kappa}_0, \tilde{\kappa}_1)$  is a part of the model and changes in these two parameters are estimated as part of the Gauss–Newton step. Critical values for the test of the unknown-mean null model based on  $\zeta_1$  are the values given as  $\tau_\mu^*$  in Table 10.A.9.

Tanaka (1990), Tsay (1993), Saikkonen and Luukkonen (1993), and Breitung (1994) have discussed alternative testing procedures.

**Example 10.4.1.** To illustrate testing for a moving average unit root we use computer generated data. The data are given in Table 10.B.3. The model for the data is

$$\begin{aligned}
 Y_t - \mu + \alpha_1(Y_{t-1} - \mu) + \alpha_2(Y_{t-2} - \mu) &= e_t + \beta_1 e_{t-1} + \beta_2 e_{t-2} \\
 &= e_t + \zeta_1 e_{t-1} + \zeta_2 (e_{t-1} - e_{t-2}),
 \end{aligned} \tag{10.4.17}$$

where the  $e_t$  are  $N(0, \sigma^2)$  random variables,  $\zeta_1 = \beta_1 + \beta_2$ , and  $\zeta_2 = -\beta_2$ . The maximum likelihood estimates of the parameters are

$$(\hat{\mu}, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\beta}_1, \hat{\beta}_2) = (-0.021, -1.375, 0.601, -0.471, -0.401). \\
 (0.060) \quad (0.100) \quad (0.099) \quad (0.118) \quad (0.117)$$

The roots of the moving average polynomial are 0.911 and -0.440. To test the hypothesis that one of the roots of the moving average polynomial is one, we create the variable  $X_t = \sum_{j=1}^t Y_j$  and estimate the parameters of the model

$$\begin{aligned}
 X_t &= \kappa_0 + \kappa_1 t + u_t, \\
 u_t + \alpha_1 u_{t-1} + \alpha_2 u_{t-2} &= e_t + \zeta_2 e_{t-1},
 \end{aligned} \tag{10.4.18}$$

where  $\kappa_0 = -e_0 - \zeta_2 e_{-1}$  and  $\kappa_1 = \mu(1 + \alpha_1 + \alpha_2)$ , by maximum likelihood. The estimates are

$$(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\zeta}_2, \tilde{\kappa}_0, \tilde{\kappa}_1) = (-1.385, 0.573, 0.481, 0.532, -0.009). \\
 (0.100) \quad (0.098) \quad (0.110) \quad (1.427) \quad (0.024)$$

We construct a test for the null model of the form (10.4.17). The dependent variable for the Gauss–Newton step is the residual  $\tilde{e}_t$  obtained from the fit of model (10.4.18). We write the regression equation as

$$\begin{aligned}
 \tilde{e}_t &= \Delta \kappa_0 + (t-2) \Delta \kappa_1 + \Delta \alpha_1 Q_{\alpha_1,t} + \Delta \alpha_2 Q_{\alpha_2,t} \\
 &\quad + \Delta \zeta_1 Q_{\zeta_1,t} + \Delta \zeta_2 Q_{\zeta_2,t} + e_t.
 \end{aligned} \tag{10.4.19}$$

The derivatives defining the explanatory variables are given by the recursive equations

$$\begin{aligned}
 Q_{\alpha_1,t} &= -Y_{t-1} + Q_{\alpha_1,t-1} - \tilde{\zeta}_2 (Q_{\alpha_1,t-1} - Q_{\alpha_1,t-2}) \\
 &= -Y_1, \quad t = 2, \\
 Q_{\alpha_2,t} &= -Y_{t-2} + Q_{\alpha_2,t-1} - \tilde{\zeta}_2 (Q_{\alpha_2,t-1} - Q_{\alpha_2,t-2}), \\
 Q_{\zeta_1,t} &= \tilde{e}_{t-1} + Q_{\zeta_1,t-1} - \tilde{\zeta}_2 (Q_{\zeta_1,t-1} - Q_{\zeta_1,t-2}), \\
 Q_{\zeta_2,t} &= \tilde{e}_{t-1} - \tilde{e}_{t-2} + Q_{\zeta_2,t-1} - \tilde{\zeta}_2 (Q_{\zeta_2,t-1} - Q_{\zeta_2,t-2}),
 \end{aligned}$$

where the equations hold for  $t \geq 3$ , and it is understood that the values are zero for  $t = 1$  and  $t = 2$  unless otherwise specified.

Regressing  $\tilde{e}_t$  on the derivatives in a regression with 98 observations, we obtain the estimated regression equation,

$$\tilde{e}_t = 0.23 - 0.003(t-2) - 0.011Q_{\alpha_1,t} - 0.016Q_{\alpha_2,t} - 0.081Q_{\zeta_1,t} + 0.058Q_{\zeta_2,t}. \quad (10.4.17)$$

The test that the moving average contains a unit root is

$$\hat{\tau}_\mu^* = -(0.051)^{-1}(0.081) = -1.59.$$

The hypothesis of a unit root is just accepted at the 5% level, because  $-1.59$  is just less than the tabular value of  $-1.563$ . The hypothesis is rejected at the 10% level because the calculated value is greater than the 10% tabular value of  $-1.711$ .

If the null model is (10.4.17) with  $(\zeta_1, \mu) = (-1, 0)$ , and the alternative is the model (10.4.7), we estimate the model in  $X_t$  that contains only  $\kappa_0$ . The maximum likelihood estimates are

$$(\tilde{\alpha}_1, \tilde{\alpha}_2, \tilde{\zeta}_2, \tilde{\kappa}_0) = (-1.385, 0.572, 0.483, 0.074). \quad (10.4.20)$$

The derivatives for  $(\alpha_1, \alpha_2, \zeta_1, \zeta_2)$  are found in the same manner as for the alternative model with a mean. One step of the Gauss–Newton procedure is computed with the intercept but no time variable. The result of the Gauss–Newton step is

$$\tilde{e}_t = -0.11 - 0.002Q_{\alpha_1,t} - 0.016Q_{\alpha_2,t} - 0.059Q_{\zeta_1,t} + 0.051Q_{\zeta_2,t}. \quad (10.4.17)$$

and the test statistic is  $\hat{\tau}^* = -1.40$ . In this case, the hypothesis of a moving average unit root is accepted at the 10% level, because the calculated value is less than the tabular value of  $-1.282$ .

If the null is (10.4.17) with  $(\zeta_1, \mu) = (-1, 0)$  and the alternative is the model (10.4.17), we use the residuals associated with the estimates (10.4.20) but include time in the Gauss–Newton step. The result is

$$\begin{aligned} \tilde{e}_t = & 0.43 - 0.006(t-2) - 0.009Q_{\alpha_1,t} \\ & (0.36) \quad (0.004) \quad (0.102) \\ & - 0.018Q_{\alpha_2,t} - 0.102Q_{\zeta_1,t} + 0.070Q_{\zeta_2,t}, \\ & (0.100) \quad (0.053) \quad (0.113) \end{aligned}$$

and the test statistic is  $\hat{\tau}_\mu^* = -1.91$ . Because the 10% critical value of the test obtained from Table 10.A.2 is  $-1.22$ , the null hypothesis is accepted at the 10% level. ▲▲

## REFERENCES

- Section 10.1.** Bhargava (1986), Dickey (1976), Dickey and Fuller (1979, 1981), Dickey, Hasza, and Fuller (1984), Dickey and Said (1982), Elliott, Rothenberg, and Stock (1992), Evans and Savin (1981a,b), Gonzalez-Farias and Dickey (1992), Lai and Wei (1985a,b), Pantula (1982, 1988a), Pantula, Gonzalez-Farias, and Fuller (1994), Phillips (1987a–c), Phillips and Perron (1988), Rao (1961), Reeves (1972), Sargan and Bhargava (1983), Stigum (1974), White (1958).
- Section 10.2.** Anderson (1959), Basawa (1987), Fuller and Hasza (1980, 1981), Hasza (1977), Rao (1961, 1978a,b), Rubin (1950), Stigum (1974), Venkataraman (1967, 1973), White (1958, 1959).
- Section 10.3.** Ahn and Reinsel (1990), Chan and Wei (1988), Fountis and Dickey (1989), Fuller, Hasza, and Goebel (1981), Johansen (1988), Johansen and Juselius (1990), Lai and Wei (1985a,b), Nagaraj and Fuller (1991), Phillips (1987a–c, 1988, 1991), Phillips and Durlauf (1986), Stock and Watson (1988).
- Section 10.4.** Arellano (1992), Arellano and Pantula (1995), Chang (1989), Tsay (1993).

## EXERCISES

**1. Prove**

$$\lim_{n \rightarrow \infty} \left[ \sum_{t=1}^n \left( \sum_{j=1}^t e_j \right)^2 \right]^{-1} \sum_{t=1}^n \left( X_0 + \sum_{j=1}^t e_j \right)^2 = 1$$

for fixed  $X_0$ , where the  $e_t$  are iid( $0, \sigma^2$ ).

**2. Let the model (10.1.1) hold with  $\rho = 1$ .**

(a) Show that

$$n \left[ \left( \sum_{t=1}^n Y_t^2 \right)^{-1} \sum_{t=2}^n Y_{t-1} Y_t - 1 \right] \xrightarrow{\mathcal{L}} -(2G)^{-1}(T^2 + 1).$$

(b) Show that

$$n \left\{ \left[ \sum_{t=2}^n Y_{t-1}^2 \sum_{t=2}^n Y_t^2 \right]^{-1/2} \sum_{t=2}^n Y_{t-1} Y_t - 1 \right\} \xrightarrow{\mathcal{L}} -(2G)^{-1}.$$

**3. Let  $Y_t$  be the  $p$ th order autoregressive process with a unit root considered in Theorem 10.1.2. Let  $\hat{m}_1$  be the largest root of the characteristic equation associated with the ordinary least squares estimator. Show that**

$$n \left\{ \left[ \left( \sum_{t=2}^n Y_t^2 \right)^{-1} \sum_{t=2}^n Y_{t-1}^2 \right]^{1/2} \hat{m}_1 - 1 \right\} \xrightarrow{\mathcal{L}} -(2G)^{-1},$$

where  $G$  is defined in Theorem 10.1.1.

4. Assume that  $Y_t = Y_{t-1} + e_t$ ,  $t = 1, 2, \dots$ , where  $Y_0 = 0$ , and  $e_t \sim NI(0, \sigma^2)$ . Show that

$$\text{Var}\{\bar{y}_{(-1)}\} = [6(n-1)]^{-1}n(2n-1)\sigma^2,$$

$$\text{Cov}\{\bar{y}_{(-1)}, \bar{e}_{(0)}\} = 0.5(n-1)^{-1}(n-2)\sigma^2,$$

$$E\left\{\sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})e_t\right\} = -0.5(n-2)\sigma^2,$$

$$\text{Var}\left\{\sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})e_t\right\} = (12)^{-1}(n+6)(n-2)\sigma^4,$$

$$E\left\{\sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})^2\right\} = 6^{-1}n(n-2)\sigma^2,$$

$$\text{Var}\left\{\sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})^2\right\} = (90)^{-1}n(n-2)(2n^2 - 4n + 9)\sigma^4,$$

$$\begin{aligned} \text{Cov}\left\{\sum_{t=2}^n (e_t - \bar{e}_{(0)})(Y_{t-1} - \bar{y}_{(-1)}), \sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})^2\right\} \\ = -(12)^{-1}n(n-2)(2n+1)\sigma^4, \end{aligned}$$

$$\text{Cov}\left\{\sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})(Y_t - \bar{y}_{(0)}), \sum_{t=2}^n (Y_{t-1} - \bar{y}_{(-1)})^2\right\} = 6^{-1}n(n-2)\sigma^4.$$

5. Assume that the time series  $Y_t$  satisfies the equation

$$Y_t = \rho Y_{t-1} + e_t, \quad t = 1, 2, \dots,$$

where  $Y_0 = 0$  and  $e_t \sim NI(0, 1)$ . Let  $\hat{\rho} = \hat{\rho}_t$  if  $\hat{\rho}_t \leq 1$  and  $\hat{\rho} = 1$  if  $\hat{\rho}_t > 1$ , where

$$\hat{\rho}_t = \left(\sum_{i=2}^{t-1} Y_i^2\right)^{-1} \sum_{i=2}^t Y_{i-1}Y_i.$$

Let  $\hat{\rho}_m$  be the value of  $\rho$  that maximizes (10.1.47). Show that the limiting distribution of  $n(\hat{\rho}^* - 1)$  has a smaller second moment about zero than the limiting distribution of  $n(\hat{\rho}_m - 1)$  when  $\rho^0 = 1$ .

6. Show that, under normality, the likelihood ratio statistic for testing  $H_0: \alpha_1 = \alpha_1^0$  against  $H_A: \alpha_1 \neq \alpha_1^0$  for the model (10.2.6) is a function of  $\hat{\tau}$  defined in Corollary 10.2.1.2.
7. Plot the residuals from the model of Example 10.1.1 against time. Do you believe the original errors are identically distributed over time? Do you believe the original errors are normally distributed? Fit autoregressive models of orders 2, 3, and 4 to the square roots of the original observations. Plot the residuals from the second order model.

8. Let the assumptions of Theorem 10.3.1 hold for the model

$$Y_t = \beta_0 + \beta' X_{t-1} + \theta_1 Y_{t-1} + e_t.$$

Show that

$$t_{1\mu} \xrightarrow{\mathcal{L}} \rho_{ue} \hat{\tau}_{\mu l} + (1 - \rho_{ue}^2)^{1/2} d,$$

where

$$t_{1\mu} = [\hat{V}\{\hat{\theta}_1\}]^{-1/2}(\hat{\theta}_1 - 1),$$

$\hat{V}\{\hat{\theta}_1\}$  is the ordinary least squares estimator of the variance of the ordinary least squares estimator  $\hat{\theta}_1$ ,  $\hat{\tau}_{\mu l}$  has the distribution of Table 10.A.2,  $\hat{\tau}_{\mu l} \xrightarrow{\mathcal{L}} \hat{\tau}_{\mu l}$ , and  $d$  is a  $N(0, 1)$  random variable independent of  $\hat{\tau}_{\mu l}$ .

9. Use the facts that  $\hat{K}' \hat{A} \hat{K} = \hat{S}_{hh}$  and  $\hat{K}' \Sigma_{ee} \hat{K} = I$ , where  $\hat{A} = \text{diag}(\hat{\lambda}_1, \hat{\lambda}_2, \dots, \hat{\lambda}_k)$ , to show that the expression (10.3.47) for  $\hat{B}_{0,12}'$  is equivalent to that in (10.3.37) when  $g = 1$ .

10. Let the following model hold:

$$\begin{aligned} Y_{1t} &= \pi_1 Y_{2,t-1} + e_{1t}, \\ Y_{2t} &= Y_{2,t-1} + e_{2t}, \end{aligned}$$

for  $t = 2, 3, \dots$ , where  $Y_{20} = 0$  and  $(e_{1t}, e_{2t})' \sim NI(\mathbf{0}, \Sigma)$ . Show that the maximum likelihood estimator of  $\pi_1$  is given by the coefficient of  $Y_{2,t-1}$  in the regression of  $Y_{1t}$  on  $(Y_{2,t-1}, Y_{2t} - Y_{2,t-1})$ . What is the coefficient of  $Y_{2t} - Y_{2,t-1}$  estimating? Show that the limiting distribution of the usual regression  $t$ -statistic for the coefficient of  $Y_{2,t-1}$  in the multiple regression is that of a  $N(0, 1)$  random variable. See Phillips (1991).

11. Let model (10.1.25) hold with  $\theta_1 = 1$  and assume  $e_t \sim NI(0, \sigma^2)$ . Show that  $\Delta \bar{Y}$  of (10.1.26) is uncorrelated with  $\hat{\theta}_1 - 1$  but that  $n^{1/2} \Delta \bar{Y}$  is not independent of  $n(\hat{\theta}_1 - 1)$  in the limit.

12. Assume that  $(Z_t, Y_t)$  of Table 10.B.4 satisfy the model

$$\begin{aligned} Z_t &= \beta_{30} + \beta_{31} Y_t + e_{3t}, \\ Y_t &= \pi_{20} + \pi_{21} Y_{t-1} + \pi_{22} \Delta Y_{t-1} + e_{2t}, \end{aligned}$$

where  $(e_{3t}, e_{2t})' \sim NI(\mathbf{0}, \Sigma_{ee})$ ,  $\pi_{21} \in (-1, 1]$ , and  $\pi_{20} = 0$  if  $\pi_{21} = 1$ . Estimate the parameters of the model by ordinary least squares and by the system method of Example 10.3.1. Also estimate the parameters subject to the restriction that  $(\pi_{20}, \pi_{21}) = (0, 1)$ .

## APPENDIX A

## Appendix 10.A. Percentiles for Unit Root Distributions

**Table 10.A.1. Empirical Cumulative Distribution of  $n(\hat{\rho} - 1)$  for  $\rho = 1$** 

Sample Size <i>n</i>	Probability of a Smaller Value								
	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
$\hat{\rho}$									
25	-11.8	-9.3	-7.3	-5.3	-0.82	1.01	1.41	1.78	2.28
50	-12.8	-9.9	-7.7	-5.5	-0.84	0.97	1.34	1.69	2.16
100	-13.3	-10.2	-7.9	-5.6	-0.85	0.95	1.31	1.65	2.09
250	-13.6	-10.4	-8.0	-5.7	-0.86	0.94	1.29	1.62	2.05
500	-13.7	-10.4	-8.0	-5.7	-0.86	0.93	1.29	1.61	2.04
$\infty$	-13.7	-10.5	-8.1	-5.7	-0.86	0.93	1.28	1.60	2.03
$\hat{\rho}_\mu$									
25	-17.2	-14.6	-12.5	-10.2	-4.22	-0.76	0.00	0.64	1.39
50	-18.9	-15.7	-13.3	-10.7	-4.29	-0.81	-0.07	0.53	1.22
100	-19.8	-16.3	-13.7	-11.0	-4.32	-0.83	-0.11	0.47	1.13
250	-20.3	-16.7	-13.9	-11.1	-4.34	-0.84	-0.13	0.44	1.08
500	-20.5	-16.8	-14.0	-11.2	-4.35	-0.85	-0.14	0.42	1.07
$\infty$	-20.6	-16.9	-14.1	-11.3	-4.36	-0.85	-0.14	0.41	1.05
$\hat{\rho}_\tau$									
25	-22.5	-20.0	-17.9	-15.6	-8.49	-3.65	-2.51	-1.53	-0.46
50	-25.8	-22.4	-19.7	-16.8	-8.80	-3.71	-2.60	-1.67	-0.67
100	-27.4	-23.7	-20.6	-17.5	-8.96	-3.74	-2.63	-1.74	-0.76
250	-28.5	-24.4	-21.3	-17.9	-9.05	-3.76	-2.65	-1.79	-0.83
500	-28.9	-24.7	-21.5	-18.1	-9.08	-3.76	-2.66	-1.80	-0.86
$\infty$	-29.4	-25.0	-21.7	-18.3	-9.11	-3.77	-2.67	-1.81	-0.88

NOTE. This table was constructed by David A. Dickey using the Monte Carlo method. Details are given in Dickey (1976). Standard errors of the estimates vary, but most are less than 0.10 for entries in the left half of the table and less than 0.02 for entries in the right half of the table. Some entries differ slightly from those of the first edition.

**Table 10.A.2. Empirical Cumulative Distribution of  $\hat{\tau}$  for  $\rho = 1$** 

Sample Size <i>n</i>	Probability of a Smaller Value								
	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
$\hat{\tau}$									
25	-2.65	-2.26	-1.95	-1.60	-0.47	0.92	1.33	1.70	2.15
50	-2.62	-2.25	-1.95	-1.61	-0.49	0.91	1.31	1.66	2.08
100	-2.60	-2.24	-1.95	-1.61	-0.50	0.90	1.29	1.64	2.04
250	-2.58	-2.24	-1.95	-1.62	-0.50	0.89	1.28	1.63	2.02
500	-2.58	-2.23	-1.95	-1.62	-0.50	0.89	1.28	1.62	2.01
$\infty$	-2.58	-2.23	-1.95	-1.62	-0.51	0.89	1.28	1.62	2.01
$\hat{\tau}_\mu$									
25	-3.75	-3.33	-2.99	-2.64	-1.53	-0.37	0.00	0.34	0.71
50	-3.59	-3.23	-2.93	-2.60	-1.55	-0.41	-0.04	0.28	0.66
100	-3.50	-3.17	-2.90	-2.59	-1.56	-0.42	-0.06	0.26	0.63
250	-3.45	-3.14	-2.88	-2.58	-1.56	-0.42	-0.07	0.24	0.62
500	-3.44	-3.13	-2.87	-2.57	-1.57	-0.44	-0.07	0.24	0.61
$\infty$	-3.42	-3.12	-2.86	-2.57	-1.57	-0.44	-0.08	0.23	0.60
$\hat{\tau}_r$									
25	-4.38	-3.95	-3.60	-3.24	-2.14	-1.14	-0.81	-0.50	-0.15
50	-4.16	-3.80	-3.50	-3.18	-2.16	-1.19	-0.87	-0.58	-0.24
100	-4.05	-3.73	-3.45	-3.15	-2.17	-1.22	-0.90	-0.62	-0.28
250	-3.98	-3.69	-3.42	-3.13	-2.18	-1.23	-0.92	-0.64	-0.31
500	-3.97	-3.67	-3.42	-3.13	-2.18	-1.24	-0.93	-0.65	-0.32
$\infty$	-3.96	-3.67	-3.41	-3.13	-2.18	-1.25	-0.94	-0.66	-0.32

NOTE. This table was constructed by David A. Dickey using the Monte Carlo method. Details are given in Dickey (1976). Standard errors of the estimates vary, but most are less than 0.014. Some entries differ slightly from those of the first edition.

**Table 10.A.3. Cumulative Distribution of Simple Symmetric  $\hat{\tau}_t$  for  $\rho = 1$** 

Sample Size <i>n</i>	Probability of a Smaller Value								
	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
<b>No Adjustment</b>									
25	-2.72	-2.35	-2.05	-1.74	-0.87	-0.43	-0.37	-0.33	-0.29
50	-2.71	-2.36	-2.08	-1.78	-0.90	-0.44	-0.38	-0.33	-0.29
100	-2.70	-2.37	-2.09	-1.79	-0.91	-0.45	-0.38	-0.34	-0.30
250	-2.70	-2.37	-2.10	-1.80	-0.92	-0.46	-0.39	-0.34	-0.30
500	-2.70	-2.37	-2.10	-1.80	-0.92	-0.46	-0.39	-0.34	-0.30
$\infty$	-2.69	-2.37	-2.10	-1.81	-0.93	-0.46	-0.39	-0.34	-0.30
<b>Mean Removed</b>									
25	-3.40	-3.02	-2.71	-2.37	-1.42	-0.83	-0.73	-0.65	-0.59
50	-3.28	-2.94	-2.66	-2.35	-1.44	-0.84	-0.73	-0.65	-0.58
100	-3.23	-2.90	-2.64	-2.34	-1.44	-0.84	-0.73	-0.65	-0.58
250	-3.20	-2.88	-2.62	-2.34	-1.45	-0.85	-0.73	-0.66	-0.58
500	-3.19	-2.88	-2.62	-2.33	-1.45	-0.85	-0.73	-0.66	-0.58
$\infty$	-3.17	-2.87	-2.62	-2.33	-1.45	-0.85	-0.73	-0.66	-0.58
<b>Linear Trend Removed</b>									
25	-4.19	-3.76	-3.45	-3.09	-2.10	-1.42	-1.28	-1.18	-1.07
50	-3.99	-3.36	-3.36	-3.04	-2.11	-1.44	-1.29	-1.18	-1.07
100	-3.89	-3.57	-3.31	-3.02	-2.12	-1.44	-1.30	-1.19	-1.07
250	-3.84	-3.54	-3.29	-3.01	-2.12	-1.45	-1.30	-1.19	-1.07
500	-3.82	-3.52	-3.28	-3.00	-2.12	-1.45	-1.30	-1.19	-1.07
$\infty$	-3.80	-3.51	-3.27	-2.99	-2.12	-1.45	-1.30	-1.19	-1.07
<b>Quadratic Trend Removed</b>									
25	-4.75	-4.30	-3.97	-3.61	-2.57	-1.84	-1.69	-1.57	-1.45
50	-4.50	-4.14	-3.85	-3.54	-2.58	-1.86	-1.70	-1.58	-1.45
100	-4.38	-4.05	-3.79	-3.50	-2.58	-1.87	-1.70	-1.58	-1.45
250	-4.30	-4.00	-3.76	-3.47	-2.58	-1.87	-1.70	-1.58	-1.44
500	-4.28	-3.99	-3.74	-3.47	-2.58	-1.87	-1.71	-1.58	-1.44
$\infty$	-4.26	-3.97	-3.73	-3.46	-2.58	-1.87	-1.71	-1.58	-1.44

NOTE. This table was constructed by David A. Dickey.

**Table 10.A.4. Cumulative Distribution of Weighted Symmetric  $\hat{\tau}_v$  for  $\rho = 1$** 

Sample Size <i>n</i>	Probability of a Smaller Value								
	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
<b>No Adjustment</b>									
25	-2.73	-2.37	-2.09	-1.80	-1.05	-0.05	0.24	0.48	0.80
50	-2.74	-2.40	-2.13	-1.85	-1.09	-0.06	0.24	0.51	0.83
100	-2.74	-2.42	-2.16	-1.88	-1.10	-0.06	0.25	0.53	0.85
250	-2.75	-2.43	-2.17	-1.89	-1.12	-0.06	0.25	0.53	0.86
500	-2.75	-2.43	-2.18	-1.90	-1.12	-0.06	0.25	0.54	0.86
$\infty$	-2.76	-2.44	-2.18	-1.90	-1.12	-0.07	0.26	0.54	0.86
<b>Mean removed</b>									
25	-3.33	-2.92	-2.60	-2.26	-1.19	-0.07	0.25	0.51	0.84
50	-3.21	-2.85	-2.57	-2.25	-1.19	-0.04	0.30	0.58	0.93
100	-3.16	-2.82	-2.55	-2.24	-1.20	-0.02	0.32	0.62	0.98
250	-3.12	-2.80	-2.54	-2.23	-1.20	-0.01	0.34	0.64	1.01
500	-3.11	-2.80	-2.53	-2.23	-1.20	-0.00	0.34	0.65	1.01
$\infty$	-3.10	-2.79	-2.52	-2.22	-1.20	-0.00	0.35	0.65	1.01
<b>Linear Trend Removed</b>									
25	-4.11	-3.70	-3.37	-3.02	-1.98	-1.07	-0.82	-0.60	-0.35
50	-3.93	-3.57	-3.28	-2.96	-1.96	-1.01	-0.72	-0.48	-0.19
100	-3.84	-3.51	-3.24	-2.94	-1.96	-0.97	-0.68	-0.42	-0.11
250	-3.78	-3.47	-3.21	-2.92	-1.95	-0.95	-0.65	-0.38	-0.06
500	-3.76	-3.45	-3.20	-2.91	-1.95	-0.95	-0.64	-0.37	-0.05
$\infty$	-3.75	-3.45	-3.19	-2.91	-1.94	-0.94	-0.63	-0.36	-0.03

NOTE. This table was constructed by Heon Jin Park.

**Table 10.A.5. Cumulative Distribution of Maximum Likelihood  $\hat{\tau}_m$  for  $\rho = 1$** 

Sample Size <i>n</i>	Probability of a Smaller Value								
	0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
<b>No Adjustment</b>									
25	-2.80	-2.44	-2.16	-1.86	-0.99	-0.40	-0.32	-0.27	-0.23
50	-2.78	-2.44	-2.17	-1.87	-1.00	-0.40	-0.32	-0.27	-0.22
100	-2.77	-2.44	-2.17	-1.88	-1.00	-0.39	-0.32	-0.27	-0.22
250	-2.76	-2.44	-2.17	-1.88	-1.00	-0.39	-0.32	-0.27	-0.22
500	-2.75	-2.44	-2.17	-1.88	-1.00	-0.39	-0.31	-0.27	-0.22
$\infty$	-2.75	-2.44	-2.17	-1.88	-1.00	-0.39	-0.31	-0.27	-0.22
<b>Mean Removed</b>									
25	-3.52	-3.11	-2.78	-2.44	-1.44	-0.65	-0.51	-0.42	-0.35
50	-3.33	-2.98	-2.69	-2.38	-1.42	-0.64	-0.51	-0.42	-0.34
100	-3.24	-2.92	-2.64	-2.34	-1.41	-0.64	-0.50	-0.42	-0.34
250	-3.19	-2.88	-2.61	-2.32	-1.40	-0.64	-0.50	-0.41	-0.34
500	-3.17	-2.86	-2.60	-2.31	-1.39	-0.64	-0.50	-0.41	-0.34
$\infty$	-3.16	-2.85	-2.59	-2.30	-1.38	-0.63	-0.50	-0.41	-0.34
<b>Linear Trend Removed</b>									
25	-4.40	-3.97	-3.62	-3.25	-2.17	-1.35	-1.15	-1.00	-0.87
50	-4.07	-3.72	-3.43	-3.11	-2.12	-1.34	-1.14	-1.00	-0.86
100	-3.92	-3.60	-3.34	-3.04	-2.10	-1.33	-1.14	-0.99	-0.86
250	-3.84	-3.54	-3.28	-3.00	-2.08	-1.32	-1.13	-0.99	-0.85
500	-3.82	-3.52	-3.27	-2.98	-2.08	-1.32	-1.13	-0.99	-0.85
$\infty$	-3.79	-3.50	-3.25	-2.97	-2.07	-1.31	-1.13	-0.99	-0.85

NOTE. This table was constructed by Heon Jin Park.

Table 10.A.6. Vector Ordinary Least Squares: No Adjustment

D.f.	$H_0$	$H_1$	Probability of a Smaller Value								
			0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
23	1	0	0.00	0.00	0.01	0.02	0.58	2.70	3.68	4.62	5.80
22	2	0	1.18	1.49	1.84	2.31	4.75	8.35	9.53	10.60	11.89
	2	1	1.01	1.28	1.58	1.99	4.16	7.42	8.48	9.47	10.59
21	3	0	5.38	6.11	6.80	7.67	11.34	15.76	17.09	18.34	19.83
	3	1	5.11	5.83	6.50	7.32	10.84	15.06	16.34	17.49	18.86
	3	2	3.34	3.82	4.30	4.90	7.48	10.63	11.58	12.37	13.32
20	4	0	12.06	13.11	14.04	15.19	19.62	24.63	26.10	27.47	29.00
	4	1	11.78	12.82	13.72	14.84	19.15	24.03	25.46	26.75	28.29
	4	2	9.78	10.67	11.46	12.44	16.17	20.27	21.43	22.51	23.70
	4	3	5.81	6.40	6.92	7.55	10.05	12.77	13.54	14.19	14.91
48	1	0	0.00	0.00	0.01	0.02	0.59	2.83	3.90	4.96	6.36
47	2	0	1.21	1.56	1.93	2.43	5.12	9.35	10.84	12.23	13.90
	2	1	1.04	1.34	1.65	2.10	4.50	8.39	9.79	11.02	12.67
46	3	0	5.80	6.64	7.43	8.44	12.84	18.53	20.38	22.06	24.13
	3	1	5.53	6.35	7.10	8.06	12.31	17.84	19.62	21.24	23.19
	3	2	3.61	4.17	4.73	5.42	8.62	12.95	14.38	15.70	17.23
45	4	0	13.55	14.85	16.04	17.47	23.31	30.26	32.42	34.36	36.67
	4	1	13.26	14.53	15.68	17.10	22.81	29.65	31.80	33.67	35.88
	4	2	11.08	12.15	13.19	14.43	19.49	25.56	27.40	29.07	31.01
	4	3	6.62	7.34	8.02	8.87	12.41	16.82	18.19	19.38	20.83
98	1	0	0.00	0.00	0.01	0.02	0.59	2.91	4.03	5.16	6.69
97	2	0	1.23	1.59	1.97	2.49	5.31	9.91	11.56	13.15	15.08
	2	1	1.06	1.37	1.69	2.15	4.67	8.91	10.49	11.98	13.84
96	3	0	6.00	6.89	7.74	8.83	13.62	20.08	22.23	24.21	26.64
	3	1	5.73	6.59	7.39	8.44	13.09	19.37	21.46	23.40	25.74
	3	2	3.74	4.34	4.94	5.69	9.23	14.29	16.00	17.59	19.56
95	4	0	14.33	15.76	17.06	18.66	25.32	33.50	36.10	38.44	41.23
	4	1	14.01	15.41	16.68	18.26	24.81	32.86	35.44	37.74	40.48
	4	2	11.72	12.93	14.07	15.46	21.32	28.61	30.94	33.02	35.46
	4	3	7.05	7.84	8.60	9.56	13.71	19.19	20.97	22.61	24.61

Table 10.A.6.—Continued

D.f.	$H_0$	$H_1$	Probability of a Smaller Value								
			0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
248	1	0	0.00	0.00	0.01	0.02	0.60	2.95	4.10	5.27	6.87
247	2	0	1.25	1.61	1.99	2.53	5.42	10.26	12.01	13.73	15.87
	2	1	1.07	1.38	1.71	2.18	4.77	9.25	10.93	12.58	14.59
246	3	0	6.12	7.03	7.92	9.05	14.11	21.07	23.42	25.60	28.28
	3	1	5.83	6.71	7.56	8.65	13.57	20.35	22.65	24.79	27.41
	3	2	3.82	4.44	5.06	5.85	9.61	15.15	17.05	18.83	21.11
245	4	0	14.80	16.32	17.69	19.39	26.57	35.59	38.49	41.15	44.27
	4	1	14.46	15.94	17.30	18.98	26.06	34.95	37.81	40.45	43.55
	4	2	12.11	13.42	14.62	16.10	22.47	30.59	33.25	35.66	38.48
	4	3	7.31	8.15	8.96	9.99	14.54	20.74	22.83	24.79	27.21
498	1	0	0.00	0.00	0.01	0.02	0.60	2.96	4.12	5.30	6.91
497	2	0	1.25	1.61	2.00	2.54	5.46	10.37	12.16	13.92	16.14
	2	1	1.08	1.39	1.72	2.19	4.80	9.36	11.08	12.75	14.86
496	3	0	6.16	7.07	7.98	9.12	14.28	21.42	23.84	26.07	28.85
	3	1	5.86	6.75	7.61	8.72	13.73	20.69	23.06	25.27	27.98
	3	2	3.84	4.47	5.09	5.90	9.74	15.43	17.42	19.28	21.66
495	4	0	14.95	16.50	17.90	19.65	26.99	36.31	39.32	42.10	45.34
	4	1	14.62	16.12	17.51	19.23	26.48	35.67	38.64	41.42	44.62
	4	2	12.23	13.57	14.80	16.33	22.86	31.27	34.04	36.58	39.56
	4	3	7.40	8.27	9.09	10.14	14.82	21.28	23.49	25.55	28.13
$\infty$	1	0	0.00	0.00	0.01	0.02	0.60	2.97	4.13	5.31	6.94
	2	0	1.25	1.62	2.01	2.55	5.49	10.48	12.31	14.10	16.42
	2	1	1.08	1.39	1.72	2.20	4.83	9.48	11.23	12.88	15.14
3	0	6.19	7.11	8.03	9.18	14.46	21.78	24.26	26.54	29.41	
	3	1	5.89	6.78	7.66	8.77	13.91	21.06	23.47	25.75	28.56
	3	2	3.87	4.49	5.13	5.95	9.88	15.71	17.82	19.77	22.22
4	0	15.10	16.67	18.12	19.91	27.41	37.04	40.17	43.08	46.47	
	1	14.78	16.29	17.73	19.50	26.89	36.40	39.51	42.44	45.69	
	2	12.36	13.72	15.00	16.56	23.25	31.96	34.81	37.54	40.71	
	3	7.50	8.39	9.21	10.31	15.11	21.83	24.18	26.31	29.07	

NOTE. This table was constructed by Heon Jin Park. See equation (10.3.48) for the definition of the statistics.

Table 10.A.7. Vector Ordinary Least Squares: Mean Removed

D.f.	$H_0$	$H_1$	Probability of a Smaller Value								
			0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
22	1	0	0.00	0.02	0.07	0.24	2.13	5.20	6.25	7.23	8.44
21	2	0	2.51	2.99	3.51	4.18	7.32	11.27	12.57	13.68	15.03
	2	1	2.07	2.46	2.85	3.38	5.79	8.92	9.89	10.75	11.75
20	3	0	7.90	8.74	9.50	10.46	14.38	18.93	20.30	21.53	23.03
	3	1	7.33	8.11	8.80	9.67	13.18	17.17	18.36	19.43	20.65
	3	2	4.53	5.05	5.53	6.13	8.59	11.41	12.23	12.93	13.73
19	4	0	15.22	16.27	17.23	18.37	22.74	27.65	29.10	30.41	32.04
	4	1	14.62	15.63	16.53	17.60	21.69	26.18	27.50	28.66	29.99
	4	2	11.78	12.62	13.37	14.28	17.69	21.34	22.41	23.30	24.32
	4	3	6.79	7.33	7.82	8.39	10.64	13.02	13.65	14.19	14.81
47	1	0	0.00	0.02	0.07	0.25	2.30	5.88	7.20	8.43	9.99
46	2	0	2.69	3.28	3.85	4.63	8.33	13.42	15.07	16.57	18.40
	2	1	2.23	2.68	3.13	3.74	6.65	10.80	12.25	13.51	15.10
45	3	0	8.92	9.93	10.91	12.13	17.22	23.46	25.41	27.19	29.30
	3	1	8.28	9.23	10.11	11.24	15.87	21.60	23.36	24.98	26.90
	3	2	5.15	5.81	6.43	7.20	10.54	14.88	16.23	17.44	18.92
44	4	0	18.02	19.46	20.75	22.33	28.58	35.83	38.06	40.07	42.40
	4	1	17.33	18.74	19.95	21.45	27.42	34.28	36.36	38.28	40.49
	4	2	14.05	15.23	16.28	17.57	22.72	28.65	30.48	32.09	33.99
	4	3	8.19	8.96	9.65	10.50	14.01	18.23	19.54	20.65	22.10
97	1	0	0.00	0.02	0.07	0.26	2.38	6.26	7.72	9.10	10.82
96	2	0	2.77	3.40	4.03	4.86	8.89	14.60	16.53	18.29	20.50
	2	1	2.31	2.79	3.27	3.93	7.12	11.90	13.59	15.11	17.07
95	3	0	9.44	10.58	11.66	13.03	18.79	26.10	28.45	30.61	33.26
	3	1	8.78	9.84	10.82	12.08	17.38	24.18	26.36	28.37	30.83
	3	2	5.48	6.21	6.91	7.77	11.65	16.93	18.67	20.30	22.26
94	4	0	19.47	21.12	22.61	24.47	31.87	40.67	43.45	45.94	48.96
	4	1	18.75	20.34	21.76	23.53	30.65	39.07	41.72	44.14	47.05
	4	2	15.24	16.59	17.83	19.35	25.59	33.08	35.44	37.58	40.13
	4	3	8.93	9.81	10.64	11.66	15.99	21.50	23.29	24.91	26.97

Table 10.A.7.—Continued

D.f.	$H_0$	$H_1$	Probability of a Smaller Value								
			0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
247	1	0	0.00	0.02	0.07	0.26	2.43	6.47	8.01	9.48	11.34
246	2	0	2.82	3.47	4.13	5.00	9.23	15.36	17.47	19.42	21.93
	2	1	2.35	2.86	3.36	4.04	7.41	12.60	14.44	16.17	18.38
245	3	0	9.76	10.99	12.13	13.58	19.77	27.83	30.45	32.88	35.92
	3	1	9.08	10.21	11.26	12.60	18.33	25.85	28.33	30.63	33.49
	3	2	5.69	6.46	7.20	8.13	12.35	18.29	20.32	22.24	24.57
244	4	0	20.38	22.15	23.78	25.81	33.97	43.88	47.05	49.89	53.48
	4	1	19.64	21.35	22.90	24.85	32.71	42.25	45.32	48.09	51.53
	4	2	15.99	17.45	18.81	20.50	27.44	36.03	38.79	41.32	44.39
	4	3	9.41	10.36	11.27	12.40	17.28	23.73	25.91	27.93	30.44
497	1	0	0.00	0.02	0.07	0.26	2.45	6.52	8.10	9.59	11.53
496	2	0	2.84	3.49	4.15	5.04	9.33	15.62	17.79	19.82	22.43
	2	1	2.36	2.88	3.39	4.07	7.50	12.83	14.72	16.54	18.85
495	3	0	9.86	11.13	12.28	13.75	20.10	28.43	31.15	33.68	36.84
	3	1	9.19	10.34	11.41	12.76	18.65	26.43	29.02	31.41	34.42
	3	2	5.76	6.54	7.30	8.25	12.59	18.77	20.91	22.92	25.38
494	4	0	20.69	22.51	24.19	26.27	34.70	45.00	48.32	51.30	55.09
	4	1	19.96	21.69	23.30	25.30	33.42	43.37	46.59	49.48	53.13
	4	2	16.25	17.75	19.15	20.90	28.08	37.07	39.98	42.65	45.93
	4	3	9.58	10.56	11.49	12.65	17.72	24.53	26.85	29.02	31.70
$\infty$	1	0	0.00	0.02	0.07	0.26	2.46	6.56	8.16	9.68	11.74
	2	0	2.86	3.51	4.17	5.07	9.41	15.89	18.11	20.21	22.94
	2	1	2.36	2.88	3.41	4.10	7.59	13.05	14.99	16.93	19.36
	3	0	9.97	11.25	12.45	13.91	20.42	29.03	31.85	34.49	37.73
493	3	1	9.29	10.46	11.56	12.92	18.96	27.02	29.70	32.18	35.32
	3	2	5.83	6.62	7.39	8.37	12.83	19.27	21.54	23.59	26.19
	4	0	21.04	22.89	24.63	26.74	35.44	46.17	49.62	52.77	56.73
492	4	1	20.31	22.07	23.73	25.76	34.14	44.55	47.90	50.91	54.79
	4	2	16.54	18.09	19.52	21.33	28.72	38.12	41.20	44.00	47.55
	4	3	9.78	10.78	11.73	12.92	18.17	25.33	27.85	30.17	33.01

NOTE. This table was constructed by Heon Jin Park. See equation (10.3.48) for the definition of the statistics.

Table 10.A.8. Vector Ordinary Least Squares: Linear Trend Removed

D.f.	$H_0$	$H_1$	Probability of a Smaller Value								
			0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
21	1	0	0.08	0.33	0.71	1.32	3.89	7.19	8.24	9.22	10.32
20	2	0	4.20	4.93	5.62	6.49	10.09	14.22	15.44	16.51	17.78
	2	1	3.24	3.73	4.20	4.79	7.31	10.30	11.18	11.96	12.83
19	3	0	10.65	11.62	12.49	13.56	17.72	22.26	23.62	24.78	26.18
	3	1	9.56	10.38	11.14	12.04	15.50	19.26	20.33	21.27	22.39
	3	2	5.73	6.26	6.75	7.35	9.70	12.22	12.95	13.54	14.21
18	4	0	18.56	19.68	20.64	21.84	26.30	31.04	32.44	33.67	35.12
	4	1	17.52	18.51	19.40	20.47	24.41	28.53	29.74	30.79	31.96
	4	2	13.77	14.55	15.28	16.17	19.37	22.62	23.56	24.34	25.22
	4	3	7.79	8.28	8.74	9.30	11.34	13.39	13.93	14.38	14.87
46	1	0	0.11	0.40	0.80	1.43	4.32	8.44	9.86	11.13	12.76
45	2	0	4.71	5.56	6.41	7.46	11.95	17.50	19.31	20.90	22.79
	2	1	3.57	4.16	4.75	5.48	8.74	13.06	14.46	15.75	17.29
44	3	0	12.52	13.79	14.96	16.38	22.08	28.75	30.81	32.67	34.83
	3	1	11.18	12.30	13.30	14.51	19.47	25.33	27.13	28.77	30.70
	3	2	6.79	7.49	8.16	8.98	12.46	16.71	18.07	19.31	20.78
43	4	0	22.95	24.58	25.98	27.73	34.47	42.06	44.25	46.18	48.65
	4	1	21.60	23.05	24.39	26.00	32.18	39.12	41.20	43.01	45.21
	4	2	17.06	18.31	19.46	20.82	26.03	31.92	33.69	35.21	37.14
	4	3	9.77	10.59	11.31	12.17	15.69	19.81	21.05	22.17	23.45
96	1	0	0.13	0.43	0.86	1.50	4.54	9.09	10.72	12.23	14.13
95	2	0	4.96	5.89	6.79	7.94	12.95	19.36	21.46	23.41	25.75
	2	1	3.76	4.39	5.02	5.84	9.51	14.62	16.36	17.98	20.02
94	3	0	13.48	14.88	16.22	17.85	24.49	32.53	35.06	37.37	40.12
	3	1	12.02	13.26	14.40	15.82	21.69	28.90	31.22	33.31	35.84
	3	2	7.32	8.12	8.89	9.85	14.04	19.47	21.28	22.90	24.95
93	4	0	25.21	27.12	28.81	30.92	39.08	48.57	51.44	54.05	57.18
	4	1	23.71	25.44	27.06	29.01	36.60	45.50	48.22	50.68	53.57
	4	2	18.80	20.32	21.70	23.35	29.88	37.63	40.07	42.16	44.76
	4	3	10.85	11.83	12.71	13.78	18.26	23.89	25.69	27.33	29.28

Table 10.A.8.—Continued

D.f.	$H_0$	$H_1$	Probability of a Smaller Value									
			0.01	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99	
246	1	0	0.14	0.44	0.88	1.54	4.68	9.49	11.26	12.93	15.03	
245	2	0	5.10	6.08	7.02	8.22	13.56	20.57	22.87	25.06	27.74	
	2	1	3.86	4.53	5.18	6.05	9.99	15.66	17.63	19.48	21.85	
244	3	0	14.05	15.55	16.98	18.75	26.02	35.01	37.89	40.52	43.72	
	3	1	12.53	13.84	15.08	16.63	23.10	31.26	33.95	36.37	39.36	
	3	2	7.63	8.51	9.34	10.40	15.05	21.33	23.47	25.39	27.84	
243	4	0	26.61	28.71	30.58	32.92	42.05	52.92	56.31	59.42	63.11	
	4	1	25.01	26.95	28.73	30.89	39.47	49.76	52.99	55.92	59.43	
	4	2	19.91	21.59	23.11	24.94	32.38	41.49	44.42	47.00	50.16	
	4	3	11.54	12.61	13.59	14.81	19.96	26.71	28.96	31.02	33.54	
496	1	0	0.14	0.45	0.89	1.55	4.72	9.62	11.44	13.17	15.34	
495	2	0	5.15	6.13	7.09	8.32	13.76	20.99	23.39	25.66	28.46	
	2	1	3.89	4.57	5.24	6.11	10.15	16.03	18.09	20.03	22.49	
494	3	0	14.23	15.77	17.23	19.05	26.55	35.87	38.88	41.64	45.02	
	3	1	12.70	14.04	15.30	16.91	23.59	32.10	34.91	37.45	40.63	
	3	2	7.72	8.64	9.49	10.58	15.40	21.98	24.25	26.28	28.89	
493	4	0	27.09	29.26	31.18	33.61	43.09	54.47	58.05	61.33	65.28	
	4	1	25.46	27.48	29.30	31.53	40.46	51.26	54.69	57.80	61.59	
	4	2	20.31	22.03	23.58	25.49	33.26	42.87	45.95	48.74	52.18	
	4	3	11.79	12.88	13.89	15.16	20.56	27.73	30.15	32.38	35.13	
$\infty$	1	0	0.14	0.45	0.89	1.55	4.77	9.77	11.63	13.41	15.67	
	2	0	5.19	6.16	7.16	8.41	13.95	21.44	24.00	26.32	29.24	
	2	1	3.91	4.60	5.29	6.17	10.30	16.43	18.59	20.64	23.15	
	3	0	14.40	16.00	17.47	19.36	27.08	36.76	39.92	42.83	46.41	
4	3	1	12.85	14.23	15.54	17.19	24.08	32.97	35.92	38.57	41.97	
	3	2	7.82	8.77	9.64	10.78	15.75	22.64	25.05	27.25	30.00	
	4	0	27.60	29.86	31.81	34.31	44.16	56.11	59.88	63.27	67.64	
	4	1	25.94	28.06	29.89	32.18	41.48	52.80	56.47	59.74	63.97	
4	2	20.75	22.50	24.06	26.05	34.15	44.30	47.50	50.57	54.44		
	4	3	12.06	13.16	14.19	15.51	21.19	28.78	31.42	33.81	36.88	

NOTE. This table was constructed by Heon Jin Park. See equation (10.3.48) for the definition of the statistics.

**Table 10.A.9. Percentiles of Test Statistics for Test of Unit Root of Moving Average**

Sample Size <i>n</i>	Probability of a Smaller Value								
	0.00	0.025	0.05	0.10	0.50	0.90	0.95	0.975	0.99
$\tau^*_\mu$ of (10.4.8)									
25	-3.87	-3.47	-3.16	-2.83	-1.88	-1.28	-1.16	-1.08	-1.00
50	-3.71	-3.37	-3.09	-2.79	-1.89	-1.28	-1.16	-1.07	-0.98
100	-3.64	-3.32	-3.06	-2.77	-1.89	-1.28	-1.16	-1.06	-0.97
250	-3.59	-3.29	-3.04	-2.76	-1.90	-1.28	-1.16	-1.06	-0.97
500	-3.58	-3.28	-3.03	-2.75	-1.90	-1.28	-1.16	-1.06	-0.97
750	-3.57	-3.28	-3.02	-2.75	-1.90	-1.28	-1.16	-1.06	-0.96
$\infty$	-3.56	-3.27	-3.02	-2.75	-1.90	-1.28	-1.16	-1.06	-0.96
$\tau^*_\mu$ of (10.4.15)									
25	-4.51	-4.09	-3.75	-3.40	-2.39	-1.69	-1.55	-1.45	-1.34
50	-4.28	-3.93	-3.64	-3.33	-2.39	-1.70	-1.56	-1.45	-1.34
100	-4.17	-3.85	-3.58	-3.29	-2.39	-1.71	-1.56	-1.45	-1.33
250	-4.10	-3.80	-3.55	-3.27	-2.40	-1.72	-1.56	-1.45	-1.33
500	-4.08	-3.78	-3.54	-3.27	-2.40	-1.72	-1.56	-1.45	-1.33
750	-4.07	-3.78	-3.53	-3.26	-2.40	-1.72	-1.57	-1.45	-1.33
$\infty$	-4.06	-3.77	-3.53	-3.26	-2.40	-1.72	-1.57	-1.45	-1.33

NOTE. This table was constructed by David A. Dickey.

## APPENDIX B

### Appendix 10.B. Data Used in Examples

**Table 10.B.1. Seasonally Adjusted Interest Rates**

Obs.	U1	U2	U3
1	4.12808	4.39345	5.02055
2	4.11977	4.10203	4.47400
3	4.02260	3.39690	3.82110
4	4.08650	3.23849	3.62040
5	3.89452	3.33345	3.65055
6	3.29188	2.52677	3.15180
7	3.13677	2.25655	2.95945
8	2.78538	2.15797	2.64600
9	2.41740	2.39310	2.71890
10	2.34835	2.29151	2.84960
11	2.35063	2.32655	2.79945
12	2.00812	2.18323	2.58820
13	1.58808	2.28345	2.70055
14	2.68977	2.56203	2.92400
15	2.20260	2.47690	2.90110
16	1.66650	2.29849	2.81040
17	2.02452	2.33345	2.79055
18	1.70188	2.39677	2.85180
19	1.06678	2.19655	2.71945
20	1.80538	2.24797	2.73600
21	1.69740	2.19310	2.73890
22	2.13835	2.29151	2.82960
23	2.53062	2.43655	2.82945
24	2.35812	2.53323	2.91820
25	2.27808	2.76345	3.26055
26	2.51977	2.87203	3.38400
27	2.88260	2.80690	3.12110
28	2.85650	2.73849	2.97040
29	2.33452	2.72345	2.98055

Table 10.B.1.—Continued

Obs.	U1	U2	U3
30	2.65188	2.79677	2.94180
31	2.61678	2.87655	3.09945
32	2.73538	2.67797	2.92600
33	2.71740	2.69310	2.84890
34	2.77835	2.73151	2.82960
35	2.85062	2.78655	2.86945
36	2.95812	2.80323	2.88820
37	3.04808	2.95345	3.07055
38	3.14977	3.06203	3.17399
39	3.16260	2.97690	3.11110
40	3.06651	3.90849	3.10040
41	3.04452	2.96345	3.13055
42	2.96188	3.05677	3.16180
43	2.92678	3.13656	3.32946
44	3.29538	3.17797	3.32601
45	3.29740	3.29310	3.42890
46	3.37835	3.44151	3.53960
47	3.39062	3.47655	3.59945
48	2.40811	3.45323	3.63820
49	3.61808	3.56344	3.75054
50	3.62977	3.67203	3.88399
51	3.61260	3.62690	3.92110
52	3.63651	3.47849	3.82040
53	3.54453	3.52345	3.78055
54	3.47188	3.54677	3.75180
55	3.32678	3.41656	3.56946
56	3.30538	3.35797	3.49601
57	3.26740	3.44309	3.58890
58	3.23835	3.56151	3.71960
59	3.43062	3.59655	3.78945
60	3.87811	3.77323	3.90820
61	4.03808	3.85344	3.98054
62	4.12977	4.07203	4.17399
63	4.22260	4.01691	4.16110
64	4.25651	3.93849	4.07040
65	4.14453	3.93345	4.03055
66	4.01188	3.86677	3.94180
67	3.99678	3.79656	3.81946
68	3.92538	3.69797	3.78601
69	3.82740	3.83309	3.94890
70	3.95835	4.02151	4.08960
71	4.01062	4.04655	4.15945
72	4.34811	4.31323	4.50820
73	4.55808	4.63344	4.76054
74	4.74977	4.79203	4.98399

**Table 10.B.1.—Continued**

Obs.	U1	U2	U3
75	4.83260	4.67691	4.95111
76	4.83651	4.62849	4.81040
77	4.94453	4.68345	4.92055
78	5.14188	4.56677	4.83180
79	5.20678	4.75656	4.86946
80	5.33538	4.81797	5.16601
81	5.21740	5.28309	5.65889
82	5.40835	5.34151	5.44960
83	5.68062	5.27655	5.41945
84	5.42811	4.89323	4.94819
85	5.07808	4.76344	4.68054
86	5.14977	4.70203	4.74399
87	4.71260	4.34691	4.32111
88	4.21651	3.84849	3.97040
89	3.98453	3.64345	3.95055
90	3.95188	3.60677	4.21181
91	3.69678	4.16656	4.82946
92	3.69538	4.12797	4.87601
93	3.81739	4.33309	4.95889
94	3.75835	4.55151	5.14960
95	4.03062	4.68656	5.31945
96	4.53811	4.90323	5.53819
97	4.73808	5.04344	5.37054
98	4.86977	5.12203	5.41399
99	5.23260	5.25691	5.56111
100	5.92651	5.38849	5.54040
101	6.16453	5.70344	5.91055
102	6.04188	5.58677	5.73181
103	5.92678	5.26656	5.31946
104	5.83538	4.94797	4.97601
105	5.59739	5.10309	5.03889
106	5.79835	5.34151	5.25960
107	5.72062	5.40656	5.43945
108	6.04811	5.89323	5.91819
109	6.43807	6.18344	6.14054
110	6.78977	6.26203	6.38399
111	6.97260	6.10691	6.34111
112	7.57651	6.11849	6.11040
113	8.71453	6.08344	6.18055
114	8.87188	6.50677	6.93181
115	8.51678	6.95656	7.11946
116	8.99538	6.83797	7.10601
117	8.96739	7.00309	7.20889
118	8.87835	6.99151	7.13960
119	8.76062	7.19656	7.28945

**Table 10.B.1.—Continued**

Obs.	U1	U2	U3
120	8.99810	7.75323	7.56819
121	9.11807	7.91344	7.58054
122	9.12977	7.27203	7.22399
123	7.94260	6.71691	6.64111
124	8.26651	6.51849	6.60040
125	7.98453	6.88344	7.19055
126	7.57188	6.74677	7.12181
127	7.11678	6.40657	6.54947
128	6.41538	6.26797	6.37601
129	6.10739	6.04309	6.24889
130	6.07835	5.90151	6.15960
131	5.51062	5.23656	5.31945
132	4.92810	4.80322	4.78819
133	4.27807	4.48343	4.47053
134	3.86977	3.84203	4.01399
135	3.89260	3.46691	3.74111
136	4.31651	3.86849	4.16040
137	4.67453	4.18344	4.71055
138	4.88188	4.81678	5.38181
139	5.21679	5.35657	5.66947
140	5.37538	4.79798	5.34601
141	5.36739	4.60309	5.04889
142	5.07835	4.45151	4.67960
143	4.82062	4.17656	4.41945
144	4.16810	3.94322	4.34819
145	3.63807	3.42343	3.89053
146	3.43977	3.34202	4.23399
147	4.0126	3.81691	4.57111
148	4.3365	3.71849	4.72040
149	4.3145	3.73344	4.53055
150	4.4319	3.97678	4.76181
151	4.4568	3.93657	4.82947
152	4.6054	3.87798	4.72601
153	4.6874	4.57309	5.29889
154	4.9183	4.73151	5.31960
155	4.9706	4.73656	5.12945
156	5.3581	5.00322	5.22819
157	6.0781	5.45343	5.65053
158	6.7298	5.74202	6.10399
159	7.2726	6.17691	6.67111
160	7.2865	6.26849	6.58040
161	7.8845	6.40344	6.70055
162	8.4619	7.25678	7.10181
163	10.3068	7.96657	7.89947
164	10.3054	8.52798	8.14601

**Table 10.B.1.—Continued**

Obs.	U1	U2	U3
165	10.5974	8.20309	7.92889
166	9.8883	7.21151	7.09960
167	9.9406	7.78656	7.32945
168	9.9781	7.38322	6.95819
169	9.7881	7.81343	7.08053
170	9.1198	7.26202	6.68399
171	9.5326	8.04692	7.48111
172	10.6765	8.33849	8.15040
173	11.3545	8.27344	8.28055
174	11.9019	7.96678	8.21181
175	12.8268	7.50657	7.96947
176	11.8154	8.81798	8.70601
177	11.1574	7.97308	8.37889
178	9.9383	7.45151	7.51960
179	9.3606	7.42656	7.21945
180	8.5581	7.08322	6.73819
181	7.2681	6.30343	6.34053
182	6.3898	5.64202	5.73399
183	5.7226	5.57692	5.84111
184	5.6565	5.61849	6.47040
185	5.2645	5.27344	5.98055
186	5.5219	5.40678	5.91181
187	6.0068	6.08657	6.56947
188	5.9454	6.29798	6.98601
189	6.0574	6.33308	7.05889
190	5.6983	5.95151	6.40960
191	5.1306	5.43656	5.99945
192	5.2281	5.37322	6.10819
193	5.0081	4.91343	5.15053
194	4.9198	5.02202	5.70399
195	5.0226	5.08692	5.96111
196	4.9865	4.86849	5.61040
197	5.3345	5.24344	6.05055
198	5.4519	5.47678	6.17181
199	5.2168	5.18657	5.74947
200	5.0954	4.99798	5.46601
201	5.0674	4.99308	5.35889
202	4.9083	4.91151	5.11960
203	4.8606	4.70656	4.92945
204	4.6781	4.28322	4.58819
205	4.7481	4.66343	5.07052
206	4.8298	4.81202	5.33399
207	4.8726	4.68692	5.33111
208	4.8965	4.54849	5.17040
209	5.3945	5.00344	5.50055

Table 10.B.1.—Continued

Obs.	U1	U2	U3
210	5.3619	5.08678	5.46181
211	5.3268	5.14658	5.49948
212	5.7054	5.34798	5.79601
213	5.9574	5.72308	5.98889
214	6.3483	6.15151	6.44960
215	6.4206	6.05656	6.44945
216	6.5881	6.00322	6.46819
217	6.8381	6.48342	6.87052
218	6.9298	6.59202	7.03399
219	6.9726	6.37692	6.96112
220	7.0565	6.29849	7.03040
221	7.4045	6.45344	7.35055
222	7.5719	6.79678	7.58181
223	7.7168	6.96658	7.71948
224	7.8454	6.93798	7.55601
225	8.2674	7.76308	7.86888
226	8.8383	7.98151	8.37960
227	9.6706	8.59656	9.12945
228	10.0581	9.01322	9.38819
229	10.2081	9.39342	9.61052
230	10.2098	9.46202	9.56399
231	10.2726	9.56692	9.52112
232	10.1765	9.46849	9.35040
233	10.2845	9.65344	9.34055
234	10.2619	9.12678	8.86181
235	10.3768	9.19658	8.79948
236	10.7454	9.37798	8.98601

SOURCE. *Banking and Monetary Statistics: 1941–1970*, Federal Reserve Systems, Washington, D.C., and *Annual Statistical Digest* (1970–1979), Federal Reserve Systems, Washington, D.C. These data differ from those in the original sources because they are seasonally adjusted in a different manner.

**Table 10.B.2. Artificial Autoregressive Moving Average Data**

Time	Observation	Time	Observation	Time	Observation
1	-0.589	35	-4.546	68	-5.698
2	-0.665	36	-5.403	69	-3.757
3	1.655	37	-4.411	70	-1.778
4	2.689	38	-3.709	71	-0.637
5	1.695	39	-0.834	72	1.473
6	-2.183	40	2.498	73	3.078
7	-6.983	41	3.103	74	3.157
8	-10.053	42	-0.698	75	3.902
9	-10.566	43	-6.606	76	5.557
10	-8.129	44	-10.295	77	6.295
11	-2.242	45	-8.954	78	4.099
12	4.792	46	-4.039	79	-0.518
13	9.154	47	2.415	80	-5.132
14	9.521	48	8.275	81	-7.221
15	6.793	49	12.414	82	-6.819
16	2.258	50	13.220	83	-5.023
17	-2.285	51	9.959	84	-3.389
18	-4.200	52	3.281	85	-1.546
19	-2.712	53	-3.934	86	1.469
20	-0.547	54	-8.842	87	3.959
21	0.389	55	-9.985	88	3.387
22	0.731	56	-7.533	89	0.777
23	-1.075	57	-3.458	90	-2.547
24	-2.228	58	2.159	91	-5.955
25	-2.305	59	8.960	92	-8.043
26	-0.519	60	14.563	93	-5.394
27	1.788	61	15.020	94	-0.869
28	4.267	62	9.890	95	2.738
29	5.913	63	3.059	96	3.708
30	5.483	64	-3.910	97	3.245
31	3.745	65	-7.551	98	-0.757
32	2.017	66	-8.689	99	-5.353
33	0.318	67	-7.357	100	-8.213
34	-1.804				

**Table 10.B.3. Computer Generated Moving Average Data**

Time	Observation	Time	Observation	Time	Observation
1	0.0824	35	-0.2092	68	-0.8659
2	-0.7421	36	1.1027	69	-0.3166
3	-2.5863	37	0.7200	70	1.2318
4	-1.0847	38	-0.1186	71	2.6148
5	0.0811	39	-0.7560	72	1.3507
6	2.1080	40	-1.1736	73	0.3786
7	1.2283	41	0.1670	74	0.0751
8	0.2222	42	0.4029	75	-0.1134
9	0.1139	43	1.4065	76	-0.9633
10	0.9652	44	-0.6458	77	-2.1476
11	1.7643	45	-1.8765	78	-1.6736
12	2.2358	46	-1.6670	79	0.2056
13	0.8803	47	-0.9132	80	2.3535
14	0.8975	48	0.9964	81	1.7980
15	0.8376	49	1.6263	82	-0.1434
16	-0.3025	50	0.5360	83	-1.4796
17	-2.5194	51	1.0236	84	-1.4369
18	-2.7883	52	0.7965	85	-0.7226
19	-3.5046	53	0.2285	86	0.4259
20	-4.1382	54	-0.2926	87	1.3133
21	-3.6002	55	-1.3110	88	1.8455
22	-1.5214	56	0.3846	89	0.5274
23	2.0135	57	1.9833	90	-1.8385
24	3.2380	58	-0.1662	91	-3.3115
25	1.5426	59	-0.0513	92	-1.7593
26	1.1661	60	3.0565	93	0.0670
27	0.6807	61	1.9968	94	0.4853
28	-0.8016	62	0.4235	95	3.0867
29	-1.5890	63	-1.3013	96	1.7763
30	-1.1519	64	-1.2351	97	1.5428
31	-1.3396	65	-1.7102	98	0.2979
32	0.0265	66	-1.2697	99	-1.4703
33	1.7851	67	-0.6817	100	-1.4294
34	0.1965				

**Table 10.B.4. Computer Generated Data for Vector Process**

Time	$X_t$	$Y_t$	$Z_t$
1	-5.59	-10.78	-6.64
2	-6.85	-12.11	-7.91
3	-7.25	-13.21	-8.12
4	-6.72	-12.48	-6.14
5	-9.56	-12.73	-9.75
6	-8.20	-13.50	-8.82
7	-7.93	-13.66	-8.06
8	-8.85	-13.72	-8.89
9	-9.55	-14.25	-9.98
10	-7.56	-12.61	-6.24
11	-6.47	-9.67	-4.13
12	-4.98	-7.05	-2.88
13	-3.14	-4.43	-1.05
14	-1.65	-2.79	-0.33
15	-0.78	-3.03	-0.98
16	-0.65	-3.47	-1.00
17	2.67	-0.30	5.21
18	5.28	5.08	9.59
19	5.96	8.73	8.88
20	8.97	12.29	11.83
21	13.18	15.86	16.03
22	14.49	19.61	17.49
23	17.45	21.46	18.93
24	19.15	22.98	20.37
25	20.99	23.74	21.60
26	20.56	23.45	20.33
27	19.87	22.11	18.80
28	19.88	21.98	19.77
29	18.84	21.47	18.44
30	18.25	20.14	17.19
31	17.46	18.61	16.24
32	16.03	17.23	14.93
33	14.89	16.32	14.17
34	14.69	14.99	13.62
35	15.46	15.41	15.80
36	14.56	15.60	14.71
37	14.45	14.62	13.66
38	14.27	14.44	14.13
39	14.63	14.98	15.06
40	14.12	16.56	15.38
41	14.89	16.39	14.75
42	15.09	16.30	15.02
43	14.16	15.56	13.57
44	13.46	13.34	11.68

**Table 10.B.4.—Continued**

Time	X <sub>t</sub>	Y <sub>t</sub>	Z <sub>t</sub>
45	13.96	12.44	13.25
46	12.93	11.53	12.20
47	10.99	11.53	10.99
48	9.51	10.19	8.44
49	9.63	8.48	8.26
50	8.58	7.14	7.50
51	8.00	5.00	6.29
52	5.16	3.39	3.88
53	3.46	1.33	1.81
54	4.60	1.13	4.43
55	3.10	1.14	3.11
56	3.53	2.69	4.77
57	5.53	4.21	6.75
58	5.20	4.70	5.59
59	6.41	4.57	6.31
60	6.20	5.29	6.78
61	6.30	5.93	6.81
62	7.63	7.08	8.55
63	8.95	9.84	11.15
64	9.25	12.11	11.07
65	11.25	14.41	13.08
66	15.56	17.82	18.29
67	17.16	21.12	19.81
68	19.77	24.07	22.12
69	19.82	24.78	20.39
70	24.38	26.91	26.08
71	23.36	28.22	24.40
72	24.19	28.20	24.18
73	25.47	28.35	25.59
74	23.57	27.55	22.94
75	22.30	25.98	21.05
76	22.44	25.37	21.95
77	23.50	26.42	24.34
78	23.66	28.76	25.53
79	25.15	30.08	26.21
80	26.57	31.39	27.62
81	27.17	32.18	27.81
82	28.03	33.44	29.04
83	28.43	33.19	28.22
84	27.36	32.43	26.74
85	27.87	30.06	25.98
86	27.17	29.50	26.72
87	26.52	29.52	26.54
88	24.90	28.90	24.40

**Table 10.B.4.** —Continued

Time	X <sub>t</sub>	Y <sub>t</sub>	Z <sub>t</sub>
89	24.56	27.58	23.50
90	24.04	26.97	23.56
91	21.84	26.34	21.33
92	23.68	26.11	23.50
93	22.50	24.44	21.16
94	20.36	24.37	20.30
95	20.53	23.21	19.60
96	20.59	22.00	19.63
97	19.75	20.66	18.67
98	17.93	18.60	16.28
99	15.65	16.22	13.75
100	15.12	14.62	13.83

# Bibliography

- Agiakloglou, C., and Newbold, P. (1992), Empirical evidence on Dickey-Fuller type tests. *J. Time Series Anal.* **13**, 471-483.
- Ahlberg, J. H., Nilson, E. N., and Walsh, J. L. (1967), *The Theory of Splines and Their Applications*. Academic Press, New York.
- Ahn, S. K., and Reinsel, G. C. (1988), Nested reduced-rank autoregressive models for multiple time series. *J. Amer. Statist. Assoc.* **83**, 849-858.
- Ahn, S. K., and Reinsel, G. C. (1990), Estimation for partially nonstationary multivariate autoregressive models. *J. Amer. Statist. Assoc.* **85**, 813-823.
- Ahtola, J., and Tiao, G. C. (1984), Parameter inference for a nearly nonstationary first order autoregressive model. *Biometrika* **71**, 263-272.
- Ahtola, J., and Tiao, G. C. (1987), Distributions of least squares estimators of autoregressive parameters for a process with complex roots on the unit circle. *J. Time Series Anal.* **8**, 1-14.
- Aigner, D. J. (1971), A compendium on estimation of the autoregressive moving average model from time series data. *Int. Econ. Rev.* **12**, 348-371.
- Akaike, H. (1969a), Fitting autoregressive models for prediction. *Ann. Inst. Statist. Math.* **21**, 243-247.
- Akaike, H. (1969b), Power spectrum estimation through autoregressive model fitting. *Ann. Inst. Statist. Math.* **21**, 407-419.
- Akaike, H. (1973), Maximum likelihood identification of Gaussian autoregressive moving average models. *Biometrika* **60**, 255-265.
- Akaike, H. (1974), Markovian representation of stochastic processes and its application to the analysis of autoregressive moving average processes. *Ann. Inst. Statist. Math.* **26**, 363-387.
- Amemiya, T. (1973), Generalized least squares with an estimated autocovariance matrix. *Econometrica* **41**, 723-732.
- Amemiya, T., and Fuller, W.A. (1967), A comparative study of alternative estimators in a distributed lag model. *Econometrica* **35**, 509-529.
- Amemiya, Y. (1990), On the convergence of the ordered roots of a sequence of determinantal equations. *Linear Algebra Appl.* **127**, 531-542.
- Amos, D. E., and Koopmans, L. H. (1963), Tables of the distribution of the coefficient of coherence for stationary bivariate Gaussian processes. Sandia Corporation Monograph SCR-483, Albuquerque, New Mexico.
- Anderson, B. D. O., and Moore, J. B. (1979), *Optimal Filtering*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Anderson, R. L. (1942), Distribution of the serial correlation coefficient. *Ann. Math. Statist.* **13**, 1-13.

- Anderson, R. L., and Anderson, T. W. (1950), Distribution of the circular serial correlation coefficient for residuals from a fitted Fourier series. *Ann. Math. Statist.* **21**, 59–81.
- Anderson, T. W. (1948), On the theory of testing serial correlation. *Skand. Aktuarietidskrift* **31**, 88–116.
- Anderson, T. W. (1952), Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann. Math. Statist.* **23**, 193–212.
- Anderson, T. W. (1959), On asymptotic distributions of estimates of parameters of stochastic difference equations. *Ann. Math. Statist.* **30**, 676–687.
- Anderson, T. W. (1962), The choice of the degree of a polynomial regression as a multiple decision problem. *Ann. Math. Statist.* **33**, 255–265.
- Anderson, T. W. (1971), *The Statistical Analysis of Time Series*. Wiley, New York.
- Anderson, T. W. (1984), *Introduction to Multivariate Statistical Analysis* (2nd ed.). Wiley, New York.
- Anderson, T. W., and Darling, D. A. (1952), Asymptotic theory of certain “goodness of fit” criteria based on stochastic processes. *Ann. Math. Statist.* **23**, 193–212.
- Anderson, T. W., and Takemura, A. (1986), Why do noninvertible estimated moving averages occur? *J. Time Series Anal.* **7**, 235–254.
- Anderson, T. W., and Taylor, J. B. (1979), Strong consistency of least squares estimators in dynamic models. *Ann. Statist.* **7**, 484–489.
- Anderson, T. W., and Walker, A. M. (1964), On the asymptotic distribution of the autocorrelations of a sample from a linear stochastic process. *Ann. Math. Statist.* **35**, 1296–1303.
- Andrews, D. W. K. (1987), Least squares regression with integrated or dynamic regressors under weak error assumptions. *Econometric Theory* **3**, 98–116.
- Arellano, C. (1992), Testing for trend stationarity versus difference stationarity. Unpublished thesis, North Carolina State University, Raleigh, North Carolina.
- Arellano, C., and Pantula, S. G., (1995), Testing for trend stationarity versus difference stationarity. *J. Time Series Anal.* **16**, 147–164.
- Ash, R. B. (1972), *Real Analysis and Probability*. Academic Press, New York.
- Ball, R. J., and St. Cyr, E. B. A. (1966), Short term employment functions in British manufacturing industry. *Rev. Econom. Stud.* **33**, 179–207.
- Bartle, R. G. (1964), *The Elements of Real Analysis*. Wiley, New York.
- Bartlett, M. S. (1946), On the theoretical specification and sampling properties of autocorrelated time series. *Suppl. J. Roy. Statist. Soc.* **8**, 27–41.
- Bartlett, M. S. (1950), Periodogram analysis and continuous spectra. *Biometrika* **37**, 1–16.
- Bartlett, M. S. (1966), *An Introduction to Stochastic Processes with Special Reference to Methods and Applications* (2nd ed.). Cambridge University Press, Cambridge.
- Bartlett, M. S. (1978), *An Introduction to Stochastic Processes* (3rd ed.). Cambridge University Press, Cambridge.
- Basawa, I. V. (1987), Asymptotic distributions of prediction errors and related tests of fit for nonstationary processes. *Ann. Statist.* **15**, 46–58.
- Basmann, R. L. (1957), A generalized classical method of linear estimation of coefficients in a structural equation. *Econometrica* **25**, 77–83.
- Beamish, N., and Priestly, M. B. (1981), A study of AR and window spectral estimation. *Appl. Statist.* **30**, No. 1.

- Bell, W. R. (1983), A computer program (TEST) for detecting outliers in time series. *1983 Proceedings of the Business and Economics Section, Amer. Statist. Assoc.*, 634–639.
- Bellman, R. (1960), *Introduction to Matrix Analysis*. McGraw-Hill, New York.
- Beran, J. (1992), Statistical methods for data with long-range dependence. *Statist. Sci.* **7**, 404–416.
- Beran, J., and Ghosh, S. (1991), Slowly decaying correlations, testing normality, nuisance parameters. *J. Amer. Statist. Assoc.* **86**, 785–791.
- Berk, K. N. (1974), Consistent autoregressive spectral estimates. *Ann. Statist.* **2**, 489–502.
- Bernstein, S. (1927), Sur l'extension du théorème limite du calcul des probabilités aux sommes de quantités dépendantes. *Math. Ann.* **97**, 1–59.
- Bhansali, R. J. (1980), Autoregressive and window estimates of the inverse correlation function. *Biometrika* **67**, 551–566.
- Bhansali, R. J. (1981), Effects of not knowing the order of an autoregressive process I. *J. Amer. Statist. Assoc.* **76**, 588–597.
- Bhansali, R. J. (1991), Consistent recursive estimation of the order of an autoregressive moving average process. *Int. Statist. Rev.* **59**, 81–96.
- Bhansali, R. J. (1993), Order selection for linear time series models: a review. In *Developments in Time Series Analysis*. (T. Subba Rao, ed.) Chapman and Hall, New York. 50–66.
- Bhargava, A. (1986), On the theory of testing for unit roots in observed time series. *Rev. Econ. Stud.* **53**, 369–384.
- Billinglsey, P. (1968), *Convergence of Probability Measures*. Wiley, New York.
- Billingsley, P. (1979), *Probability and Measure*. Wiley, New York.
- Birnbaum, Z. W. (1952), Numerical tabulation of the distribution of Kolmogorov's statistic for finite sample size. *J. Amer. Statist. Assoc.* **47**, 425–441.
- Blackman, R. B., and Tukey, J. W. (1959), *The Measurement of Power Spectra from the Point of View of Communications Engineering*. Dover, New York.
- Bloomfield, P. (1976), *Fourier Analysis of Time Series: An Introduction*. Wiley, New York.
- Bollerslev, T., Chou, R. Y., and Kroner, K. F. (1992), ARCH modeling in finance: A review of the theory and empirical evidence. *J. Econometrics* **52**, 5–59.
- Box, G. E. P. (1954), Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *Ann. Math. Statist.* **25**, 290–316.
- Box, G. E. P., Jenkins, G. M., and Reinsel, G. C. (1994), *Time Series Analysis: Forecasting and Control*. Prentice Hall, Englewood Cliffs, New Jersey.
- 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. (1977), A canonical analysis of multiple time series. *Biometrika*, **64**, 355–365.
- Breidt, F. J., and Carriquiry, A. L. (1994), Stochastic volatility models: an application to S & P 500 Futures. Unpublished manuscript, Iowa State University, Ames, Iowa.
- Breitung, J. (1994), Some simple tests of the moving-average unit root hypothesis. *J. Time Series Anal.* **15**, 351–370.
- Brillinger, D. R. (1975), *Time Series: Data Analysis and Theory*. Holt, Rinehart and

- Winston, New York.
- Brockwell, P. J., and Davis, R. A. (1991), *Time Series: Theory and Methods* (2nd ed.). Springer-Verlag, New York.
- Brown, B. M. (1971), Martingale central limit theorems. *Ann. Math. Statist.* **42**, 59–66.
- Brown, B. M., and Hewitt, J. I. (1975), Asymptotic likelihood theory for diffusion processes. *J. Appl. Prob.* **12**, 228–238.
- Brown, R. G. (1962), *Smoothing, Forecasting, and Prediction of Discrete Time Series*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Bruce, A. G., and Martin, R. D. (1989), Leave- $k$ -out diagnostics for time series (with discussion). *J. Roy. Statist. Soc. B* **51**, 363–424.
- Burg, J. P. (1975), Maximum entropy spectral analysis, Ph.D. thesis, Stanford University, Stanford, California.
- Burguet, J., Gallant, A. R., and Souza, G. (1982), On unification of the asymptotic theory of nonlinear econometric models. *Econ. Rev.* **1**, 151–190.
- Burman, J. P. (1965), Moving seasonal adjustment of economic time series. *J. Roy. Statist. Soc. A* **128**, 534–558.
- Carroll, R. J., and Ruppert, D. (1988), *Transformation and Weighting in Regression*. Chapman and Hall, New York.
- Chan, N. H. (1988), On the parameter inference for nearly nonstationary time series. *J. Amer. Statist. Assoc.* **83**, No. 403, 857–862.
- Chan, N. H., and Tsay, R. S. (1991), Asymptotic inference for non-invertible moving average time series. Unpublished manuscript. Carnegie Mellon University, Pittsburgh, Pennsylvania.
- Chan, N. H., and Wei, C. Z. (1987), Asymptotic inference for nearly nonstationary AR(1) processes. *Ann. Statist.* **15**, 1050–1063.
- Chan, N. H., and Wei, C. Z. (1988), Limiting distributions of least square estimates of unstable autoregressive processes. *Ann. Statist.* **16**, 367–401.
- Chang, I., Tiao, G. C., and Chen, C. (1988), Estimation of time series parameters in the presence of outliers. *Technometrics* **30**, 193–204.
- Chang, M. C. (1989), Testing for overdifference. Unpublished thesis, North Carolina State University, Raleigh, North Carolina.
- Chernoff, H. (1956), Large sample theory: parametric case. *Ann. Math. Statist.* **27**, 1–22.
- Chung, K. L. (1968), *A Course in Probability Theory*. Harcourt, Brace and World, New York.
- Cleveland, W. S. (1971), Fitting time series models for prediction. *Technometrics* **13**, 713–723.
- Cleveland, W. S. (1972), The inverse autocorrelations of a time series and their applications. *Technometrics* **14**, 277–293.
- Cochran, W. T., et al. (1967), What is the fast Fourier transform? *IEEE Trans. Audio and Electroacoustics AU-15*, No. 2, 45–55.
- Cochrane, D., and Orcutt, G. H. (1949), Application of least squares regression to relationships containing autocorrelated error terms. *J. Amer. Statist. Assoc.* **44**, 32–61.
- Colley, J. W., Lewis, P. A. W., and Welch, P. D. (1967), Historical notes on the Fast Fourier Transform. *IEEE Trans. Audio and Electroacoustics AU-15*, No. 2, 76–79.
- Cox, D. D. (1991), Gaussian likelihood estimation for nearly nonstationary AR(1) processes. *Ann. Statist.* **19**, 1129–1142.

- Cox, D. R. (1977), Discussion of papers by Campbell and Walker and Roy. *Statist. Soc. A* **140**, 453–454.
- Cox, D. R. (1984), Long-range dependence: a review. In *Statistics: An Appraisal* (H. A. David and H. T. David, eds.). Iowa State University Press, Ames, Iowa.
- Cox, D. R., and Miller, H. D. (1965), *The Theory of Stochastic Processes*. Wiley, New York.
- Cramér, H. (1940), On the theory of stationary random processes. *Ann. Math.* **41**, 215–230.
- Cramér, H. (1946), *Mathematical Methods of Statistics*. Princeton University Press, Princeton, New Jersey.
- Crowder, M. J. (1980), On the asymptotic properties of least squares estimators in autoregression. *Ann. Statist.* **8**, 132–146.
- Cryer, J. D., and Ledolter, J. (1981), Small sample properties of the maximum likelihood estimator in the MA(1) model. *Biometrika* **68**, 691–694.
- Cryer, J. D., Nankervis, J. C., and Savin, N. E. (1989), Mirror-image distributions in AR(1) models. *Econometric Theory* **5**, 36–52.
- Dahlhaus, R. (1989), Efficient parameter estimation for self-similar processes. *Ann. Statist.* **17**, 1749–1766.
- Davies, R. B. (1973), Asymptotic inference in stationary Gaussian time-series. *Adv. Appl. Prob.* **5**, 469–497.
- Davis, H. T. (1941), *The Analysis of Economic Time Series*. Principia Press, Bloomington, Indiana.
- Davisson, L. D. (1965), The prediction error of stationary Gaussian time series of unknown covariance, *IEEE Trans. Inform. Theory* **11**, 527–532.
- DeGracie, J. S., and Fuller, W. A. (1972), Estimation of the slope and analysis of covariance when the concomitant variable is measured with error. *J. Amer. Statist. Assoc.* **67**, 930–937.
- DeJong, D. N., Nankervis, J. C., Savin, N. E., and Whiteman, C. H. (1992a), Integration versus trend-stationarity in macroeconomic time series. *Econometrica* **60**, 423–434.
- DeJong, D. N., Nankervis, J. C., Savin, N. E., and Whiteman, C. H. (1992b), The power problems of unit root tests for time series with autoregressive errors. *J. Econometrics* **53**, 323–343.
- Denby, L., and Martin, R. D. (1979), Robust estimation of the first order autoregressive parameter. *J. Amer. Statist. Assoc.* **74**, 140–146.
- Deo, R. (1995), Tests for unit roots in multivariate autoregressive processes. Ph.D. dissertation, Iowa State University, Ames, Iowa.
- Dhrymes, P. J. (1971), *Distributed Lags: Problems of Estimation and Formulation*. Holden-Day, San Francisco.
- Diananda, P. H. (1953), Some probability limit theorems with statistical applications. *Proc. Cambridge Philos. Soc.* **49**, 239–246.
- Dickey, D. A. (1976), Estimation and hypothesis testing in nonstationary time series. Ph.D. thesis, Iowa State University, Ames, Iowa.
- Dickey, D. A., Bell, W. R., and Miller, R. B. (1986), Unit roots series model: Tests and implications. *Amer. Statistician* **40**, 12–26.
- Dickey, D. A., and Fuller, W. A. (1979), Distribution of the estimators for autoregressive time series with a unit root. *J. Amer. Statist. Assoc.* **74**, 427–431.

- Dickey, D. A., and Fuller, W. A. (1981), Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica* **49**, 1057–1072.
- 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.
- Dickey, D. A., and Pantula, S. G. (1987), Determining the order of differencing in autoregressive processes. *J. Business and Econ. Statist.* **5**, 455–461.
- Dickey, D. A., and Said, S. E. (1982), Testing ARIMA ( $p, 1, q$ ) versus ARMA ( $p + 1, q$ ). In *Applied Time Series Analysis* (O. D. Anderson, ed.), North Holland, Amsterdam.
- Diderrich, G. T. (1985), The Kalman filter from the perspective of Goldberger-Theil estimators. *Amer. Statistician* **39**, 193–198.
- Diebold, F. X., and Nerlove, M. (1990), Unit roots in economic time series: a selective survey. *Econometrica* **8**, 3–69.
- Draper, N. R., and Smith, H. (1981), *Applied Regression Analysis* (2nd ed.). Wiley, New York.
- Dufour, J. M., and King, M. L. (1991), Optimal invariant tests for the autocorrelation coefficient in linear regressions with stationary or nonstationary AR(1) errors. *J. Econometrics* **47**, 115–143.
- Duncan, D. B., and Horn, S. D. (1972), Linear dynamic recursive estimation from the viewpoint of regression analysis. *J. Amer. Statist. Assoc.* **67**, 815–821.
- Durbin, J. (1959), Efficient estimation of parameters in moving-average models. *Biometrika* **46**, 306–316.
- Durbin, J. (1960), Estimation of parameters in time-series regression models. *J. Roy. Statist. Soc. B* **22**, 139–153.
- Durbin, J. (1961), Efficient fitting of linear models for continuous stationary time series from discrete data. *Bull. Int. Statist. Inst.* **38**, 273–281.
- Durbin, J. (1962), Trend elimination by moving-average and variate-difference filters. *Bull. Int. Statist. Inst.* **39**, 131–141.
- Durbin, J. (1963), Trend elimination for the purpose of estimating seasonal and periodic components of time series. *Proc. Symp. Time Series Analysis, Brown University* (M. Rosenblatt, ed.). Wiley, New York, 3–16.
- Durbin, J. (1967), Tests of serial independence based on the cumulated periodogram. *Bull. Int. Statist. Inst.* **42**, 1039–1049.
- Durbin, J. (1968), The probability that the sample distribution function lies between two parallel straight lines. *Ann. Math. Statist.* **39**, 398–411.
- Durbin, J. (1969), Tests for serial correlation in regression analysis based on the periodogram of least-squares residuals. *Biometrika* **56**, 1–15.
- Durbin, J. (1973), *Distribution Theory for Tests Based on the Sample Distribution Function*. Regional Conference Series in Applied Mathematics No. 9, SIAM, Philadelphia.
- Durbin, J., and Cordero, M. (1993), Handling structural shifts, outliers and heavy-tailed distributions in state space time series models. Unpublished manuscript, London School of Economics and Political Science, London.
- Durbin, J., and Watson, G. S. (1950), Testing for serial correlation in least squares regression, I. *Biometrika* **37**, 409–428.

- Durbin, J., and Watson, G. S. (1951), Testing for serial correlation in least squares regression, II. *Biometrika* **38**, 159–178.
- Dvoretzky, A. (1972), Asymptotic normality for sums of dependent random variables. *Sixth Berkeley Symp. Math. Statist. Prob.* **2**, University of California Press, Berkeley, 513–535.
- Eberlain, E. (1986), On strong invariance principles under dependence assumptions. *Ann. Probab.* **14**, 260–270.
- Eicker, F. (1963), Asymptotic normality and consistency of the least squares estimators for families of linear regressions. *Ann. Math. Statist.* **34**, 447–456.
- Eicker, F. (1967), Limit theorems for regressions with unequal and dependent errors. *Proc. Fifth Berkeley Symp. Math. Statist. Prob.* **1**, University of California Press, Berkeley, 59–82.
- Eisenhart, C., Hastay, M. W., and Wallis, W. A. (1947), *Selected Techniques of Statistical Analysis*. McGraw-Hill, New York.
- Elliott, G. (1993), Efficient tests for a unit root when the initial observation is drawn from its unconditional distribution. Unpublished manuscript. Harvard University, Cambridge, Massachusetts.
- Elliott, G., Rothenberg, T. J., and Stock, J. H. (1992), Efficient tests for an autoregressive unit root. Unpublished paper presented at the NBER-NSF Time Series Seminar, Chicago.
- Elliott, G., and Stock, J. H. (1992), Efficient tests for an autoregressive unit root. Paper presented at the NBER-NSF Time Series Seminar, Chicago.
- Eltinge, J. L. (1991), Exponential convergence properties of autocovariance matrix inverses and latent vector prediction coefficients. Unpublished manuscript, Department of Statistics, Texas A & M University, College Station, Texas.
- Elton, C., and Nicholson, M. (1942), The ten-year cycle in numbers of lynx in Canada. *J. Animal Ecology* **11**, 215–244.
- Engle, R. F. (1982), Autoregressive conditional heteroscedasticity with estimates of UK inflation. *Econometrica* **50**, 987–1007.
- Engle, R. F., and Bollerslev (1986), Modelling the persistence of conditional variances. *Econometric Reviews* **5**, 1–50.
- 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., and Kraft, D. F. (1981), Multiperiod forecast error variances of inflation estimated from ARCH models. *Proc. ASA-Census-NBER Conference on Applied Time Series Analysis of Economic Data* (A. Zellner, ed.), 176–177.
- Evans, G. B. A., and Savin, N. E. (1981a), The calculation of the limiting distribution of the least squares estimator of the parameter in a random walk model. *Ann. Statist.* **9**, 1114–1118.
- Evans, G. B. A., and Savin, N. E. (1981b), Testing for unit roots: 1. *Econometrica* **49**, 753–777.
- Evans, G. B. A., and Savin, N. E. (1984), Testing for unit roots: 2. *Econometrica* **52**, 1241–1249.
- Ezekiel, M., and Fox, K. A. (1959), *Methods of Correlation and Regression Analysis*. Wiley, New York.

- Fieller, E. C. (1932), The distribution of the index in a normal bivariate population. *Biometrika* 24, 428-440.
- Fieller, E. C. (1954), Some problems in interval estimation. *J. Roy. Statist. Soc. B* 26, 175-185.
- Findley, D. F. (1980), Large sample behavior of the  $S$ -array of seasonally nonstationary ARMA series. In *Time Series Analysis* (O. D. Anderson and M. R. Perryman, eds.), North Holland, Amsterdam, 163-170.
- Findley, D. F. (1985), On the unbiasedness property of AIC for exact or approximating linear stochastic time series models. *J. Time Series Anal.* 6, 229-252.
- Findley, D. F., and Wei, C. Z. (1989), Beyond chi-square: likelihood ratio procedures for comparing non-nested possibly incorrect regressors. Unpublished manuscript, U.S. Bureau of the Census, Washington.
- Finkbeiner, D. T. (1960), *Introduction to Matrices and Linear Transformations*. W. H. Freeman, San Francisco.
- Fisher, R. A. (1929), Tests of significance in harmonic analysis. *Proc. Roy. Soc. London Ser. A* 125, 54-59.
- Fishman, G. S. (1969), *Spectral Methods in Econometrics*. Harvard University Press, Cambridge, Massachusetts.
- Fountis, N. G. (1983), Testing for unit roots in multivariate autoregressions. Unpublished Ph.D. thesis, North Carolina State University, Raleigh, North Carolina.
- Fountis, N. G., and Dickey, D. A. (1989), Testing for a unit root nonstationarity in multivariate autoregressive time series. *Ann. Statist.* 17, 419-428.
- Fox, A. J. (1972), Outliers in time series. *J. Roy. Statist. Soc. B* 34, 350-363.
- Fox, R., and Taqqu, M. S. (1985), Noncentral limit theorems for quadratic forms in random variables having long-range dependence. *Ann. Prob.* 13, 428-446.
- Fox, R., and Taqqu, M. S. (1986), Large sample properties of parameter estimates for strongly dependent stationary Gaussian time series. *Ann. Statist.* 14, 517-532.
- Franklin, J. N. (1968), *Matrix Theory*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Friedman, M., and Schwartz, A. J. (1963), *A Monetary History of the United States 1867-1960*. Princeton University Press, Princeton, New Jersey.
- Fuller, W. A. (1969), Grafted polynomials as approximating functions. *Australian J. Agr. Econ.* 13, 35-46.
- Fuller, W. A. (1979), Testing the autoregressive process for a unit root. Paper presented at the 42nd Session of the International Statistical Institute, Manila.
- Fuller, W. A. (1980), The use of indicator variables in computing predictions. *J. Econometrics* 12, 231-243.
- Fuller, W. A. (1984), Nonstationary autoregressive time series. In *Handbook of Statistics, 5, Time Series in the Time Domain* (E. J. Hannan, P. R. Krishnaiah, and M. M. Rao, eds.). North Holland, Amsterdam, 1-23.
- Fuller, W. A. (1987), *Measurement Error Models*. Wiley, New York.
- Fuller, W. A., and Hasza, D. P. (1980), Predictors for the first-order autoregressive process. *J. Econometrics* 13, 139-157.
- Fuller, W. A., and Hasza, D. P. (1981), Properties of predictors for autoregressive time series. *J. Amer. Statist. Assoc.* 76, 155-161.

- Fuller, W. A., Hasza, D. P., and Goebel, J. J. (1981), Estimation of the parameters of stochastic difference equations. *Ann. Statist.* **9**, 531–543.
- Fuller, W. A., and Martin, J. E. (1961), The effects of autocorrelated errors on the statistical estimation of distributed lag models. *J. Farm Econ.* **43**, 71–82.
- Galbraith, R. F., and Galbraith, J. I. (1974), On the inverses of some patterned matrices arising in the theory of stationary time series. *J. Appl. Prob.* **11**, 63–71.
- Gallant, A. R. (1971), *Statistical Inference for Nonlinear Regression Models*. Ph.D. thesis, Iowa State University, Ames, Iowa.
- Gallant, A. R. (1975a), Nonlinear regression. *Amer. Statistician* **29**, 73–81.
- Gallant, A. R. (1975b), The power of the likelihood ratio test of location in nonlinear regression models. *J. Amer. Statist. Assoc.* **70**, 199–203.
- Gallant, A. R. (1987), *Nonlinear Statistical Models*. Wiley, New York.
- Gallant, A. R., and Fuller, W. A. (1973), Fitting segmented polynomial regression models whose join points have to be estimated. *J. Amer. Statist. Assoc.* **68**, 144–147.
- Gallant, A. R., and White, H. (1988), *A Unified Theory of Estimation and Inference for Nonlinear Dynamic Models*. Basil Blackwell, New York.
- Gay, R., and Heyde, C. C. (1990), On a class of random field models which allows long range dependence. *Biometrika* **77**, 401–403.
- Geweke, J., and Porter-Hudak, S. (1983), The estimation and application of long memory time series models. *J. Time Series Anal.* **4**, 221–237.
- Giraitis, L., and Surgailis, D. (1990), A central limit theorem for quadratic forms in strongly dependent linear variables and its application to asymptotic normality of Whittle's estimate. *Prob. Theory and Related Fields* **86**, 87–104.
- Gnedenko, B. V. (1967), *The Theory of Probability* (translated by B. D. Seckler). Chelsea, New York.
- Goldberg, S. (1958), *Introduction to Difference Equations*. Wiley, New York.
- Gonzalez-Farias, G. M. (1992), A new unit root test for autoregressive time series. Ph.D. dissertation, North Carolina State University, Raleigh, North Carolina.
- Gonzalez-Farias, G., and Dickey, D. A. (1992), An unconditional maximum likelihood test for a unit root. *1992 Proceedings of the Business and Economic Statistics Section, Amer. Statist. Assoc.* 139–143.
- Gould, J. P., and Nelson, C. R. (1974), The stochastic structure of the velocity of money. *Amer. Econ. Rev.* **64**, 405–417.
- Gradshteyn, I. S., and Ryzhik, I. M. (1965), (4th ed.). *Tables of Integrals, Series and Products*. Academic Press, New York.
- Granger, C. W. J. (1980), Long memory relationships and the aggregation of dynamic models. *J. Econometrics* **14**, 227–238.
- Granger, C. W. J. (1981), Some properties of time series data and their use in econometric model specification. *J. Econometrics* **16**, 121–130.
- Granger, C. W. J., and Andersen, A. P. (1978), *An Introduction to Bilinear Time Series Models*. Vandenhoeck and Ruprecht, Göttingen.
- Granger, C. W. J., and Hatanaka, M. (1964), *Spectral Analysis of Economic Time Series*. Princeton University Press, Princeton, New Jersey.
- Granger, C. W. J., and Joyeux, R. (1980), An introduction to long-range time series models and fractional differencing. *J. Time Series Anal.* **1**, 15–30.

- Granger, C. W. J., and Teräsvirta, T. (1993), *Modelling Nonlinear Economic Relationships*. Oxford University Press, Oxford.
- Granger, C. W. J., and Weiss, A. A. (1983), Time series analysis of error correction models. In *Studies in Economic Time Series and Multivariate Statistics* (S. Karlin, T. Amemiya and L.A. Goodman, eds.), Academic Press, New York, 255–278.
- Grenander, U. (1950), Stochastic processes and statistical inference. *Ark. Mat.* **1**, 195–275.
- Grenander, U. (1954), On the estimation of regression coefficients in the case of an autocorrelated disturbance. *Ann. Math. Statist.* **25**, 252–272.
- Grenander, U. (1957), Modern trends in time series analysis. *Sankhyā* **18**, 149–158.
- Grenander, U., and Rosenblatt, M. (1957), *Statistical Analysis of Stationary Time Series*. Wiley, New York.
- Grether, D. M., and Nerlove, M. (1970), Some properties of “optimal” seasonal adjustment. *Econometrica* **38**, 682–703.
- Greville, T. N. E. (1969), *Theory and Applications of Spline Functions*. Academic Press, New York.
- Griliches, Z. (1967), Distributed lags: a survey. *Econometrica* **35**, 16–49.
- Grizzle, J. E., and Allen, D. M. (1969), Analysis of growth and dose response curves. *Biometrics* **25**, 357–381.
- Gumpertz, M. L., and Pantula, S. G. (1992), Nonlinear regression with variance components. *J. Amer. Statist. Assoc.* **87**, 201–209.
- Guo, L., Huang, D. W., and Hannan, E. J. (1990), On ARX( $\infty$ ) approximation. *J. Multivariate Anal.* **32**, 17–47.
- Hall, A. (1992), Testing for a unit root in time series with pretest data-based model selection. *J. Business and Econ. Statist.* **12**, 461–470.
- Hall, P., and Heyde, C. C. (1980), *Martingale Limit Theory and its Application*. Academic Press, New York.
- Hampel, F. R., Ronchetti, E. M., Rousseeuw, P. J., and Stahel, W. A. (1986), *Robust Statistics: The Approach Based on Influence Functions*. Wiley, New York.
- Hannan, E. J. (1956), The estimation of relationships involving distributed lags. *Econometrica* **33**, 206–224.
- Hannan, E. J. (1958), The estimation of spectral density after trend removal. *J. Roy. Statist. Soc. B* **20**, 323–333.
- Hannan, E. J. (1960), *Time Series Analysis*. Methuen, London.
- Hannan, E. J. (1961), A central limit theorem for systems of regressions. *Proc. Cambridge Phil. Soc.* **57**, 583–588.
- Hannan, E. J. (1963), The estimation of seasonal variation in economic time series. *J. Amer. Statist. Assoc.* **58**, 31–44.
- Hannan, E. J. (1969), A note on an exact test for trend and serial correlation. *Econometrica* **37**, 485–489.
- Hannan, E. J. (1970), *Multiple Time Series*. Wiley, New York.
- Hannan, E. J. (1971), Nonlinear time series regression, *J. Appl. Prob.* **8**, 767–780.
- Hannan, E. J. (1979), The central limit theorem for time series regression. *Stoch. Processes Appl.* **9**, 281–289.
- Hannan, E. J. (1980), The estimation of the order of an ARMA process. *Ann. Statist.* **8**, 1071–1081.

- Hannan, E. J. (1987), Rational transfer function approximation. *Statist. Sci.* **2**, 135–161.
- Hannan, E. J., and Deistler, M. (1988), *The Statistical Theory of Linear Systems*. Wiley, New York.
- Hannan, E. J., Dunsmuir, W. T. M., and Deistler, M. (1980), Estimation of vector ARMAX models. *J. Multivariate Anal.* **10**, 275–295.
- Hannan, E. J., and Heyde, C. C. (1972), On limit theorems for quadratic funtions of discrete time series. *Ann. Math. Statist.* **43**, 2058–2066.
- Hannan, E. J., and Kavalieris, L. (1984), A method for autoregressive-moving average estimation. *Biometrika* **72**, 273–280.
- Hannan, E. J., and Nicholls, D. F. (1972), The estimation of mixed regression, autoregression, moving average and distributed lag models. *Econometrica* **40**, 529–548.
- Hannan, E. J., and Quinn, B. G. (1979), The determination of the order of an autoregression. *J. Roy. Statist. Soc. B* **41**, 190–195.
- Hannan, E. J., and Rissanen, J. (1982), Recursive estimation of mixed autoregressive-moving average order. *Biometrika* **69**, 81–94. Correction (1983), **70**, 303.
- Hansen, J., and Lebedeff, S. (1987), Global trends of measured surface air temperature. *J. Geophys. Res.* **92**, 13345–13372.
- Hansen, J., and Lebedeff, S. (1988), Global surface air temperatures: update through 1987. *Geophys. Res. Lett.* **15**, 323–326.
- Hansen, M. H., Hurwitz, W. N., and Madow, W. G. (1953), *Sample Survey Methods and Theory, II*. Wiley, New York.
- Harris, B. (1967), *Spectral Analysis of Time Series*. Wiley, New York.
- Hart, B. I. (1942), Significance levels for the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **13**, 445–447.
- Hartley, H. O. (1961), The modified Gauss–Newton method for the fitting of nonlinear regression functions by least squares. *Technometrics* **3**, 269–280.
- Hartley, H. O., and Booker, A. (1965), Nonlinear least squares estimation. *Ann. Math. Statist.* **36**, 638–650.
- Harvey, A. C. (1984), A unified view of statistical forecasting procedures (with comments). *J. Forecasting* **3**, 245–283.
- Harvey, A. C. (1989). *Forecasting, Structural Time Series Models and the Kalman Filter*. Cambridge University Press, Cambridge.
- Harvey, A. C., and Durbin, J. (1986), The effects of seat belt legislation on British road casualties: a case study in structural time series modelling (with discussion). *J. Roy. Statist. Soc. A* **149**, 187–227.
- Harvey, A. C., and Pierse, R. G. (1984), Estimating missing observations in economic time series. *J. Amer. Statist. Assoc.* **79**, 125–131.
- Harvey, A. C., Ruiz, E., and Shephard, N. (1994), Multivariate stochastic variance models. *Rev. Econ. Stud.* **61**, 247–264.
- Harville, D. A. (1986), Using ordinary least squares software to compute combining intra-interblock estimates of treatment contrasts. *Amer. Statistician* **40**, 153–157.
- Haslett, J., and Raftery, A. E. (1989), Space–time modelling with long-memory dependence: assessing Ireland's wind power resource (with discussion). *J. Roy. Statist. Soc. C* **38**, 1–21.

- Hasza, D. P. (1977), Estimation in nonstationary time series. Ph.D. thesis, Iowa State University, Ames, Iowa.
- Hasza, D. P. (1980), A note on maximum likelihood estimation for the first-order autoregressive process. *Comm. Statist. Theory and Methods* **A9**, No. 13, 1411–1415.
- Hasza, D. P., and Fuller, W. A. (1979), Estimation for autoregressive processes with unit roots. *Ann. Statist.* **7**, 1106–1120.
- Hasza, D. P., and Fuller, W. A. (1982), Testing for nonstationary parameter specifications in seasonal time series models. *Ann. Statist.* **10**, 1209–1216.
- Hatanaka, M. (1973), On the existence and the approximation formulae for the moments of the  $k$ -class estimators. *Econ. Stud. Quart.* **24**, 1–15.
- Hatanaka, M. (1974), An efficient two-step estimator for the dynamic adjustment model with autoregressive errors. *J. Econometrics* **2**, 199–220.
- Herrndorf, N. (1984), A functional central limit theorem for weakly dependent sequences of random variables. *Ann. Prob.* **12**, 141–153.
- Higgins, M. L., and Bera, A. K. (1992), A Class of Nonlinear ARCH Models. *Int. Econ. Rev.* **33**, 137–158.
- Hildebrand, F. B. (1968), *Finite-Difference Equations and Simulations*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Hildreth, C. (1969), Asymptotic distribution of maximum likelihood estimators in a linear model with autoregressive disturbances. *Ann. Math. Statist.* **40**, 583–594.
- Hillmer, S. C., Bell, W. R., and Tiao, G. C. (1983), Modeling considerations in the seasonal adjustment of economic time series. In *Applied Time Series Analysis of Economic Data* (A. Zellner, ed.). United States Bureau of the Census, Washington.
- Hinich, M. (1982), Testing for Gaussianity and linearity of a time series. *J. Time Series Anal.* **3**, 169–176.
- Hoeffding, W., and Robbins, H. (1948), The central limit theorem for dependent random variables. *Duke Math. J.* **15**, 773–780.
- Hosking, J. R. M. (1981), Fractional differencing. *Biometrika* **68**, 165–176.
- Hosking, J. R. M. (1982), Some models of persistence in time series. In *Time Series Analysis: Theory and Practice 1* (O. D. Anderson, ed.). North-Holland, Amsterdam.
- Huang, D., and Anh, V. V. (1993), Estimation of the nonstationary factor in ARUMA models. *J. Time Series Anal.* **14**, 27–46.
- Huber, P. J. (1981), *Robust Statistics*. Wiley, New York.
- Ibragimov, I. A., and Linnik, Y. V. (1971), *Independent and Stationary Sequences of Random Variables*. Wolters-Noordhoff, Groningen.
- Jacquier, E., Polson, N. G., and Rossi, P. E. (1994), Bayesian analysis of stochastic volatility models. *J. Business and Econ. Statist.* **12**, 371–389.
- Jenkins, G. M. (1961), General considerations in the analysis of spectra. *Technometrics* **3**, 133–166.
- Jenkins, G. M., and Watts, D. G. (1968), *Spectral Analysis and its Application*. Holden-Day, San Francisco.
- Jennrich, R. I. (1969), Asymptotic properties of non-linear least squares estimators. *Ann. Math. Statist.* **40**, 633–643.
- Jobson, J. D. (1972), Estimation for linear models with unknown diagonal covariance matrix. Unpublished Ph.D. thesis, Iowa State University, Ames, Iowa.

- Jobson, J. D., and Fuller, W. A. (1980), Least squares estimation when the covariance matrix and the parameter vector are functionally related. *J. Amer. Statist. Assoc.* **75**, 176–181.
- Johansen, S. (1988), Statistical analysis of cointegration vectors. *J. Econ. Dynamics and Control* **12**, 231–254.
- Johansen, S. (1991), Estimation and hypothesis testing of cointegration vectors 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 Bull. Econ. and Statist.* **52**, 169–210.
- Johnston, J. (1984), *Econometric Methods* (3rd ed.). McGraw-Hill, New York.
- Jolley, L. B. W. (1961), *Summation of Series* (2nd ed.). Dover, New York.
- Jones, D. A. (1978), Nonlinear autoregressive processes. *Proc. Roy. Soc. A* **360**, 71–95.
- Jones, R. H. (1980), Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics* **22**, 389–395.
- Jorgenson, D. W. (1964), Minimum variance, linear, unbiased seasonal adjustment of economic time series. *J. Amer. Statist. Assoc.* **59**, 681–724.
- Judge, G. C., Griffiths, W. E., Hill, R. C., Lüthkepohl, H., and Lee, T-C. (1985), *The Theory and Practice of Econometrics*. Wiley, New York.
- Kalman, R. E. (1960), A new approach to linear filtering and prediction problems. *Trans. ASME J. Basic Eng.* **82**, 35–45.
- Kalman, R. E. (1963), New methods in Wiener filtering theory. In *Proceedings of the First Symposium on Engineering Applications of Random Function Theory and Probability* (J. L. Bodanoff and F. Kazin, eds.). Wiley, New York.
- Kalman, R. E., and Bucy, R. S. (1961), New results in linear filtering and prediction theory. *Trans. Amer. Soc. Mech. Eng. J. Basic Eng.* **83**, 95–108.
- Kawashima, H. (1980), Parameter estimation of autoregressive integrated processes by least squares. *Ann. Statist.* **8**, 423–435.
- Kempthorne, O., and Folks, L. (1971), *Probability, Statistics and Data Analysis*. Iowa State University Press, Ames, Iowa.
- Kendall, M. G. (1954), Note on bias in the estimation of autocorrelation. *Biometrika* **41**, 403–404.
- Kendall, M. G., and Stuart, A. (1966), *The Advanced Theory of Statistics* 3. Hafner, New York.
- King, M. L. (1980), Robust tests for spherical symmetry and their application to least squares regression. *Ann. Statist.* **8**, 1265–1271.
- King, M. L. (1988), Towards a theory of point optimal testing. *Econ. Rev.* **6**, 169–218.
- Kitigawa, G. (1987), Non-Gaussian state-space modelling of nonstationary time series (with discussion). *J. Amer. Statist. Assoc.* **82**, 1032–1063.
- Klimko, L. A., and Nelson, P. I. (1978), On conditional least squares estimation for stochastic processes. *Ann. Statist.* **6**, 629–642.
- Kohn, R., and Ansley, C. F. (1986), Estimation, prediction, and interpolation for ARIMA models with missing data. *J. Amer. Statist. Assoc.* **81**, 751–761.
- Koopmans, L. H. (1974), *The Spectral Analysis of Time Series*. Academic Press, New York.

- Koopmans, T. (1942), Serial correlation and quadratic forms in normal variables. *Ann. Math. Statist.* **13**, 14–33.
- Koopmans, T. C., Rubin, H., and Leipnik, R. B. (1950), Measuring the equation systems of dynamic economics. In *Statistical Inference in Dynamic Economic Models* (T. C. Koopmans, ed.). Wiley, New York.
- Kramer, K. H. (1963), Tables for constructing confidence limits on the multiple correlation coefficient. *J. Amer. Statist. Assoc.* **58**, 1082–1085.
- Krämer, W. (1986), Least squares regression when the independent variable follows an ARIMA process. *J. Amer. Statist. Assoc.* **81**, 150–154.
- Kreiss, J. P. (1987), On adaptive estimation in stationary ARMA processes. *Ann. Statist.* **15**, 112–133.
- Lai, T. L., and Siegmund, D. (1983), Fixed accuracy estimation of an autoregressive parameter. *Ann. Statist.* **11**, 478–485.
- Lai, T. L., and Wei, C. Z. (1982a), Asymptotic properties of projections with applications to stochastic regression problems. *J. Multivariate Anal.* **12**, 346–370.
- Lai, T. L., and Wei, C. Z. (1982b), Least squares estimates in stochastic regression models with applications to identification and control of dynamic systems. *Ann. Statist.* **10**, 154–166.
- Lai, T. L., and Wei, C. Z. (1982c), A note on martingale difference sequences satisfying the local Marcinkiewicz–Zygmund condition. Columbia University Tech. Report, New York.
- Lai, T. L., and Wei, C. Z. (1985a), Asymptotic properties of general autoregressive models and strong consistency of least squares estimates of their parameters. *J. Multivariate Anal.* **13**, 1–23.
- Lai, T. L., and Wei, C. Z. (1985b), Asymptotic properties of multivariate weighted sums with applications to stochastic regression in linear dynamic systems. In *Multivariate Analysis VI* (P. R. Krishnaiah, ed.). North-Holland, Amsterdam, 375–393.
- Lighthill, M. J. (1970), *Introduction to Fourier Analysis and Generalised Functions*. Cambridge University Press, Cambridge.
- Little, R. J. A., and Rubin, D. B. (1987), *Statistical Analysis with Missing Data*. Wiley, New York.
- Liviatan, N. (1963), Consistent estimation of distributed lags. *Int. Econ. Rev.* **4**, 44–52.
- Ljung, G. M. (1993), On outlier detection in time series. *J. Roy. Statist. Soc. B* **55**, 559–567.
- Ljung, G. M., and Box, G. E. P. (1978), On a measure of lack of fit in time series models. *Biometrika* **72**, 299–315.
- Loève, M. (1963), *Probability Theory* (3rd ed.). Van Nostrand, Princeton, New Jersey.
- Lomnicki, Z. A., and Zaremba, S. K. (1957), On estimating the spectral density function of a stochastic process. *J. Roy. Statist. Soc. B* **19**, 13–37.
- Lovell, M. C. (1963), Seasonal adjustment of economic time series and multiple regression analysis. *J. Amer. Statist. Assoc.* **58**, 993–1010.
- MacNeil, I. B. (1978), Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *Ann. Statist.* **6**, 422–433.

- Macpherson, B. D. (1981), Properties of estimators for the parameter of the first order moving average process. Ph.D. Thesis, Iowa State University, Ames, Iowa.
- Macpherson, B. D., and Fuller, W. A. (1983), Consistency of the least squares estimator of the first order moving average parameter. *Ann. Statist.* **11**, 326–329.
- Magnus, J. R., and Neudecker, H. (1988), *Matrix Differential Calculus with Applications in Statistics and Econometrics*. Wiley, New York.
- Malinvaud, E. (1970a), *Statistical Methods in Econometrics*. North-Holland, Amsterdam.
- Malinvaud, E. (1970b), The consistency of non-linear regressions. *Ann. Math. Statist.* **41**, 956–969.
- Mallows, C. L. (1973), Some comments on  $C_p$ . *Technometrics* **15**, 661–675.
- Mandelbrot, B.B. (1969), Long-run linearity, locally Gaussian process, H-spectra and infinite variances. *Int. Econ. Rev.* **10**, 82–111.
- Mandelbrot, B. B., and Van Ness, J. W. (1968), Fractional Brownian motions, fractional noises and applications. *SIAM Rev.* **10**, 422–437.
- Mann, H. B., and Wald, A. (1943a), On the statistical treatment of linear stochastic difference equations. *Econometrica* **11**, 173–220.
- Mann, H. B., and Wald, A. (1943b), On stochastic limit and order relationships. *Ann. Math. Statist.* **14**, 217–226.
- Marden, M. (1966), *Geometry of Polynomials*. Amer. Math. Soc., Providence, Rhode Island.
- Mariott, F. H. C., and Pope, J. A. (1954), Bias in the estimation of autocorrelations. *Biometrika* **41**, 390–402.
- Marple, S. L. (1987), *Digital Spectral Analysis with Applications*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Martin, R. D. (1980), Robust estimation of autoregressive models. In *Directions in Time Series* (D. R. Brillinger and G. C. Tiao, eds.), Institute of Mathematical Statistics, Hayward, California, 228–254.
- Martin, R. D., and Yohai, V. J. (1986), Influence functionals for time series. *Ann. Statist.* **14**, 781–818.
- McClave, J. T. (1974), A comparison of moving average estimation procedures. *Comm. Statist.* **3**, 865–883.
- McDougall, A. J. (1994), Robust methods for autoregressive moving average estimation. *J. Roy. Statist. Soc. B* **56**, 189–207.
- McLeish, D. L. (1974), Dependent central limit theorems and invariance principles. *Ann. Prob.* **2**, 620–628.
- McLeish, D. L. (1975), A maximal inequality and dependent strong laws. *Ann. Prob.* **3**, 829–839.
- McNeill, I. B. (1978), Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times. *Ann. Statist.* **6**, 422–433.
- Meinholt, R. J., and Singpurwalla, N. D. (1983), Understanding the Kalman filter. *Amer. Statistician* **37**, 123–127.
- Melino, A., and Turnbull, S. (1990), Pricing foreign currency options with stochastic volatility. *J. Econometrics* **45**, 239–265.
- Miller, K. S. (1963), *Linear Differential Equations in the Real Domain*. Norton, New York.

- Miller, K. S. (1968), *Linear Difference Equations*. Benajmin, New York.
- Moran, P. A. P. (1947), Some theorems on time series I. *Biometrika* **34**, 281–291.
- Moran, P. A. P. (1953), The statistical analysis of the Canadian lynx cycle, I: structure and prediction. *Australian J. Zool.* **1**, 163–173.
- Morettin, P. A. (1984), The Levinson algorithm and its applications in time series analysis. *Int. Statist. Rev.* **52**, 83–92.
- Muirhead, C. R. (1986), Distinguishing outlier types in time series. *J. Roy. Statist. Soc. B* **48**, 39–47.
- Muirhead, R. J. (1982), *Aspects of Multivariate Statistical Theory*. Wiley, New York.
- Nabeya, S., and Tanaka, K. (1990a), A general approach to the limiting distribution for estimators in time series regression with nonstable autoregressive errors. *Econometrica* **58**, 145–163.
- Nabeya, S., and Tanaka, K. (1990b), Limiting powers of unit root tests in time series regression. *J. Econometrics* **46**, 247–271.
- Nagaraj, N. K. (1990), Results on estimation for vector unit root processes. Unpublished manuscript, Iowa State University, Ames, Iowa.
- Nagaraj, N. K., and Fuller, W. A. (1991), Estimation of the parameters of linear time series models subject to nonlinear restrictions. *Ann. Statist.* **19**, 1143–1154.
- Nagaraj, N. K., and Fuller, W. A. (1992), Least squares estimation of the linear model with autoregressive errors. In *New Directions in Time Series Analysis, Part I* (D. Brillinger et al., eds.), IMA Volumes in Mathematics and Its Applications, Vol. 45, Springer, New York, 215–225.
- Nankervis, J. C., and Savin, N. E. (1988), The Student's *t* approximation in a stationary first order autoregressive model. *Econometrica* **56**, 119–145.
- Narasimham, G. V. L. (1969), Some properties of estimators occurring in the theory of linear stochastic processes. In *Lecture Notes in Operations Research and Mathematical Economics* (M. Beckman and H. P. Künzi, eds.). Springer, Berlin.
- Nelson, C. R. (1974), The first-order moving average process: identification, estimation and prediction. *J. Econometrics* **2**, 121–141.
- Nelson, C. R., and Plosser, C. I. (1982), Trends and random walks in macroeconomic time series: some evidence and implications. *J. Monetary Econ.* **10**, 129–162.
- Nerlove, M. (1964a), Notes on Fourier series and analysis for economists, Multilith prepared under Grant NSF-GS-142 to Stanford University, Palo Alto, California.
- Nerlove, M. (1964b), Spectral analysis of seasonal adjustment procedures. *Econometrica* **32**, 241–286.
- Newbold, P. (1974), The exact likelihood function for a mixed autoregressive-moving average process. *Biometrika* **61**, 423–426.
- Newey, W. K., and West, K. D. (1987), A simple, positive definite, heteroscedasticity and autocorrelation consistent covariance matrix. *Econometrica* **55**, 703–708.
- Newton, H. J. (1988), *Timeslab: A Time Series Analysis Laboratory*. Wadsworth and Brooks/Cole, Pacific Grove, California.
- Newton, H. J., and Pagano, M. (1984), Simultaneous confidence bands for autoregressive spectra. *Biometrika* **71**, 197–202.
- Nicholls, D. F. (1976), The efficient estimator of vector linear time series models. *Biometrika* **63**, 381–390.

- Nicholls, D. F., Pagan, A. R., and Terrell, R. D. (1975), The estimation and use of models with moving average disturbance terms: a survey. *Int. Econ. Rev.* **16**, 113–134.
- Nicholls, D. F., and Quinn, B. G. (1982), *Random Coefficient Autoregressive Models: An Introduction*, Springer Lecture Notes in Statistics, No. 11. Springer, New York.
- Nold, F. C. (1972), A bibliography of applications of techniques of spectral analysis to economic time series. Technical Report No. 66, Institute for Mathematical Studies in the Social Sciences, Stanford University, Stanford, California.
- Nyblom, J., and Makelainen, T. (1983), Comparisons of tests for the presence of random walk coefficients in a simple linear model. *J. Amer. Statist. Assoc.* **78**, 856–864.
- Olshen, R. A. (1967), Asymptotic properties of the periodogram of a discrete stationary process. *J. Appl. Prob.* **4**, 508–528.
- Orcutt, G. H., and Winokur, H. S. (1969), First order autoregression: inference, estimation, and prediction. *Econometrica* **37**, 1–14.
- Otnes, R. K., and Enochson, L. (1972), *Digital Time Series Analysis*. Wiley, New York.
- Otto, M. C., and Bell, W. R. (1990), Two issues in time series outlier detection using indicator variables. *1990 Proc. Business and Economics Section, Amer. Statist. Assoc.*, 182–187.
- Ozaki, T. (1980), Non-linear threshold models for non-linear random vibrations. *J. Appl. Prob.* **17**, 84–93.
- Pantula, S. G. (1982), Properties of the parameters of auto-regressive time series. Ph.D. Thesis, Iowa State University, Ames, Iowa.
- Pantula, S. G. (1986), On asymptotic properties of the least squares estimators for autoregressive time series with a unit root. *Sankhyā Ser. A* **48**, 208–218.
- Pantula, S. G. (1988a), Statistical time series. Unpublished manuscript for Statistics 613, North Carolina State University, Raleigh, North Carolina.
- Pantula, S. G. (1988b), Estimation of autoregressive models with ARCH errors. *Sankhyā Ser. B*, **50**, 119–138.
- Pantula, S. G. (1989), Testing for unit roots in time series data. *Econometric Theory* **5**, 256–271.
- Pantula, S. G., and Fuller, W. A. (1985), Mean estimation bias in least squares estimation of autoregressive processes. *J. Econometrics* **27**, 99–121.
- Pantula, S. G., and Fuller, W. A. (1993), The large sample distribution of the roots of the second order autoregressive polynomial. *Biometrika* **80**, 919–923.
- Pantula, S. G., Gonzalez-Farias, G., and Fuller, W. A. (1994), A comparison of unit root criteria. *J. Business and Econ. Statist.* **13**, 449–459.
- Park, H. J. (1990), Alternative estimators of the parameters of the autoregressive process. Ph.D. dissertation, Iowa State University, Ames, Iowa.
- Park, H. J., and Fuller, W. A. (1995), Alternative estimators and unit root tests for the autoregressive process. *J. Time Series Anal.* **16**, 415–429.
- Park, J., and Phillips, P. C. B. (1988), Statistical inference in regressions with integrated processes: Part I. *Econometric Theory* **4**, 468–497.
- Park, J., and Phillips, P. C. B. (1989), Statistical inference in regressions with integrated processes: Part II. *Econometric Theory* **5**, 95–131.
- Parzen, E. (1957), On consistent estimates of the spectrum of a stationary time series. *Ann. Math. Statist.* **28**, 329–348.

- Parzen, E. (1958), Conditions that a stochastic process be ergodic. *Ann. Math. Statist.* **29**, 299–301.
- Parzen, E. (1961), Mathematical considerations in the estimation of spectra. *Technometrics* **3**, 167–190.
- Parzen, E. (1962), *Stochastic Processes*. Holden-Day, San Francisco.
- Parzen, E. (1969), Multiple time series modeling. In *Multivariate Analysis II* (P. R. Krishnaiah, ed.), Academic Press, New York.
- Parzen, E. (1974), Some recent advances in time series modeling. *IEEE Trans. Automatic Control AC-19*, 723–729.
- Parzen, E. (1977), Multiple time series modeling: determining the order of the approximating autoregressive scheme. In *Multivariate Analysis IV* (P. R. Krishnaiah, ed.). North Holland, Amsterdam, 283–295.
- Peña, D., and Guttman, I. (1989), Bayesian approach to robustifying the Kalman filter. In *Bayesian Analysis of Time Series and Dynamic Models* (J. C. Spall, ed.), Marcel Dekker, New York, 227–253.
- Perron, P. (1989), The calculation of the limiting distribution of the least-squares estimator in a near-integrated model. *Econometric Theory* **5**, 241–255.
- Phillips, P. C. B. (1979), The sampling distribution of forecasts from a first-order autoregression. *J. Econometrics* **9**, 241–262.
- Phillips, P. C. B. (1986), Understanding spurious regressions in econometrics. *J. 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 of autoregression. *Biometrika* **74**, 535–548.
- Phillips, P. C. B. (1987c), Asymptotic expansions in nonstationary vector autoregressions. *Econometric Theory* **2**, 45–68.
- Phillips, P. C. B. (1988), Weak convergence to the matrix stochastic integral  $\int_0^1 B dB'$ . *J. Multivariate Anal.* **24**, 252–264.
- Phillips, P. C. B. (1991), Optimal inference in cointegrated systems. *Econometrica* **59**, 283–306.
- Phillips, P. C. B. (1994), Some exact distribution theory for maximum likelihood estimators of cointegrating coefficients in error correction models. *Econometrica* **62**, 73–93.
- Phillips, P. C. B., and Durlauf, S. N. (1986), Multiple time series regression with integrated processes. *Rev. Econ. Stud.* **53**, 473–496.
- Phillips, P. C. B., and Hansen, B. E. (1990), Statistical inference in instrumental variables regression with  $I(1)$  processes. *Rev. Econ. Stud.* **57**, 99–125.
- Phillips, P. C. B., and Ouliaris, S. (1990), Asymptotic properties of residual based tests for cointegration. *Econometrica* **58**, 165–193.
- Phillips, P. C. B., and Perron, P. (1988), Testing for a unit root in time series regression. *Biometrika* **75**, 335–346.
- Phillips, P. C. B., and Solo, V. (1992), Asymptotics for linear processes. *Ann. Statist.* **20**, 971–1001.

- Pierce, D. A. (1970a), A duality between autoregressive and moving average processes concerning their least squares parameter estimates. *Ann. Math. Statist.* **41**, 422–426.
- Pierce, D. A. (1970b), Rational distributed lag models with autoregressive–moving average errors. *1970 Proc. Business and Economics Statistics Section, Amer. Statist. Assoc.* 591–596.
- Pierce, D. A. (1971), Least squares estimation in the regression model with autoregressive–moving average errors. *Biometrika* **58**, 299–312.
- Pierce, D. A. (1972), Least squares estimation in dynamic-disturbance time series models. *Biometrika* **59**, 73–78.
- Plosser, C. I., and Schwert, G. W. (1977), Estimation of noninvertible moving average process: the case of overdifferencing. *J. Econometrics* **6**, 199–224.
- Poirier, D. J. (1973), Piecewise regression using cubic splines. *J. Amer. Statist. Assoc.* **68**, 515–524.
- Pollard, D. (1984), *Convergence of Stochastic Processes*. Springer, New York.
- Potscher, B. M., and Prucha, I. R. (1991a), Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part I: consistency and approximation concepts. *Econometric Rev.* **10**, 125–216.
- Potscher, B. M., and Prucha, I. R. (1991b), Basic structure of the asymptotic theory in dynamic nonlinear econometric models, part II: asymptotic normality. *Econometric Rev.* **10**, 253–325.
- Pratt, J. W. (1959), On a general concept of “in probability.” *Ann. Math. Statist.* **30**, 549–558.
- Prest, A. R. (1949), Some experiments in demand analysis. *Rev. Econ. and Statist.* **31**, 33–49.
- Priestly, M. B. (1981), *Spectral Analysis and Time Series*. Academic Press, New York.
- Priestly, M. B. (1988), *Non-linear and Non-stationary Time Series Analysis*. Academic Press, San Diego.
- Quenouille, M. H. (1957), *The Analysis of Multiple Time Series*. Hafner, New York.
- Quinn, B. G. (1982), A note on the existence of strictly stationary solutions to bilinear equations. *J. Times Series Anal.* **3**, 249–252.
- Rao, C. R. (1959), Some problems involving linear hypothesis in multivariate analysis. *Biometrika* **46**, 49–58.
- Rao, C. R. (1965), The theory of least squares when the parameters are stochastic and its application to the analysis of growth curves. *Biometrika* **52**, 447–458.
- Rao, C. R. (1967), Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In *Fifth Berkeley Symp. Math. Statist. Prob.* **1**, 355–372.
- Rao, C. R. (1973), *Linear Statistical Inference and its Applications* (2nd ed.). Wiley, New York.
- Rao, J. N. K., and Tintner, G. (1963). On the variate difference method. *Australian J. Statist.* **5**, 106–116.
- Rao, M. M. (1961), Consistency and limit distributions of estimators of parameters in explosive stochastic difference equations. *Ann. Math. Statist.* **32**, 195–218.
- Rao, M. M. (1978a), Asymptotic distribution of an estimator of the boundary parameter of an unstable process. *Ann. Statist.* **6**, 185–190. Correction (1980), **8**, 1403.

- Rao, M. M. (1978b), Covariance analysis of nonstationary time series. In *Developments in Statistics* (P. R. Krishnaiah, ed.). Academic Press, New York.
- Reeves, J. E. (1972), The distribution of the maximum likelihood estimator of the parameter in the first-order autoregressive series. *Biometrika* **59**, 387–394.
- Reinsel, G. C. (1993), *Elements of Multivariate Time Series Analysis*. Springer, New York.
- Reinsel, G. C., and Ahn, S. K. (1989), Asymptotic distribution of the likelihood ratio test for cointegration in the nonstationary vector AR model. Unpublished report, University of Wisconsin, Madison.
- Robinson, P. M. (1972), Nonlinear regression for multiple time series. *J. Appl. Prob.* **9**, 758–768.
- Robinson, P. M. (1994a), Semiparametric analysis of long-memory time series. *Ann. Statist.* **22**, 515–539.
- Robinson, P. M. (1994b), Efficient tests of nonstationary hypotheses. *J. Amer. Statist. Assoc.* **89**, 1420–1437.
- Rogosinski, W. (1959), *Fourier Series* (translated by H. Cohn and F. Steinhardt). Chelsea, New York.
- Rosenblatt, M. (ed.) (1963), *Proc. Symp. Time Series Analysis*, Brown University, June 11–14, 1962, Wiley, New York.
- Rothenberg, T. J. (1984), Approximate normality of generalized least squares estimates. *Econometrica* **52**, 811–825.
- Royden, H. L. (1989), *Real Analysis* (3rd ed.). Macmillan, New York.
- Rubin, D. B. (1976), Inference and missing data. *Biometrika* **63**, 581–592.
- Rubin, H. (1950), Consistency of maximum-likelihood estimates in the explosive case. In *Statistical Inference in Dynamic Economic Models* (T. C. Koopmans, ed.). Wiley, New York.
- Rutherford, D. E. (1946), Some Continuant Determinants Arising in Physics and Chemistry. *Proceedings of the Royal Society of Edinburgh, Sect. A.* **62**, 229–236.
- Sage, A. P. (1968), *Optimum Systems Control*. Prentice-Hall, Englewood Cliffs, New Jersey.
- Said, E. S., and Dickey, D. A. (1984), Testing for unit roots in autoregressive-moving average models of unknown order. *Biometrika* **71**, 599–607.
- Saikkonen, P. and Luukonen, R. (1993), Testing for a moving average unit root in autoregressive integrated moving average models. *J. Amer. Statist. Assoc.* **88**, 596–601.
- Salem, A. S. (1971), Investigation of alternative estimators of the parameters of autoregressive processes. Unpublished M.S. thesis, Iowa State University, Ames, Iowa.
- Sallas, W. M., and Harville, D. A. (1981), Best linear recursive estimation for mixed linear models. *J. Amer. Statist. Assoc.* **76**, 860–869.
- Samaranayake, V. A., and Hasza, D. P. (1983), The asymptotic properties of the sample autocorrelations for a multiple autoregressive process with one unit root. University of Missouri, Rolla, Missouri.
- Sanger, T. M. (1992), Estimated generalized least squares estimation for the heterogeneous measurement error model. Ph.D. thesis, Iowa State University, Ames, Iowa.
- Sargan, J. D. (1958), The estimation of economic relationships using instrumental variables. *Econometrica* **26**, 393–415.

- Sargan, J. D. (1959), The estimation of relationships with autocorrelated residuals by the use of instrumental variables. *J. Roy. Statist. Soc. B* **21**, 91–105.
- Sargan, J. D., and Bhargava, A. (1983), Testing residuals from least squares regression for being generated by the Gaussian random walk. *Econometrica* **51**, 153–174.
- Sarkar, S. (1990), Nonlinear least squares estimators with differential rates of convergence. Ph.D. thesis, Iowa State University, Ames, Iowa.
- SAS (1984), *SAS/ETS User's Guide*, Version 5 ed., SAS Institute, Cary, N.C.
- SAS Institute Inc. (1989), *SAS/STAT User's Guide*, Version 6 (4th ed.), Vol. 2. SAS Institute, Cary, North Carolina.
- Scheffé, H. (1970), Multiple testing versus multiple estimation. Improper confidence sets. Estimation of directions and ratios. *Ann. Math. Statist.* **41**, 1–29.
- Schmidt, P., and Phillips, P. C. B. (1992), Testing for a unit root in the presence of deterministic trends. *Oxford Bull. Econ. and Statist.* **54**, 257–288.
- Schwartz, G. (1978), Estimating the dimension of a model. *Ann. Statist.* **6**, 461–464.
- Schwartz, G. W. (1989), Tests for unit roots: a Monte Carlo investigation. *J. Business and Econ. Statist.* **7**, 147–160.
- Scott, D. J. (1973), Central limit theorems for martingales and for processes with stationary increments using a Skorokhod representation approach. *Adv. Appl. Prob.* **5**, 119–137.
- Sen, D. L., and Dickey, D. A. (1987), Symmetric test for second differencing in univariate time series. *J. Business and Econ. Statist.* **5**, 463–473.
- Shaman, P., and Stine, R. A. (1988), The bias of autoregressive coefficient estimators. *J. Amer. Statist. Assoc.* **83**, 842–848.
- Shibata, R. (1976), Selection of the order of an autoregressive model by Akaike's information criterion. *Biometrika* **63**, 117–126.
- Shibata, R. (1986), Consistency of model selection and parameter estimation. In *Essays in Time Series and Allied Processes* (J. M. Gani and M. B. Priestley, eds.), Appl. Prob. Trust, Sheffield, 126–141.
- Shin, D. (1990), Estimation for the autoregressive moving average model with a unit root. Ph.D. thesis, Iowa State University, Ames, Iowa.
- Shin, D., and Fuller, W. A. (1991), Estimation for the autoregressive moving average with an autoregressive unit root. Unpublished manuscript, Iowa State University, Ames, Iowa.
- Shively, T. S. (1988), An exact test for a stochastic coefficient in a time series regression model. *J. Time Series Anal.* **9**, 81–88.
- Sims, C. A., Stock, J. H., and Watson, M. W. (1990), Inference in linear time series models with some unit roots. *Econometrica* **58**, 113–144.
- Singleton, R. C. (1969), An algorithm for computing the mixed radix fast Fourier transform. *IEEE Trans. Audio and Electroacoustics AU-17*, No. 2, 93–103.
- Solo, V. (1984), The order of differencing in ARIMA models. *J. Amer. Statist. Assoc.* **79**, 916–921.
- Sorenson, H. W. (1966), Kalman filtering techniques. In *Advances in Control Systems* **3** (D. T. Leondis, ed.). Academic Press, New York.
- Stephenson, J. A., and Farr, H. T. (1972), Seasonal adjustment of economic data by application of the general linear statistical model. *J. Amer. Statist. Assoc.* **67**, 37–45.

- Stigum, B. P. (1974), Asymptotic properties of dynamic stochastic parameter estimates (III). *J. Multivariate Anal.* **4**, 351–381.
- Stigum, B. P. (1975), Asymptotic properties of autoregressive integrated moving average processes. *Stochastic Properties Appl.* **3**, 315–344.
- Stigum, B. P. (1976), Least squares and stochastic difference equations. *J. Econometrics* **4**, 349–370.
- Stine, R. A., and Shaman, P. (1990), Bias of autoregressive spectral estimators. *J. Amer. Statist. Assoc.* **85**, 1091–1098.
- Stock, J. H. (1987), Asymptotic properties of least squares estimators of cointegrating vectors. *Econometrica* **55**, 1035–1056.
- Stock, J. H., and Wason, W. W. (1988), Testing for common trends. *J. Amer. Statist. Assoc.* **83**, 1097–1107.
- Stout, W. F. (1968), Some results on the complete and almost sure convergence of linear combinations of independent variables and martingale differences. *Ann. Math. Statist.* **39**, 1549–1562.
- Strasser, H. (1986), Martingale difference arrays and stochastic integrals. *Prob. Theory and Related Fields* **72**, 83–98.
- Stuart, A., and Ord, J. K. (1987), *Kendall's Advanced Theory of Statistics*, Vol. 1. Oxford University Press, New York.
- Stuart, A., and Ord, J. K. (1991), *Kendall's Advanced Theory of Statistics*, Vol. 2. Oxford University Press, New York.
- Subba Rao, T., and Gabr, M. M. (1984), *An Introduction to Bispectral Analysis and Bilinear Time Series Models*. Springer Lecture Notes in Statistics, Springer, New York.
- Sweeting, T. J. (1980), Uniform asymptotic normality of the maximum likelihood estimator. *Ann. Statist.* **8**, 1375–1381.
- Tanaka, K. (1990), Testing for a moving average unit root. *Econometric Theory* **6**, 433–444.
- Taqqu, M. S. (1988), Self-similar processes. In *Encyclopedia of Statistical Sciences*, Vol. 8. Wiley, New York, 352–357.
- Theil, H. (1961), *Economic Forecasts and Policy*. North-Holland, Amsterdam.
- Thompson, L. M. (1969), Weather and technology in the production of wheat in the United States. *J. Soil and Water Conservation* **24**, No. 6, 219–224.
- Thompson, L. M. (1973), Cyclical weather patterns in the middle latitudes. *J. Soil and Water Conservation* **28**, No. 2, 87–89.
- Tiao, G. C., and Tsay, R. S. (1983), Consistency properties of least squares estimates of autoregressive parameters in ARMA models. *Ann. Statist.* **11**, 856–871.
- Tintner, G. (1940), *The Variate Difference Method*. Principia Press, Bloomington, Indiana.
- Tintner, G. (1952), *Econometrics*. Wiley, New York.
- Tolstov, G. P. (1962), *Fourier Series* (translated by R. A. Silverman). Prentice-Hall, Englewood Cliffs, New Jersey.
- Tong, H. (1981), A note on a Markov bilinear stochastic process in discrete time. *J. Time Series Anal.* **2**, 279–284.
- Tong, H. (1983), *Threshold Models in Non-linear Time Series Analysis*. Lecture Notes in Statistics, No. 21. Springer, Heidelberg.

- Tong, H. (1990), *Non-linear Time Series: A Dynamical System Approach*. Oxford University Press, New York.
- Torres, V. A. (1986), A note on teaching infinite moving averages. *Amer. Statistician* **40**, 40–41.
- Toyooka, Y. (1982), Second-order expansion of mean squared error matrix of generalized least squares estimator with estimated parameters. *Biometrika* **69**, 269–273.
- Toyooka, Y. (1990), No estimation effect up to the second order for the covariance parameter of GLSE and MLE in a regression with a general covariance structure. *Math. Japonica* **35**, 745–757.
- Tsay, R. S. (1986), Nonlinearity tests for time series. *Biometrika* **73**, 461–466.
- Tsay, R. S. (1988), Outliers, level shifts and variance changes in time series. *J. Forecasting* **7**, 1–20.
- Tsay, R. S. (1993), Testing for noninvertible models with applications. *J. Business and Econ. Statist.* **11**, 225–233.
- Tucker, H. G. (1967), *A Graduate Course in Probability*. Academic Press, New York.
- Tukey, J. W. (1961), Discussion, emphasizing the connection between analysis of variance and spectrum analysis. *Technometrics* **3**, 1–29.
- Ulyrych, T. J., and Bishop, T. N. (1975), Maximum entropy spectral analysis and autoregressive decomposition. *Rev. Geophys. and Space Phys.* **13**, 183–200.
- Van der Genugten, B. B. (1983), The asymptotic behavior of the estimated generalized least squares method in the linear regression model. *Statist. Neerlandica* **37**, 127–141.
- Varadarajan, V. S. (1958), A useful convergence theorem. *Sankhyā* **20**, 221–222.
- Venkataraman, K. N. (1967), A note on the least square estimators of the parameters of a second order linear stochastic difference equation. *Calcutta Statist. Assoc. Bull.* **16**, 15–28.
- Venkataraman, K. N. (1968), Some limit theorems on a linear stochastic difference equation with a constant term, and their statistical applications. *Sankhyā Ser. A* **30**, 51–74.
- Venkataraman, K. N. (1973), Some convergence theorems on a second order linear explosive stochastic difference equation with a constant term. *J. Indian Statist. Assoc.* **11**, 47–69.
- von Neumann, J. (1941), Distribution of ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **12**, 367–395.
- von Neumann, J. (1942), A further remark concerning the distribution of the ratio of the mean square successive difference to the variance. *Ann. Math. Statist.* **13**, 86–88.
- Wahba, G. (1968), On the distribution of some statistics useful in the analysis of jointly stationary time series. *Ann. Math. Statist.* **39**, 1849–1862.
- Wahba, G. (1990), *Spline Models for Observational Data*. SIAM, Philadelphia.
- Wald, A. (1949), Note on the consistency of the maximum likelihood estimate. *Ann. Math. Statist.* **20**, 595–601.
- Walker, A. M. (1961), Large-sample estimation of parameters for moving-average models. *Biometrika* **48**, 343–357.
- Walker, A. M. (1964), Asymptotic properties of least-squares estimates of the spectrum of a stationary non-deterministic time series. *J. Australian Math. Soc.* **4**, 363–384.
- Wallis, K. F. (1967), Lagged dependent variables and serially correlated errors: a reappraisal of three pass least squares. *Rev. Econ. Statist.* **49**, 555–567.

- Wallis, K. F. (1972), Testing for fourth order autocorrelation in quarterly regression equations. *Econometrica* **40**, 617–638.
- Watson, G. S. (1955), Serial correlation in regression analysis I. *Biometrika* **42**, 327–341.
- Watson, G. S., and Hannan, E. J. (1956), Serial correlation in regression analysis II. *Biometrika* **43**, 436–445.
- Wegman, E. J. (1974), Some results on nonstationary first order autoregression. *Technometrics* **16**, 321–322.
- Weiss, A. A. (1984), ARMA models with ARCH errors. *J. Time Series Anal.* **5**, 129–143.
- Weiss, A. A. (1986), ARCH and bilinear time series models: comparison and combination. *J. Business and Econ. Statist.* **4**, 59–70.
- Weiss, L. (1971), Asymptotic properties of maximum likelihood estimators in some nonstandard cases. *J. Amer. Statist. Assoc.* **66**, 345–350.
- Weiss, L. (1973), Asymptotic properties of maximum likelihood estimators in some nonstandard cases, II. *J. Amer. Statist. Assoc.* **68**, 428–430.
- White, H., and Domowitz, I. (1984), Nonlinear regression with dependent observations. *Econometrica* **52**, 143–161.
- White, J. S. (1958), The limiting distribution of the serial correlation coefficient in the explosive case. *Ann. Math. Statist.* **29**, 1188–1197.
- White, J. S. (1959), The limiting distribution of the serial correlation coefficient in the explosive case II. *Ann. Math. Statist.* **30**, 831–834.
- Whittle, P. (1951), *Hypothesis Testing in Time Series Analysis*. Almqvist and Wiksell, Uppsala.
- Whittle, P. (1953), The analysis of multiple stationary time series. *J. Roy. Statist. Soc. B* **15**, 125–139.
- Whittle, P. (1963), *Prediction and Regulation by Linear Least-Square Methods*. Van Nostrand, Princeton, New Jersey.
- Wilks, S. S. (1962), *Mathematical Statistics*. Wiley, New York.
- Williams, E. J. (1959), *Regression Analysis*. Wiley, New York.
- Williamson, M. (1972), *The Analysis of Biological Populations*. Edward Arnold, London.
- Wilson, G. T. (1969), Factorization of the generating function of a pure moving average process. *SIAM J. Numer. Anal.* **6**, 1–7.
- Winckel, M. A., and Reinsel, G. C. (1986), An exact maximum likelihood estimation procedure for regression–ARMA time series models with possibly nonconsecutive data. *J. Roy. Statist. Soc. B* **48**, 303–313.
- Withers, C. S. (1981), Conditions for linear processes to be strong mixing. *Z. Wahr. verw. Geb.* **57**, 477–480.
- Wold, H. O. A. (1938), *A Study in the Analysis of Stationary Time Series* (2nd ed., 1954), Almqvist and Wiksell, Uppsala.
- Wold, H. O. A. (1965), *Bibliography on Time Series and Stochastic Processes*. Oliver and Boyd, Edinburgh.
- Woodward, W. A., and Gray, H. L. (1991). “Global warming” and the problem of testing for trend in time series data. Department of Statistical Science, Southern Methodist University.

- Wu, C. F. (1981), Asymptotic theory for nonlinear least squares estimation. *Ann. Math. Statist.* **9**, 501–513.
- Wu, L. S.-Y., Hosking, J. R. M., and Ravishanker, N. (1993), Reallocation outliers in time series. *Appl. Statist.* **42**, 301–313.
- Yaglom, A. M. (1962), *An Introduction to the Theory of Stationary Random Functions* (translated by R.A. Silverman). Prentice-Hall, Englewood Cliffs, New Jersey.
- Yajima, Y. (1985), On estimation of long-memory time series models. *Australian J. Statist.* **27**, 303–320.
- Yajima, Y. (1988), On estimation of a regression model with long-memory stationary errors. *Ann. Statist.* **16**, 791–807.
- Yajima, Y. (1989), A central limit theorem of Fourier transforms of strongly dependent stationary processes. *J. Time Series Anal.* **10**, 375–383.
- Yajima, Y. (1991), Asymptotic properties of the LSE in a regression model with long-memory stationary errors. *Ann. Statist.* **19**, 158–177.
- Yamamoto, T. (1976), Asymptotic mean square prediction error for an autoregressive model with estimated coefficients. *J. Appl. Statist.* **25**, 123–127.
- Yap, S. F., and Reinsel, G. C. (1995a), Results on estimation and testing for a unit root in the nonstationary autoregressive moving-average model. *J. Time Series Anal.* **16**, 339–353.
- Yap, S. F., and Reinsel, G. C. (1995b), Estimation and testing for unit roots in a partially nonstationary vector autoregressive moving average model. *J. Amer. Statist. Assoc.* **90**, 253–267.
- Yule, G. U. (1927), On a method of investigating periodicities in distributed series with special reference to Wolfer's sunspot numbers. *Philos. Trans. Roy. Soc. London Ser. A* **226**, 267–298.
- Zygmund, A. (1959), *Trigonometric Series*. Cambridge University Press, Cambridge.

# Index

- Absolutely summable sequence, 28  
of matrices, 76  
*see also* Sequence
- ACF, *see* Autocorrelation function
- Additive outlier, 460
- AIC, 438
- Alias, 136
- Amplitude, 14
- Amplitude spectrum, 174. *See also* Cross amplitude spectrum
- Analysis of variance, 356
- AR, *see* Autoregressive time series
- ARCH, 110 (Exercise 43), 494
- ARIMA, 101
- Arithmetic means, method of, 129. *See also* Cesáro summability
- ARMA, 70. *See also* Autoregressive moving average time series
- Associated polynomial, *see* Characteristic equation
- Autocorrelation function, 7–10  
definition, 7  
estimation, 317  
bias, 317  
covariance of estimators, 317  
distribution of estimators, 335  
least squares residuals, 484–486
- partial, 10–12, 24, 27
- spectral representation, 127
- for vector valued time series, 17  
*see also* Autocovariance
- Autocovariance, 4  
of autoregressive moving average time series, 72
- of autoregressive time series, 40, 55, 62  
difference equation for, 55, 62
- of complex valued time series, 12
- estimation, 313
- bias, 316  
consistency, 331, 343  
covariance of estimators, 314–315  
distribution, 333  
least squares residuals, 484–486  
variance of, 314–315
- of moving average time series, 22  
properties of, 7–10
- of second order autoregressive process, 55
- spectral representation, 127
- of vector process, 15–17
- Autoregression, on residuals, 519–520. *See also* Autoregressive time series, estimation
- Autoregressive moving average time series, 70–75  
autocorrelation function, 72  
autoregressive representation, 74  
estimation, 431–443  
properties, 432–434  
model selection, 437–438  
moving average representation, 72–73  
prediction, 88–90  
spectral density, 161  
state space representation, 200  
vector, 79
- Autoregressive time series, 36–41, 54–58  
autocorrelation function, 40, 56, 61  
backward representation, 63–64  
estimation for, 404–421  
bias, 412  
distribution, 408  
empirical density, 412  
first order, 404–407  
higher order, 407–418  
least squares, 407–413  
maximum likelihood, 405, 413–414  
multivariate, 419–421

- Autoregressive time series, estimation for**  
*(Continued)*
- nonlinear, 451
  - order of, 412, 437–439
  - residual mean square, 408
  - from residuals, 519–520
  - first order, 39–41
  - infinite order, 65
  - likelihood for, 405, 413–414
  - model selection, 412, 437–439
  - moving average representation, 39, 59, 63
  - nonstationary, 546, 583
  - order of, 39
  - partial correlation in, 40, 62
  - prediction for, 83–85
    - with estimated parameters, 444–449
  - as representation for spectral density, 165
  - root representation for, 67–68
  - second order, 54–58
    - autocorrelation, 56
    - autocovariance function, 55, 61–62
    - correlogram, 57
  - spectral density of, 159
  - threshold, 454
  - unit root, 64, 546, 555–560, 563–564
  - vector, 77–78
    - estimation, 419–421
- Autoregressive transformation, 520.** *See also* Gram Schmidt orthogonalization
- Auxiliary equation, 46.** *See also* Characteristic equation
- Backward shift operator, 43–44, 67**
- autoregressive moving average representation, 74
- Bandpassed white noise, 212**
- Bartlett window, 384**
- Basis vectors, 112**
- Bessel's inequality, 118**
- Bilinear model, 457**
- Bounded in probability, 216.** *See also* Order in probability
- Bricks, pottery, glass and cement, 535**
- Burg estimator, 418**
- Causal, 108 (Exercise 32)**
- Centered moving average, 501**
- Central limit theorem, 233**
- for ARMA parameters, 432
  - for finite moving average series, 320
  - for infinite moving average series, 326
- Liapounov, 233**
- Lindeberg, 233**
- for linear functions, 329**
- for martingale differences, 235**
- for  $m$ -dependent random variables, 321**
- multivariate, 234**
- for regression coefficients, 478**
- for sample autocovariance, 333**
- Cesáro summability, 129**
- for Fourier Series, 129
- Characteristic equation, of autoregressive series, 54, 59, 64**
- of difference equation, 46
  - of moving average time series, 65
    - roots greater than one, 68–70
  - roots, 46–48
    - of vector difference equation, 50
- Characteristic function, 9, 133, 143**
- Characteristic roots, of autoregressive series, 46–48**
- of circulant matrix, 151
  - of explosive process, 611–612
  - of matrix, 298
  - of moving average time series, 65
  - of unit root process, 591
- Chebyshev's inequality, 219**
- Circulant, 150**
- characteristic roots of, 151
  - characteristic vectors of, 151
- Circular matrix, 150**
- symmetric, 151
- Coherency, 172**
- inequality, 172
    - see also* Squared coherency
- Coincident spectral density, 169**
- Cointegrated time series, 608**
- Cointegrating vectors, 608**
- estimators, 617, 627
- Complete convergence, 228.** *See also* Convergence in distribution
- Complete system of functions, 119**
- Complex exponential, 9**
- Complex multivariate normal, 207**
- (Exercise 10)
- Complex trigonometric series, 130**
- Complex valued time series, 12–13**
- Component model, 475, 509.** *See also* Structural model
- Conjugate transpose, 151, 170**
- Continuity theorem, 230**
- Continuous process, 5**
- Convergence, complete, 228**
- in distribution, 227
    - of functions, 222
    - of matrix functions, 233
  - in law, 228

- in mean square (squared mean), 31–32, 35, 120
- of moving average, 31–37
- in probability, 215
- in  $r$ -th mean, 221
- of series, 27–29
- in squared mean (mean square), 31–32, 35, 120
- of sum of random variables, 31–37
- weak, 236
- Convolution**, 30, 136
  - Fourier transform of, 136–138
  - of sequence, 30
- Correlation**, cross, 17
- Correlation function**, 7–9. *See also Autocorrelation function; Cross correlation*
- Correlation matrix**, 17
- Correlation, partial**, 10–12
  - autoregressive time series, 40–41
  - moving average time series, 24
  - see also Partial autocorrelation*
- Correlogram**, 57. *See also Autocorrelation function*
- Cospectrum**, 169
  - estimation of, 391
- Covariance**, 4. *See also Autocovariance; Cross covariance*
- Cross amplitude spectrum**, 174
  - confidence interval for, 394
  - estimation, 393–394
- Cross correlation**, 17
  - estimation, 341–342
    - consistency of, 343
    - covariance of, 342
    - distribution of, 345
  - Cross covariance, 15
    - estimation, 339–343
      - consistency, 343
      - covariance of, 342
- Cross periodogram**, 388
  - definition, 388
  - expected value of, 388
  - smoothed, 389–394
    - expected value of, 390
    - variance of, 390
- Cross spectral function**, 169
- Cycle**, 14
- Delay**, 167
  - frequency response function of, 167
  - phase angle of, 167
- Delta function**, 141 (Exercise 8)
- Des Moines River**, 184, 192, 346, 394
  - sample correlations, 347
- Determinantal equation**, 50, 77, 78
- Deterministic time series**, 81, 96
- Diagnostic checking**, *see Residuals, least squares*
- Difference**, 41, 507–509
  - of autoregressive moving average series, 516
  - effects of, 513–514, 516
  - fractional, 100
  - gain of operator, 514–515
  - of integrated time series, 508
  - of lag H, 508
  - as moving average, 508
  - of periodic functions, 508
  - of polynomials, 46, 508
  - of time series, 507
- Difference equation**, 41–54
  - auxiliary equation, 46
  - characteristic equation, 46
  - determinantal equation, 50
  - general solution, 45
  - homogeneous, 44
  - Jordan canonical form, 52
  - particular solution, 45
  - reduced, 44
  - second order, 47–48
    - solution of  $n$ -th order, 48
    - solution of second order, 47–48
  - solution of vector, 50–51, 53–54
  - stochastic, 39
  - systems of, 49–54
  - vector, 49–54
    - vector  $n$ -th order, 53–54
- Dirac's delta function**, 141 (Exercise 8)
- Distributed lags**, *see Regression with lagged variables*
- Distribution function**, 2–3
- Donsker's theorem**, 237
- Drift**, random walk, 565
- Durbin–Levinson algorithm**, 82
- Durbin's moving average estimator**, 425
  - error in, 426
- Durbin–Watson  $d$** , 486
  - $t$  approximation for, 488
- Ensemble**, 3
- Ergodic**, 308
- Error spectral density**, 177
- Error spectrum**, 177
  - estimation of, 392
- Estimated generalized least squares**, 280
  - distribution of, 280
  - based on residuals, 288
- Estimator**, 79. *See also various models and parameters*

- Even function, 8–9  
 Event, 1  
 Expectation, of products of means, 242  
   of sequence of functions, 243  
 Explanatory variable, 605  
 Explosive autoregression, 583–596  
   estimation, 584–596  
   prediction, 595–596  
 Extrapolation, 79–94  
   polynomial, 482  
   *see also* Prediction
- Fast Fourier transform, 357  
 Fieller's confidence interval, 393  
 Filter, 79, 165  
   delay, 167  
   frequency response function, 167  
   gain of, 167  
   Kalman, 188  
   linear, 165, 497–509  
   matrix, 179  
   for mean, 497–502  
   mean square error of, 184  
   moving average, 497–509  
   phase of, 167  
   power transfer function, 167  
   for signal measurement, 181  
   time invariant, 166  
   transfer function, 167  
   *see also* Moving average; Difference;  
   Autoregressive transformation
- Fisher's periodogram test, 363  
 table, 364  
 Folding frequency, *see* Nyquist frequency  
 Forecast, 79–94. *See also* Prediction
- Fourier coefficients, 114  
   analysis of variance, 356  
   complex, 130–131  
   of finite vector, 114  
   of function on an interval, 117  
   multivariate, 385–386  
     distribution of, 389  
   as regression coefficients, 114–115, 356  
   of time series, 355–356  
   uniqueness, 119
- Fourier Integral Theorem, 132
- Fourier series, Cesáro summability, 129  
   convergence, 121–130  
   integral formula, 123  
   of periodic function, 125  
   *see also* Fourier transform; Trigonometric polynomial
- Fourier transform, 132  
   of convolution, 136–138
- of correlation function, 127, 132  
 of covariance function, 357. *See also*  
   Spectral density  
 of estimated covariance function, 357. *See*  
   *also* Periodogram  
 inverse transform, 133  
 pair, 132  
 of product, 138–139  
 on real line, 132  
 table of, 134
- Fourier transform pair, 132
- Fractional difference, 100
- Frequency, 14  
 Frequency domain, 143  
 Frequency response function, 167  
 Fubini's theorem, 34
- Function, complete system of, 119  
   delta, 141 (Exercise 8)  
   distribution, 2–3  
   even, 8–9  
   expectation of, 243  
   generalized, 140 (Exercise 8)  
   integrable, 117  
   odd, 8–9  
   periodic, 13, 14  
   piecewise smooth, 121  
   sample, 3  
   smooth, 121  
   symmetric distribution, 10  
   trigonometric, 13
- Gain, 174  
 bivariate time series, 174  
 confidence interval for, 394  
 of difference operator, 514  
 estimation, 394  
 of filter, 167  
 of trend removal filter, 513–514
- Gauss-Newton estimator, 269, 437  
 estimator of variance of, 271  
 moving average time series, 422–424  
 one step, 269  
 unit root moving average, 630
- Generalized function, 140 (Exercise 8)
- Generalized inverse, 80
- Generalized least squares, 280  
   efficiency, 479–480  
   estimated, 280, 520–522  
   for trend, 476
- Gibb's phenomenon, 140 (Exercise 3)
- Grafted polynomials, 480–484  
   definition, 481  
   for trend, 482–484
- Gram-Schmidt orthogonalization, 86, 520

- Helly–Bray theorem, 230  
 Hermitian matrix, 170  
 Heterogeneous variances, 488, 492–497  
 Hölder inequality, 243  
 Homogeneous difference equation, 44. *See also* Difference equation
- Implicit price deflator, 593  
 Impulse response function, 529  
 Index set, 3, 5  
 Initial condition equation, 188  
 Innovation algorithm, 86  
 Innovation outlier, 460, 461  
 Input–output system, 174  
 Instrumental variables, 273, 532–534  
     distribution of estimator, 275  
     estimator, 274  
     example of, 277  
     for lagged dependent variables, 532–534  
 Integral formula, Fourier Series, 123  
 Integrated spectrum, 144  
 Integrated time series, 502  
     moving average for, 502–503  
     *see also* Unit root autoregression  
 Inverse autocorrelation function, 211  
 Inverse transform, 133  
 Invertible moving average, 66. *See also* Moving average time series
- Joint distribution function, 3  
 Jordan canonical form, 52  
 Jump discontinuity, 122
- Kalman filter, 188, 497  
     initial condition equation for, 188  
     with missing data, 198  
     for predictions, 198, 202  
     for unit root process, 195  
     updating equations, 190  
     *see also* State space
- Kernel, 382. *See also* Window
- Kolmogorov–Smirnov test, 363
- Kronecker's Lemma, 126
- Lagged value, 42  
     in regression, 530
- Lag window, 382
- Law of large numbers, 253–255. *See also* Convergence in probability
- Least squares estimator, 251  
     for autoregressive time series, 405–412. *See also* Autoregressive time series, estimation
- of autoregressive unit root, 603  
 distribution of, 256, 478  
 efficiency of, 312, 479–480, 518–519  
 of explosive process, 584, 591  
 Fourier coefficients, 114, 117  
 of mean, 311  
 of mean function, 476–480  
 moving average, 497  
 for moving average time series, 421–429.  
     *See also* Moving average time series, estimation  
 nonlinear, 251. *See also* Gauss–Newton estimator  
 for prediction, 79–80. *See also* Prediction residuals, 484–497  
     autoregressive parameters from, 490  
     heterogeneous variances, 488, 492–497  
 with time series errors, 518–529  
 for trend, 476  
 with trigonometric functions, 117  
 variance of, 256  
 variance estimator, 260  
 of vector unit root process, 578–629  
     *see also* Generalized least squares; Maximum likelihood; various models
- Lebesgue–Stieltjes integral, 9
- Liapounov central limit theorem, 233
- Likelihood estimation, *see* Maximum likelihood estimation
- Limited information maximum likelihood, 274
- Lindeberg central limit theorem, 233
- Linear filter, 165, 497–509
- Linear operator, *see* Filter
- Line spectrum, 146
- Long memory time series, 98–101  
     estimation for, 466–471  
     spectral density of, 168
- Lynx, 455
- Martingale difference, 235
- Martingale process, 234
- Matrix, absolutely summable sequence, 76  
     circulant, 150  
     circular, 150  
     circular symmetric, 151
- covariance of least squares estimator, 256.  
     *See also* Least squares estimator; Gauss–Newton estimator; Generalized least squares; Instrumental variables
- covariance of  $n$  observations, 80  
 covariance of prediction errors, 90–91  
 covariance of vector time series, 15  
 diagonalization of covariance, 151–154

- Matrix (Continued)**
- difference equations, 49–54
  - Hermitian, 170
  - Jordan canonical form, 52
  - positive definite, 154
  - positive semidefinite, 154
- Maximum likelihood estimation,**
- autoregression, 405, 413–414
  - autoregressive moving average, 431, 434
  - lagged dependent variable, 535–538
  - regression model with autoregressive errors, 526–527
  - unit root autoregressive process, 546–547, 574–578
- m*-dependent**, 321
- Mean**, best linear estimator of, 311
- central limit theorems for, 321, 329
  - consistency of, 309, 325–326
  - distribution of sample, 321
  - efficiency of sample, 312
  - estimation of, 308
  - function, 475–476
  - local approximation to, 497–499
  - moving average estimator, 497–502
  - structural model for, 509–513
  - product of sample, 242
  - of time series, 4
  - variance of, 309
- Mean prediction error**, 438
- Mean square approximation**, 120
- Mean square continuous**, 6
- Mean square convergence**, 31–32, 35
- Mean square differentiable**, 7
- Measurement error**, 181
- Measurement equation**, 187
- Missing data**, Kalman filter for, 198
- maximum likelihood with, 458–459
- Moorman Manufacturing Company**, 278
- Moving average**, 21–26, 31–39, 497–509
- centered, 501
  - computation of, 499–500, 506–507
  - covariance function, 22
  - covariance of, 33–39, 501–502
  - difference as, 508
  - filter, 497–509
  - for integrated time series, 502
  - least squares, 499
  - for mean, 498–501
  - mean square convergence of, 31–32, 35
  - for nonstationarity, 502
  - for polynomial mean, 501, 504
  - for seasonal, 504–507, 515
  - for seasonal component, 504–507
  - spectral density, effect on, 513–514
- for trend, 498–501, 513
- see also* Filter; Difference
- Moving average process**, 21–26. *See also* Moving average time series
- Moving average time series**, 21–26
- autocorrelation function, 22–23
  - autoregressive representation, 65
  - covariance function, 22–23
  - estimation for, 421–429
  - autocovariance estimator, 421–422
  - distribution, 424, 432
  - Durbin's procedure, 425
  - empirical density, 429
  - first order, 421–425
  - higher order, 430–432
  - initial estimators, 425
  - residual mean square, 424
- first order**, 23–25, 66
- finite**, 21–22
- infinite**, 23
- invertible**, 66
- one sided**, 22
- order of**, 22
- partial correlation in**, 23, 65
- prediction for**, 85–88
- as representation for spectral density**, 163
- as representation for general time series**, 96
- root representation for**, 66
- spectral density of**, 156
- unit root**, 66, 629–636
- vector**, 75–76
- Multiplicative model**, 440
- Multivariate random walk**, 596–599
- Multivariate spectral estimation**, 385–399
- Multivariate time series**, 15–17. *See also* Vector valued time series
- NI( $0, \sigma^2$ )**, 19 (Exercise 9)
- Noise to signal ratio**, 177
- Nondeterministic**, 81, 86, 94
- Nonlinear model**, 250
- autoregression, 451
  - estimator for, 251, 269
- Nonlinear regression**, 250. *See also* Gauss–Newton estimator
- Nonlinearity**, test for, 453–454
- Nonnegative definite**, 7, 9
- Nonstationary autoregression**, 547–596. *See also* Explosive autoregression; Unit root autoregression
- Nonstationary time series**, 4, 475, 547
- types of, 475
  - see also* Nonstationary autoregression
- Nyquist frequency**, 136

- Observation equation, 187  
Odd function, 8, 9  
One-step Gauss-Newton, 269  
One-step prediction, *see* Prediction  
Order, of autoregressive moving average time series, 70  
of autoregressive time series, 39  
of difference equation, 41, 53  
of expected value of functions, 214-250  
of expected value of moments, 242  
of magnitude, 214  
of moving average time series, 22  
in probability, 215  
of trigonometric polynomial, 117  
Ordinary independent variable, 530  
Ordinate, periodogram, 358  
Orthogonal basis, 112  
Orthogonal system of functions, 112, 116  
Orthogonal vectors, 112  
Orthogonality, of trigonometric functions, 112-114, 116  
Outliers, 460-466  
additive, 460  
innovation, 461
- PACF, *see* Partial autocorrelation  
Parseval's theorem, 120  
Partial autocorrelation, 10-12  
autoregressive time series, 40, 62  
estimators, 416  
moving average time series, 23-24, 65  
Partial regression, 10  
Period, 13  
Periodic function, 13  
difference of, 508  
Periodic time series, 13-14  
Periodicity, test for, 363  
Periodogram, 357-363  
chi-square distribution of, 360  
confidence interval based on, 375  
covariance of ordinates, 369  
cross, 388. *See also* Cross periodogram  
cumulative, 363  
definition, 357  
distribution of, 360  
expected value of, 359  
as Fourier transform of sample covariances, 357  
Fisher's test, 363  
Kolmogorov-Smirnov test, 363  
smoothed, 371-374  
distribution of, 372, 374  
expected value of, 372  
variance of, 372  
test for periodicity, 363
- Phase angle, 14  
of delay, 167  
of filter, 167  
Phase spectrum, 174  
confidence interval for, 393  
estimation, 392  
Pig feeding experiment, 277  
Polynomial, extrapolation, 482  
grafted, 480  
moving average for, 501, 504  
trend, difference of, 508  
trigonometric, 112  
Positive definite Hermitian matrix, 170  
Positive semidefinite, 7  
Hermitian matrix, 170  
Power spectrum, *see* Spectral density  
Power transfer function, 167  
Prediction, 79-94  
for autoregressive moving average, 88-94  
for autoregressive time series, 83-85  
confidence interval for, 90  
estimated explosive process, 595-596  
with estimated parameters, 443-457  
estimated unit root process, 582-583  
first order autoregressive, 83-84  
with Kalman filter, 198, 202  
linear, 80  
minimum mean square error, 80  
for moving average time series, 85-88  
polynomial, 482  
recursive equations for, 86-87  
variance of, 80  
Probability limit, 215  
of functions, 222  
Probability space, 1-2  
Process, 3-5. *See also* Time series; various types  
Product, expectation of, 242
- Quadrature spectral density, 169  
Quadrature spectrum, 169  
estimation of, 391
- Random variable, 2  
Random coefficient autoregression, 457  
Random walk, 546  
with drift, 565  
estimation, 546  
vector, 596-599  
*see also* Unit root  
Realization, 3, 5, 81  
Regression, with lagged variables, 530-538  
for time series, 518-529  
*see also* Least squares estimator;  
Generalized least squares; Partial regression

- Regular time series, 81, 86, 94  
 Residuals, least squares, 484–497. *See also*  
   Least squares estimator  
 Response function, *see* Frequency response  
   function  
 Roots, of characteristic equation, 46–48  
   of matrices, 298  
   of polynomial, 291  
   *see also* Characteristic roots
- Sample function, 3  
 Seasonal adjustment, 504–507. *See also*  
   Filter; Moving average  
 Seasonal component, 14, 475, 504–507  
   moving average estimation, 504–507  
   *see also* Mean; Filter; Moving  
   average  
 Sequence, 26–39  
   absolutely summable, 28  
   convergent, 28  
   convolution of, 30  
   of expectations, 243  
   infinite, 26–27  
   of matrices, 76  
   of uncorrelated random variables, 5  
   white noise, 5  
 Serial correlation, *see* Autocorrelation; Cross  
   correlation  
 Serial covariance, *see* Autocovariance; Cross  
   covariance  
 Series, infinite, 27  
   absolutely convergent, 28  
   convergent, 28  
   divergent, 28  
 Fourier, 123–130. *See also* Fourier series  
   trigonometric, 117  
 Shift operator, *see* Backward shift operator;  
   Lagged value  
 Short memory process, 98  
 Sigma-algebra, 2  
 Sigma-field, 2  
 Signal measurement, 181  
 Singular time series, 81, 96  
 Skew symmetric, 13  
 Smoothed periodogram, 371–374  
 Smoothing, *see* Moving average; Filter;  
   Periodogram  
 Spectral density, 132  
   approximation for, 163, 165  
   autoregressive estimator, 385  
   of autoregressive moving average, 161  
   autoregressive representation, 165  
   of autoregressive time series, 159  
   coincident, 169  
   confidence interval for, 375  
   covariance estimator of, 382  
   distribution of, 382–384  
   windows for, 384–385  
 of convolution, 138  
 cross, 169. *See also* Cross spectrum  
   of differenced time series, 513–514  
 error, 177  
 estimation, 355–385. *See also* Cross  
   periodogram; Periodogram; Spectrum  
   estimation  
 existence of, 127  
 of filtered time series, 156, 166–167,  
   179–180, 513–514  
 of long memory process, 168  
 moving average representation, 163  
 of moving average time series, 156  
 properties of, 127  
 quadrature, 169  
 rational, 161  
 of seasonally adjusted time series, 515–516  
 of vector autoregressive process, 180  
 of vector moving average process, 180  
 of vector time series, 172  
   estimation, 385–399  
   *see also* Spectrum  
 Spectral distribution function, 144  
   vector process, 172  
 Spectral kernel, 382  
 Spectral window, 382. *See also* Window  
 Spectrum, 132  
   amplitude, 174  
   cospectrum, 169  
   cross, 169  
   discrete, 146  
   error, 177  
   estimation of, 392  
 estimation, 355–385. *See also* Cross  
   periodogram; Periodogram; Spectral  
   density, covariance estimator  
 integrated, 144  
 line, 146  
 phase, 174  
 quadrature, 169  
   *see also* Spectral density  
 Spline functions, 484. *See also* Grafted  
   polynomials  
 Spirits, 522, 528  
 Squared coherency, 172, 175  
   confidence interval, 392  
   estimation, 391  
   multiple, 391–392  
   test of zero coherency, 392  
 Starting equation, 188  
 State equation, 187  
 State space, 187

- measurement equation, 187
- representation for ARMA, 200
- state equation, 187
- Stationarity, 3–4**
  - covariance, 4
  - second order, 4
  - strictly, 3
  - weakly, 4
  - wide sense, 4
- see also* Nonstationary time series
- Stochastic difference equation, 39**
- estimation, 404–407. *See also* Autoregressive time series, estimation
- vector, 77
- see also* Autoregressive time series
- Stochastic process, 3–5**
  - continuous, 6
  - definition, 3
  - mean square continuous, 6
- see also* Time series
- Stochastic volatility model, 494–497**
- Strictly stationary, 3**
- Structural model, 509–513**
- Symmetric estimators, 414–416**
  - simple, 414
  - simple for unit root, 568–573
  - weighted, 414–416
- weighted for unit root, 570–574
- Taylor's series, for expectations, 214–250**
  - for random functions, 224–226
- Threshold autoregression, 454**
  - smoothed, 455
- Time domain, 143**
- Time series, causal, 108 (Exercise 32)**
  - complex, 12–13
  - covariance stationary, 4
  - continuous, 5
  - definition, 3
  - deterministic, 81
  - long memory, 98–101
  - nonsingular, 82
  - nonstationary, 4
  - periodic, 13
  - realization of, 81
  - regular, 81, 94
  - short memory, 98
  - singular, 81, 96
  - stationary, 3, 4
  - strictly stationary, 3
- see also* Autoregressive time series; Moving average time series; various time series entries
- Transfer function, 167**
  - model, 529
- of trend removal filter, 514
- Transformation, autoregressive:**
  - Gram-Schmidt orthogonalization, 86
  - see also* Filter; Moving average
- Transform pair, 132**
- Transition equation, 187**
- Treasury bill rate, 564, 577, 581, 624**
- Trend, 4, 475**
  - filter for, 497–502
  - global estimation of, 476–480
  - grafted polynomial, 480–482
  - moving average estimation, 497–502
  - structural model for, 509–513
- see also* Mean; Least squares estimator; Moving average
- Trigonometric functions, 14, 112**
  - amplitude, 14
  - completeness of, 119
  - frequency, 14
  - orthogonality of, on integers, 112–114
    - on interval, 116
  - period, 14
  - phase, 14
  - polynomial, 112, 355
  - sum on integers, 112–114
- see also* Fourier series; Fourier transform
- Trigonometric polynomial, 112**
  - order of, 117
- Trigonometric series, 117**
  - complex representation, 130
- Two-stage least squares, 274, 532**
- Unemployment rate, 336, 379, 412**
  - autoregressive representation, 412
  - correlogram for, 339–340
  - monthly, 379
- Unit root, autoregression, 63–64, 546**
  - autoregressive estimation, 547–583
    - distribution, 550–551, 556, 561–563
    - drift, 565–568
    - first order, 547–563
    - higher order, 555–560
    - maximum likelihood, 573–577
    - mean estimated, 560–565
    - median unbiased, 578–579
    - trend estimated, 567–568
  - explanatory variable, 603–606
  - first order, 546
  - Kalman filter for, 195
  - moments of, 547, 639 (Exercise 4)
  - moving average, 66, 629–636
  - prediction for, 573–577
  - test for, univariate, 553, 555, 558, 568–577
  - tables, 641–645

- Unit root (*Continued*)**
- test for, vector, 622
  - tables, 646–652
  - vector autoregression, 596–629
- Unobserved components, 509**
- Updating equation, 190**
- Variance, 4**
- heterogeneous, 488, 492–497
  - see also* Autocovariance; Cross covariance
- Vector process, 15–17, 75–79. *See also* Vector valued time series**
- Vector valued time series, 15–17, 75–79**
- autoregressive, 77–78
  - unit root, 607–610
  - covariance matrix, 15
  - covariance stationary, 15
  - estimation, 419–421. *See also* Cross covariance, estimation; Cross periodogram
  - moving average, 75–76
  - spectral density, 172
    - estimation, 385–399
  - unit roots, tests for, 622
    - tables, 646–652
- Wiener process, 238, 596–599**
- see also* Cross correlation; cross spectral
- function**
- von Neumann ratio, 319, 486**
- t* approximation for, 488
- Weak convergence, 236**
- Weakly stationary, 4**
- Weighted average. *see* Filter; Moving average**
- Wheat yields, 510, 527**
- White noise, 5**
- band passed, 212
- Wiener process, 236**
- Window, 382, 384**
- Bartlett, 384
  - Blackman-Tukey, 385
  - hamming, 385
  - hanning, 385
  - lag, 382
  - Parzen, 385
  - rectangular, 384
  - spectral, 382
  - triangular, 384
  - truncated, 384
- Wold decomposition, 94–98**
- Yule-Walker equations, 55, 408**
- multivariate, 78

## WILEY SERIES IN PROBABILITY AND STATISTICS

ESTABLISHED BY WALTER A. SHEWHART AND SAMUEL S. WILKS

Editors

*Vic Barnett, Ralph A. Bradley, Nicholas I. Fisher, J. Stuart Hunter,  
J. B. Kadane, David G. Kendall, David W. Scott, Adrian F. M. Smith,  
Jozef L. Teugels, Geoffrey S. Watson*

### *Probability and Statistics*

ANDERSON · An Introduction to Multivariate Statistical Analysis, *Second Edition*

\*ANDERSON · The Statistical Analysis of Time Series

ARNOLD, BALAKRISHNAN, and NAGARAJA · A First Course in Order Statistics  
BACCELLI, COHEN, OLSDER, and QUADRAT · Synchronization and Linearity:

An Algebra for Discrete Event Systems

BARNETT · Comparative Statistical Inference, *Second Edition*

BERNARDO and SMITH · Bayesian Statistical Concepts and Theory

BHATTACHARYYA and JOHNSON · Statistical Concepts and Methods

BILLINGSLEY · Convergence of Probability Measures

BILLINGSLEY · Probability and Measure, *Second Edition*

BOROVKOV · Asymptotic Methods in Queueing Theory

BRANDT, FRANKEN, and LISEK · Stationary Stochastic Models

CAINES · Linear Stochastic Systems

CAIROLI and DALANG · Sequential Stochastic Optimization

CHEN · Recursive Estimation and Control for Stochastic Systems

CONSTANTINE · Combinatorial Theory and Statistical Design

COOK and WEISBERG · An Introduction to Regression Graphics

COVER and THOMAS · Elements of Information Theory

\*DOOB · Stochastic Processes

DUDEWICZ and MISHRA · Modern Mathematical Statistics

ETHIER and KURTZ · Markov Processes: Characterization and Convergence

FELLER · An Introduction to Probability Theory and Its Applications, Volume 1,  
*Third Edition, Revised; Volume II, Second Edition*

FULLER · Introduction to Statistical Time Series, *Second Edition*

FULLER · Measurement Error Models

GIFI · Nonlinear Multivariate Analysis

GUTTORP · Statistical Inference for Branching Processes

HALD · A History of Probability and Statistics and Their Applications before 1750

HALL · Introduction to the Theory of Coverage Processes

HANNAN and DEISTLER · The Statistical Theory of Linear Systems

HEDAYAT and SINHA · Design and Inference in Finite Population Sampling

HOEL · Introduction to Mathematical Statistics, *Fifth Edition*

HUBER · Robust Statistics

IMAN and CONOVER · A Modern Approach to Statistics

JUREK and MASON · Operator-Limit Distributions in Probability Theory

KAUFMAN and ROUSSEEUW · Finding Groups in Data: An Introduction to Cluster  
Analysis

LARSON · Introduction to Probability Theory and Statistical Inference, *Third Edition*

LESSLER and KALSBECK · Nonsampling Error in Surveys

LINDVALL · Lectures on the Coupling Method

MANTON, WOODBURY, and TOLLEY · Statistical Applications Using Fuzzy Sets

MORGENTHALER and TUKEY · Configural Polysampling: A Route to Practical  
Robustness

MUIRHEAD · Aspects of Multivariate Statistical Theory

\*Now available in a lower priced paperback edition in the Wiley Classics Library.

*Probability and Statistics (Continued)*

- OLIVER and SMITH · Influence Diagrams, Belief Nets and Decision Analysis  
\*PARZEN · Modern Probability Theory and Its Applications  
PILZ · Bayesian Estimation and Experimental Design in Linear Regression Models  
PRESS · Bayesian Statistics: Principles, Models, and Applications  
PUKELSHEIM · Optimal Experimental Design  
PURI and SEN · Nonparametric Methods in General Linear Models  
PURI, VILAPLANA, and WERTZ · New Perspectives in Theoretical and Applied Statistics  
RAO · Asymptotic Theory of Statistical Inference  
RAO · Linear Statistical Inference and Its Applications, *Second Edition*  
RENCHER · Methods of Multivariate Analysis  
ROBERTSON, WRIGHT, and DYKSTRA · Order Restricted Statistical Inference  
ROGERS and WILLIAMS · Diffusions, Markov Processes, and Martingales, Volume II: Itô Calculus  
ROHATGI · An Introduction to Probability Theory and Mathematical Statistics  
ROSS · Stochastic Processes  
RUBINSTEIN · Simulation and the Monte Carlo Method  
RUZSA and SZEKELY · Algebraic Probability Theory  
SCHEFFE · The Analysis of Variance  
SEBER · Linear Regression Analysis  
SEBER · Multivariate Observations  
SEBER and WILD · Nonlinear Regression  
SERFLING · Approximation Theorems of Mathematical Statistics  
SHORACK and WELLNER · Empirical Processes with Applications to Statistics  
SMALL and MCLEISH · Hilbert Space Methods in Probability and Statistical Inference  
STAPLETON · Linear Statistical Models  
STAUDTE and SHEATHER · Robust Estimation and Testing  
STOYANOV · Counterexamples in Probability  
STYAN · The Collected Papers of T. W. Anderson: 1943–1985  
WHITTAKER · Graphical Models in Applied Multivariate Statistics  
YANG · The Construction Theory of Denumerable Markov Processes

*Applied Probability and Statistics*

- ABRAHAM and LEDOLTER · Statistical Methods for Forecasting  
AGRESTI · Analysis of Ordinal Categorical Data  
AGRESTI · Categorical Data Analysis  
ANDERSON and LOYNES · The Teaching of Practical Statistics  
ANDERSON, AUQUIER, HAUCK, OAKES, VANDAELE, and WEISBERG · Statistical Methods for Comparative Studies  
\*ARTHANARI and DODGE · Mathematical Programming in Statistics  
ASMUSSEN · Applied Probability and Queues  
\*BAILEY · The Elements of Stochastic Processes with Applications to the Natural Sciences  
BARNETT · Interpreting Multivariate Data  
BARNETT and LEWIS · Outliers in Statistical Data, *Second Edition*  
BARTHOLOMEW, FORBES, and MCLEAN · Statistical Techniques for Manpower Planning, *Second Edition*  
BATES and WATTS · Nonlinear Regression Analysis and Its Applications  
BECHHOFER, SANTNER, and GOLDSMAN · Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons  
BELSLEY · Conditioning Diagnostics: Collinearity and Weak Data in Regression  
BELSLEY, KUH, and WELSCH · Regression Diagnostics: Identifying Influential Data and Sources of Collinearity

\*Now available in a lower priced paperback edition in the Wiley Classics Library.

*Applied Probability and Statistics (Continued)*

- BERRY, CHALONER, and GEWEKE · Bayesian Analysis in Statistics and  
Econometrics: Essays in Honor of Arnold Zellner
- BHAT · Elements of Applied Stochastic Processes, *Second Edition*
- BHATTACHARYA and WAYMIRE · Stochastic Processes with Applications
- BIEMER, GROVES, LYBERG, MATHIOWETZ, and SUDMAN · Measurement  
Errors in Surveys
- BIRKES and DODGE · Alternative Methods of Regression
- BLOOMFIELD · Fourier Analysis of Time Series: An Introduction
- BOLLEN · Structural Equations with Latent Variables
- BOULEAU · Numerical Methods for Stochastic Processes
- BOX · R. A. Fisher, the Life of a Scientist
- BOX and DRAPER · Empirical Model-Building and Response Surfaces
- BOX and DRAPER · Evolutionary Operation: A Statistical Method for Process  
Improvement
- BOX, HUNTER, and HUNTER · Statistics for Experimenters: An Introduction to  
Design, Data Analysis, and Model Building
- BROWN and HOLLANDER · Statistics: A Biomedical Introduction
- BUCKLEW · Large Deviation Techniques in Decision, Simulation, and Estimation
- BUNKE and BUNKE · Nonlinear Regression, Functional Relations and Robust  
Methods: Statistical Methods of Model Building
- CHATTERJEE and HADI · Sensitivity Analysis in Linear Regression
- CHATTERJEE and PRICE · Regression Analysis by Example, *Second Edition*
- CLARKE and DISNEY · Probability and Random Processes: A First Course with  
Applications, *Second Edition*
- COCHRAN · Sampling Techniques, *Third Edition*
- \*COCHRAN and COX · Experimental Designs, *Second Edition*
- CONOVER · Practical Nonparametric Statistics, *Second Edition*
- CONOVER and IMAN · Introduction to Modern Business Statistics
- CORNELL · Experiments with Mixtures, Designs, Models, and the Analysis of Mixture  
Data, *Second Edition*
- COX · A Handbook of Introductory Statistical Methods
- \*COX · Planning of Experiments
- COX, BINDER, CHINNAPPA, CHRISTIANSON, COLLEDGE, and KOTT ·  
Business Survey Methods
- CRESSIE · Statistics for Spatial Data, *Revised Edition*
- DANIEL · Applications of Statistics to Industrial Experimentation
- DANIEL · Biostatistics: A Foundation for Analysis in the Health Sciences, *Sixth Edition*
- DAVID · Order Statistics, *Second Edition*
- \*DEGROOT, FIENBERG, and KADANE · Statistics and the Law
- \*DEMING · Sample Design in Business Research
- DILLON and GOLDSTEIN · Multivariate Analysis: Methods and Applications
- DODGE and ROMIG · Sampling Inspection Tables, *Second Edition*
- DOWDY and WEARDEN · Statistics for Research, *Second Edition*
- DRAPER and SMITH · Applied Regression Analysis, *Second Edition*
- DUNN · Basic Statistics: A Primer for the Biomedical Sciences, *Second Edition*
- DUNN and CLARK · Applied Statistics: Analysis of Variance and Regression, *Second  
Edition*
- ELANDT-JOHNSON and JOHNSON · Survival Models and Data Analysis
- EVANS, PEACOCK, and HASTINGS · Statistical Distributions, *Second Edition*
- FISHER and VAN BELLE · Biostatistics: A Methodology for the Health Sciences
- FLEISS · The Design and Analysis of Clinical Experiments
- FLEISS · Statistical Methods for Rates and Proportions, *Second Edition*
- FLEMING and HARRINGTON · Counting Processes and Survival Analysis
- FLURY · Common Principal Components and Related Multivariate Models

\*Now available in a lower priced paperback edition in the Wiley Classics Library.

*Applied Probability and Statistics (Continued)*

- GALLANT · Nonlinear Statistical Models
- GLASSERMAN and YAO · Monotone Structure in Discrete-Event Systems
- GROSS and HARRIS · Fundamentals of Queueing Theory, *Second Edition*
- GROVES · Survey Errors and Survey Costs
- GROVES, BIEMER, LYBERG, MASSEY, NICHOLLS, and WAKSBERG · Telephone Survey Methodology
- HAHN and MEEKER · Statistical Intervals: A Guide for Practitioners
- HAND · Discrimination and Classification
- \*HANSEN, HURWITZ, and MADOW · Sample Survey Methods and Theory, Volume 1: Methods and Applications
- \*HANSEN, HURWITZ, and MADOW · Sample Survey Methods and Theory, Volume II: Theory
- HEIBERGER · Computation for the Analysis of Designed Experiments
- HELLER · MACSYMA for Statisticians
- HINKELMAN and KEMPTHORNE · Design and Analysis of Experiments, Volume 1: Introduction to Experimental Design
- HOAGLIN, MOSTELLER, and TUKEY · Exploratory Approach to Analysis of Variance
- HOAGLIN, MOSTELLER, and TUKEY · Exploring Data Tables, Trends and Shapes
- HOAGLIN, MOSTELLER, and TUKEY · Understanding Robust and Exploratory Data Analysis
- HOCHBERG and TAMHANE · Multiple Comparison Procedures
- HOEL · Elementary Statistics, *Fifth Edition*
- HOGG and KLUGMAN · Loss Distributions
- HOLLANDER and WOLFE · Nonparametric Statistical Methods
- HOSMER and LEMESHOW · Applied Logistic Regression
- HØYLAND and RAUSAND · System Reliability Theory: Models and Statistical Methods
- HUBERTY · Applied Discriminant Analysis
- IMAN and CONOVER · Modern Business Statistics
- JACKSON · A User's Guide to Principle Components
- JOHN · Statistical Methods in Engineering and Quality Assurance
- JOHNSON · Multivariate Statistical Simulation
- JOHNSON and KOTZ · Distributions in Statistics  
Continuous Univariate Distributions—2  
Continuous Multivariate Distributions
- JOHNSON, KOTZ, and BALAKRISHNAN · Continuous Univariate Distributions, Volume 1, *Second Edition*
- JOHNSON, KOTZ, and KEMP · Univariate Discrete Distributions, *Second Edition*
- JUDGE, GRIFFITHS, HILL, LÜTKEPOHL, and LEE · The Theory and Practice of Econometrics, *Second Edition*
- JUDGE, HILL, GRIFFITHS, LÜTKEPOHL, and LEE · Introduction to the Theory and Practice of Econometrics, *Second Edition*
- JUREČKOVÁ and SEN · Robust Statistical Procedures: Asymptotics and Interrelations
- KADANE · Bayesian Methods and Ethics in a Clinical Trial Design
- KALBFLEISCH and PRENTICE · The Statistical Analysis of Failure Time Data
- KASPRZYK, DUNCAN, KALTON, and SINGH · Panel Surveys
- KISH · Statistical Design for Research
- \*KISH · Survey Sampling
- LANGE, RYAN, BILLARD, BRILLINGER, CONQUEST, and GREENHOUSE · Case Studies in Biometry
- LAWLESS · Statistical Models and Methods for Lifetime Data
- LEBART, MORINEAU, and WARWICK · Multivariate Descriptive Statistical Analysis: Correspondence Analysis and Related Techniques for Large Matrices
- LEE · Statistical Methods for Survival Data Analysis, *Second Edition*
- LEPAGE and BILLARD · Exploring the Limits of Bootstrap

\*Now available in a lower priced paperback edition in the Wiley Classics Library.

*Applied Probability and Statistics (Continued)*

- LEVY and LEMESHOW · Sampling of Populations: Methods and Applications  
LINHART and ZUCCHINI · Model Selection  
LITTLE and RUBIN · Statistical Analysis with Missing Data  
MAGNUS and NEUDECKER · Matrix Differential Calculus with Applications in Statistics and Econometrics  
MAINDONALD · Statistical Computation  
MALLOWS · Design, Data, and Analysis by Some Friends of Cuthbert Daniel  
MANN, SCHAFER, and SINGPURWALLA · Methods for Statistical Analysis of Reliability and Life Data  
MASON, GUNST, and HESS · Statistical Design and Analysis of Experiments with Applications to Engineering and Science  
McLACHLAN · Discriminant Analysis and Statistical Pattern Recognition  
MILLER · Survival Analysis  
MONTGOMERY and MYERS · Response Surface Methodology: Process and Product in Optimization Using Designed Experiments  
MONTGOMERY and PECK · Introduction to Linear Regression Analysis, *Second Edition*  
NELSON · Accelerated Testing, Statistical Models, Test Plans, and Data Analyses  
NELSON · Applied Life Data Analysis  
OCHI · Applied Probability and Stochastic Processes in Engineering and Physical Sciences  
OKABE, BOOTS, and SUGIHARA · Spatial Tesselations: Concepts and Applications of Voronoi Diagrams  
OSBORNE · Finite Algorithms in Optimization and Data Analysis  
PANKRATZ · Forecasting with Dynamic Regression Models  
PANKRATZ · Forecasting with Univariate Box-Jenkins Models: Concepts and Cases  
PORT · Theoretical Probability for Applications  
PUTERMAN · Markov Decision Processes: Discrete Stochastic Dynamic Programming  
RACHEV · Probability Metrics and the Stability of Stochastic Models  
RÉNYI · A Diary on Information Theory  
RIPLEY · Spatial Statistics  
RIPLEY · Stochastic Simulation  
ROSS · Introduction to Probability and Statistics for Engineers and Scientists  
ROUSSEEUW and LEROY · Robust Regression and Outlier Detection  
RUBIN · Multiple Imputation for Nonresponse in Surveys  
RYAN · Statistical Methods for Quality Improvement  
SCHUSS - Theory and Applications of Stochastic Differential Equations  
SCOTT · Multivariate Density Estimation: Theory, Practice, and Visualization  
SEARLE · Linear Models  
SEARLE · Linear Models for Unbalanced Data  
SEARLE · Matrix Algebra Useful for Statistics  
SEARLE, CASELLA, and McCULLOCH · Variance Components  
SKINNER, HOLT, and SMITH · Analysis of Complex Surveys  
STOYAN · Comparison Methods for Queues and Other Stochastic Models  
STOYAN, KENDALL, and MECKE · Stochastic Geometry and Its Applications  
STOYAN and STOYAN · Fractals, Random Shapes and Point Fields  
THOMPSON · Empirical Model Building  
THOMPSON · Sampling  
TIERNEY · LISP-STAT: An Object-Oriented Environment for Statistical Computing and Dynamic Graphics  
TIJMS · Stochastic Modeling and Analysis: A Computational Approach  
TITTERINGTON, SMITH, and MAKOV · Statistical Analysis of Finite Mixture Distributions  
UPTON and FINGLETON · Spatial Data Analysis by Example, Volume 1: Point Pattern and Quantitative Data

*Applied Probability and Statistics (Continued)*

- UPTON and FINGLETON · Spatial Data Analysis by Example, Volume II:  
Categorical and Directional Data
- VAN RIJCKEVORSEL and DE LEEUW · Component and Correspondence Analysis
- WEISBERG · Applied Linear Regression, *Second Edition*
- WESTFALL and YOUNG · Resampling-Based Multiple Testing: Examples and  
Methods for *p*-Value Adjustment
- WHITTLE · Optimization Over Time: Dynamic Programming and Stochastic Control,  
Volume I and Volume II
- WHITTLE · Systems in Stochastic Equilibrium
- WONNACOTT and WONNACOTT · Econometrics, *Second Edition*
- WONNACOTT and WONNACOTT · Introductory Statistics, *Fifth Edition*
- WONNACOTT and WONNACOTT · Introductory Statistics for Business and  
Economics, *Fourth Edition*
- WOODING · Planning Pharmaceutical Clinical Trials: Basic Statistical Principles
- WOOLSON · Statistical Methods for the Analysis of Biomedical Data

*Tracts on Probability and Statistics*

- BILLINGSLEY · Convergence of Probability Measures
- TOUTENBURG · Prior Information in Linear Models

This Page intentionally left blank

This Page intentionally left blank

This Page intentionally left blank



9 780471 552390