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
Introduction to Statistical Time Series
WILEY SERIES IN PROBABILITY AND STATISTICS

Established by WALTER A. SHEWHART and SAMUEL S. WILKS


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
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.
 I. Time-series analysis. 2. Regression analysis. I. Title
II. Series.
QA280.F84 1996
519.5'5—dc20  95-14875

10 9 8 7 6 5 4 3
To Evelyn
This Page intentionally left blank
Contents

Preface to the First Edition  xi
Preface to the Second Edition  xiii
List of Principal Results  xv
List of Examples  xxı

1. Introduction  1
   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

2. Moving Average and Autoregressive Processes  21
   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
<table>
<thead>
<tr>
<th>2.11</th>
<th>Long Memory Processes</th>
<th>98</th>
</tr>
</thead>
<tbody>
<tr>
<td>References</td>
<td></td>
<td>101</td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
<td>101</td>
</tr>
<tr>
<td>3.</td>
<td>Introduction to Fourier Analysis</td>
<td>112</td>
</tr>
<tr>
<td>3.1</td>
<td>Systems of Orthogonal Functions—Fourier Coefficients</td>
<td>112</td>
</tr>
<tr>
<td>3.2</td>
<td>Complex Representation of Trigonometric Series</td>
<td>130</td>
</tr>
<tr>
<td>3.3</td>
<td>Fourier Transform—Functions Defined on the Real Line</td>
<td>132</td>
</tr>
<tr>
<td>3.4</td>
<td>Fourier Transform of a Convolution</td>
<td>136</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>139</td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
<td>139</td>
</tr>
<tr>
<td>4.</td>
<td>Spectral Theory and Filtering</td>
<td>143</td>
</tr>
<tr>
<td>4.1</td>
<td>The Spectrum</td>
<td>143</td>
</tr>
<tr>
<td>4.2</td>
<td>Circulants—Diagonalization of the Covariance Matrix of Stationary Process</td>
<td>149</td>
</tr>
<tr>
<td>4.3</td>
<td>The Spectral Density of Moving Average and Autoregressive Time Series</td>
<td>155</td>
</tr>
<tr>
<td>4.4</td>
<td>Vector Processes</td>
<td>169</td>
</tr>
<tr>
<td>4.5</td>
<td>Measurement Error—Signal Detection</td>
<td>181</td>
</tr>
<tr>
<td>4.6</td>
<td>State Space Models and Kalman Filtering</td>
<td>187</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>205</td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
<td>205</td>
</tr>
<tr>
<td>5.</td>
<td>Some Large Sample Theory</td>
<td>214</td>
</tr>
<tr>
<td>5.1</td>
<td>Order in Probability</td>
<td>214</td>
</tr>
<tr>
<td>5.2</td>
<td>Convergence in Distribution</td>
<td>227</td>
</tr>
<tr>
<td>5.3</td>
<td>Central Limit Theorems</td>
<td>233</td>
</tr>
<tr>
<td>5.4</td>
<td>Approximating a Sequence of Expectations</td>
<td>240</td>
</tr>
<tr>
<td>5.5</td>
<td>Estimation for Nonlinear Models</td>
<td>250</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Estimators that Minimize an Objective Function</td>
<td>250</td>
</tr>
<tr>
<td>5.5.2</td>
<td>One-Step Estimation</td>
<td>268</td>
</tr>
<tr>
<td>5.6</td>
<td>Instrumental Variables</td>
<td>273</td>
</tr>
<tr>
<td>5.7</td>
<td>Estimated Generalized Least Squares</td>
<td>279</td>
</tr>
<tr>
<td>5.8</td>
<td>Sequences of Roots of Polynomials</td>
<td>290</td>
</tr>
<tr>
<td>References</td>
<td></td>
<td>299</td>
</tr>
<tr>
<td>Exercises</td>
<td></td>
<td>299</td>
</tr>
</tbody>
</table>
CONTENTS

6. Estimation of the Mean and Autocorrelations 308
   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

7. The Periodogram, Estimated Spectrum 355
   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

8. Parameter Estimation 404
   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

9. Regression, Trend, and Seasonality 475
   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
   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

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

<table>
<thead>
<tr>
<th>Number</th>
<th>Topic</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.3.1</td>
<td>Finite index set, 4</td>
</tr>
<tr>
<td>1.3.2</td>
<td>White noise, 5</td>
</tr>
<tr>
<td>1.3.3</td>
<td>A nonstationary time series, 5</td>
</tr>
<tr>
<td>1.3.4</td>
<td>Continuous time series, 6</td>
</tr>
<tr>
<td>2.1.1</td>
<td>First order moving average time series, 23</td>
</tr>
<tr>
<td>2.1.2</td>
<td>Second order moving average time series, 24</td>
</tr>
<tr>
<td>2.3.1</td>
<td>First order autoregressive time series, 40</td>
</tr>
<tr>
<td>2.5.1</td>
<td>Correlogram of time series, 57</td>
</tr>
<tr>
<td>2.9.1</td>
<td>Prediction for unemployment rate, 90</td>
</tr>
<tr>
<td>2.9.2</td>
<td>Prediction for autoregressive moving average, 91</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Spectral distribution function composed of jumps, 147</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Spectral distribution function of series with white noise component, 148</td>
</tr>
<tr>
<td>4.5.1</td>
<td>Filter for time series observed subject to measurement error, 184</td>
</tr>
<tr>
<td>4.6.1</td>
<td>Kalman Filter for Des Moines River data, 192</td>
</tr>
<tr>
<td>4.6.2</td>
<td>Kalman filter, Des Moines River, mean unknown, 193</td>
</tr>
<tr>
<td>4.6.3</td>
<td>Kalman filter, autoregressive unit root, 195</td>
</tr>
<tr>
<td>4.6.4</td>
<td>Kalman filter for missing observations, 198</td>
</tr>
<tr>
<td>4.6.5</td>
<td>Predictions constructed with the Kalman filter, 202</td>
</tr>
<tr>
<td>5.1.1</td>
<td>Taylor expansion about a random point, 227</td>
</tr>
<tr>
<td>5.4.1</td>
<td>Approximation to expectation, 248</td>
</tr>
<tr>
<td>5.5.1</td>
<td>Transformation of model with different rates of convergence, 265</td>
</tr>
<tr>
<td>5.5.2</td>
<td>Failure of convergence of second derivative of nonlinear model, 267</td>
</tr>
<tr>
<td>5.5.3</td>
<td>Gauss–Newton estimation, 272</td>
</tr>
<tr>
<td>5.6.1</td>
<td>Instrumental variable estimation, 277</td>
</tr>
<tr>
<td>6.3.1</td>
<td>Sample autocorrelations and means of unemployment rate, 336</td>
</tr>
</tbody>
</table>
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 \( \Omega = \{ \text{one spot shows, two }
\text{spots show, \ldots, six spots show} \} \) 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 $A_1, A_2, \ldots$ 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 $A_1, A_2, \ldots$ 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 $F_x(\alpha) = P(\{\omega: X(\omega) \leq x\})$ 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 \((\Omega, \mathcal{F}, P)\) 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 \((\Omega, \mathcal{F}, P)\). 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 \(\{X_1, X_2, \ldots, X_n\}\) from the collection \(\{X_t : t \in T\}\) is defined by

\[
F_{X_1, X_2, \ldots, X_n}(x_1, x_2, \ldots, x_n) = P(\omega : X(t_1, \omega) \leq x_1, \ldots, X(t_n, \omega) \leq x_n). \tag{1.2.1}
\]

A time series is called strictly stationary if

\[
F_{X_{t_1+h}, X_{t_2+h}, \ldots, X_{t_n+h}}(x_1, x_2, \ldots, x_n) = F_{X_{t_1}, X_{t_2}, \ldots, X_{t_n}}(x_1, x_2, \ldots, x_n),
\]

where the equality must hold for all possible (nonempty finite distinct) sets of indices \(t_1, t_2, \ldots, t_n\) and \(t_1 + h, t_2 + h, \ldots, t_n + h\) in the index set and all \((x_1, x_2, \ldots, x_n)\) in the range of the random variable \(X_t\). Note that the indices \(t_1, t_2, \ldots, t_n\) 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 \((X_{t_1}, X_{t_2}, \ldots, X_{t_n})\). 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 \textit{weakly stationary} if:

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

As before \( t_1, t_2, \ldots, t_n \) 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 \textit{autocovariance} of \( X_t \). When there is no danger of confusion, we shall abbreviate the expression to \textit{covariance}.

The terms \textit{stationary in the wide sense}, \textit{covariance stationary}, \textit{second order stationary}, and \textit{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

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