

Multivariate Time Series Analysis and Applications

WILEY SERIES IN PROBABILITY AND STATISTICS

Established by *Walter A. Shewhart and Samuel S. Wilks*

Editors: *David J. Balding, Noel A. C. Cressie, Garrett M. Fitzmaurice, Geof H. Givens, Harvey Goldstein, Geert Molenberghs, David W. Scott, Adrian F. M. Smith, Ruey S. Tsay*

Editors Emeriti: *J. Stuart Hunter, Iain M. Johnstone, Joseph B. Kadane, Jozef L. Teugels*

The *Wiley Series in Probability and Statistics* is well established and authoritative. It covers many topics of current research interest in both pure and applied statistics and probability theory. Written by leading statisticians and institutions, the titles span both state-of-the-art developments in the field and classical methods.

Reflecting the wide range of current research in statistics, the series encompasses applied, methodological and theoretical statistics, ranging from applications and new techniques made possible by advances in computerized practice to rigorous treatment of theoretical approaches. This series provides essential and invaluable reading for all statisticians, whether in academia, industry, government, or research.

A complete list of titles in this series can be found at
<http://www.wiley.com/go/wsps>

Multivariate Time Series Analysis and Applications

William W.S. Wei

*Department of Statistical Science
Temple University, Philadelphia, PA, USA*

WILEY

This edition first published 2019
© 2019 John Wiley & Sons Ltd

All rights reserved. No part of this publication may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise, except as permitted by law. Advice on how to obtain permission to reuse material from this title is available at <http://www.wiley.com/go/permissions>.

The right of William W.S. Wei to be identified as the author of this work has been asserted in accordance with law.

Registered Offices

John Wiley & Sons, Inc., 111 River Street, Hoboken, NJ 07030, USA
John Wiley & Sons Ltd, The Atrium, Southern Gate, Chichester, West Sussex, PO19 8SQ, UK

Editorial Office

9600 Garsington Road, Oxford, OX4 2DQ, UK

For details of our global editorial offices, customer services, and more information about Wiley products visit us at www.wiley.com.

Wiley also publishes its books in a variety of electronic formats and by print-on-demand. Some content that appears in standard print versions of this book may not be available in other formats.

Limit of Liability/Disclaimer of Warranty

MATLAB® is a trademark of The MathWorks, Inc. and is used with permission. The MathWorks does not warrant the accuracy of the text or exercises in this book. This work's use or discussion of MATLAB® software or related products does not constitute endorsement or sponsorship by The MathWorks of a particular pedagogical approach or particular use of the MATLAB® software. While the publisher and authors have used their best efforts in preparing this work, they make no representations or warranties with respect to the accuracy or completeness of the contents of this work and specifically disclaim all warranties, including without limitation any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives, written sales materials or promotional statements for this work. The fact that an organization, website, or product is referred to in this work as a citation and/or potential source of further information does not mean that the publisher and authors endorse the information or services the organization, website, or product may provide or recommendations it may make. This work is sold with the understanding that the publisher is not engaged in rendering professional services. The advice and strategies contained herein may not be suitable for your situation. You should consult with a specialist where appropriate. Further, readers should be aware that websites listed in this work may have changed or disappeared between when this work was written and when it is read. Neither the publisher nor authors shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

Library of Congress Cataloging-in-Publication data applied for

Hardback: 9781119502852

Cover design: Wiley

Cover image: Courtesy of William W.S. Wei

Set in 10/12pt Times by SPi Global, Pondicherry, India

10 9 8 7 6 5 4 3 2 1

To my Grandpa, Parents, Wife, Sons,
and Daughter

Contents

| | |
|---|-------------|
| <i>About the author</i> | xiii |
| <i>Preface</i> | xv |
| <i>About the Companion website</i> | xvii |
| 1 Fundamental concepts and issues in multivariate time series analysis | 1 |
| 1.1 Introduction | 1 |
| 1.2 Fundamental concepts | 3 |
| 1.2.1 Correlation and partial correlation matrix functions | 3 |
| 1.2.2 Vector white noise process | 7 |
| 1.2.3 Moving average and autoregressive representations of vector processes | 7 |
| Projects | 9 |
| References | 9 |
| 2 Vector time series models | 11 |
| 2.1 Vector moving average processes | 11 |
| 2.2 Vector autoregressive processes | 14 |
| 2.2.1 Granger causality | 18 |
| 2.3 Vector autoregressive moving average processes | 18 |
| 2.4 Nonstationary vector autoregressive moving average processes | 21 |
| 2.5 Vector time series model building | 21 |
| 2.5.1 Identification of vector time series models | 21 |
| 2.5.2 Sample moments of a vector time series | 22 |
| 2.5.2.1 <i>Sample mean and sample covariance matrices</i> | 22 |
| 2.5.2.2 <i>Sample correlation matrix function</i> | 23 |
| 2.5.2.3 <i>Sample partial correlation matrix function and extended cross-correlation matrices</i> | 24 |
| 2.5.3 Parameter estimation, diagnostic checking, and forecasting | 24 |
| 2.5.4 Cointegration in vector time series | 25 |
| 2.6 Seasonal vector time series model | 26 |
| 2.7 Multivariate time series outliers | 27 |
| 2.7.1 Types of multivariate time series outliers and detections | 27 |
| 2.7.2 Outlier detection through projection pursuit | 29 |
| 2.8 Empirical examples | 32 |
| 2.8.1 First model of US monthly retail sales revenue | 32 |
| 2.8.2 Second model of US monthly retail sales revenue | 43 |
| 2.8.3 US macroeconomic indicators | 47 |
| 2.8.4 Unemployment rates with outliers | 58 |
| Software code | 65 |
| Projects | 100 |
| References | 101 |

| | | |
|---------|--|-----|
| 3 | Multivariate time series regression models | 105 |
| 3.1 | Introduction | 105 |
| 3.2 | Multivariate multiple time series regression models | 105 |
| 3.2.1 | The classical multiple regression model | 105 |
| 3.2.2 | Multivariate multiple regression model | 106 |
| 3.3 | Estimation of the multivariate multiple time series regression model | 108 |
| 3.3.1 | The Generalized Least Squares (GLS) estimation | 108 |
| 3.3.2 | Empirical Example I – U.S. retail sales and some national indicators | 109 |
| 3.4 | Vector time series regression models | 114 |
| 3.4.1 | Extension of a VAR model to VARX models | 114 |
| 3.4.2 | Empirical Example II – VARX models for U.S. retail sales and some national indicators | 115 |
| 3.5 | Empirical Example III – Total mortality and air pollution in California | 120 |
| | Software code | 129 |
| | Projects | 137 |
| | References | 137 |
| 4 | Principle component analysis of multivariate time series | 139 |
| 4.1 | Introduction | 139 |
| 4.2 | Population PCA | 140 |
| 4.3 | Implications of PCA | 141 |
| 4.4 | Sample principle components | 142 |
| 4.5 | Empirical examples | 145 |
| 4.5.1 | Daily stock returns from the first set of 10 stocks | 145 |
| 4.5.1.1 | <i>The PCA based on the sample covariance matrix</i> | 147 |
| 4.5.1.2 | <i>The PCA based on the sample correlation matrix</i> | 150 |
| 4.5.2 | Monthly Consumer Price Index (CPI) from five sectors | 152 |
| 4.5.2.1 | <i>The PCA based on the sample covariance matrix</i> | 153 |
| 4.5.2.2 | <i>The PCA based on the sample correlation matrix</i> | 154 |
| | Software code | 157 |
| | Projects | 160 |
| | References | 161 |
| 5 | Factor analysis of multivariate time series | 163 |
| 5.1 | Introduction | 163 |
| 5.2 | The orthogonal factor model | 163 |
| 5.3 | Estimation of the factor model | 165 |
| 5.3.1 | The principal component method | 165 |
| 5.3.2 | Empirical Example I – Model 1 on daily stock returns from the second set of 10 stocks | 166 |
| 5.3.3 | The maximum likelihood method | 169 |
| 5.3.4 | Empirical Example II – Model 2 on daily stock returns from the second set of 10 stocks | 173 |
| 5.4 | Factor rotation | 175 |
| 5.4.1 | Orthogonal rotation | 176 |
| 5.4.2 | Oblique rotation | 176 |
| 5.4.3 | Empirical Example III – Model 3 on daily stock returns from the second set of 10 stocks | 177 |

| | | |
|---------|---|-----|
| 5.5 | Factor scores | 178 |
| 5.5.1 | Introduction | 178 |
| 5.5.2 | Empirical Example IV – Model 4 on daily stock returns from the second set of 10 stocks | 179 |
| 5.6 | Factor models with observable factors | 181 |
| 5.7 | Another empirical example – Yearly U.S. sexually transmitted diseases (STD) | 183 |
| 5.7.1 | Principal components analysis (PCA) | 183 |
| 5.7.1.1 | <i>PCA for standardized Z_t</i> | 183 |
| 5.7.1.2 | <i>PCA for unstandardized Z_t</i> | 186 |
| 5.7.2 | Factor analysis | 186 |
| 5.8 | Concluding remarks | 193 |
| | Software code | 194 |
| | Projects | 200 |
| | References | 201 |
| 6 | Multivariate GARCH models | 203 |
| 6.1 | Introduction | 203 |
| 6.2 | Representations of multivariate GARCH models | 204 |
| 6.2.1 | VEC and DVEC models | 204 |
| 6.2.2 | Constant Conditional Correlation (CCC) models | 206 |
| 6.2.3 | BEKK models | 207 |
| 6.2.4 | Factor models | 208 |
| 6.3 | O-GARCH and GO-GARCH models | 209 |
| 6.4 | Estimation of GO-GARCH models | 210 |
| 6.4.1 | The two-step estimation method | 210 |
| 6.4.2 | The weighted scatter estimation method | 211 |
| 6.5 | Properties of the weighted scatter estimator | 213 |
| 6.5.1 | Asymptotic distribution and statistical inference | 213 |
| 6.5.2 | Combining information from different weighting functions | 214 |
| 6.6 | Empirical examples | 215 |
| 6.6.1 | U.S. weekly interest over time on six exercise items | 215 |
| 6.6.1.1 | <i>Choose a best VAR/VARMA model</i> | 216 |
| 6.6.1.2 | <i>Finding a VARMA-ARCH/GARCH model</i> | 218 |
| 6.6.1.3 | <i>The fitted values from VAR(1)-ARCH(1) model</i> | 221 |
| 6.6.2 | Daily log-returns of the SP 500 index and three financial stocks | 222 |
| 6.6.3 | The analysis of the Dow Jones Industrial Average of 30 stocks | 225 |
| | Software code | 229 |
| | Projects | 234 |
| | References | 234 |
| 7 | Repeated measurements | 237 |
| 7.1 | Introduction | 237 |
| 7.2 | Multivariate analysis of variance | 239 |
| 7.2.1 | Test treatment effects | 239 |
| 7.2.2 | Empirical Example I – First analysis on body weight of rats under three different treatments | 241 |
| 7.3 | Models utilizing time series structure | 243 |
| 7.3.1 | Fixed effects model | 243 |

| | | |
|---------|---|-----|
| 7.3.2 | Some common variance–covariance structures | 247 |
| 7.3.3 | Empirical Example II – Further analysis on body weight of rats under three different treatments | 250 |
| 7.3.4 | Random effects and mixed effects models | 252 |
| 7.4 | Nested random effects model | 253 |
| 7.5 | Further generalization and remarks | 254 |
| 7.6 | Another empirical example – the oral condition of neck cancer patients | 255 |
| | Software code | 257 |
| | Projects | 258 |
| | References | 258 |
| 8 | Space–time series models | 261 |
| 8.1 | Introduction | 261 |
| 8.2 | Space–time autoregressive integrated moving average (STARIMA) models | 262 |
| 8.2.1 | Spatial weighting matrix | 262 |
| 8.2.2 | STARIMA models | 265 |
| 8.2.3 | STARMA models | 266 |
| 8.2.4 | ST-ACF and ST-PACF | 267 |
| 8.3 | Generalized space–time autoregressive integrated moving average (GSTARIMA) models | 272 |
| 8.4 | Iterative model building of STARMA and GSTARMA models | 273 |
| 8.5 | Empirical examples | 273 |
| 8.5.1 | Vehicular theft data | 273 |
| 8.5.2 | The annual U.S. labor force count | 279 |
| 8.5.3 | U.S. yearly sexually transmitted disease data | 281 |
| | Software code | 289 |
| | Projects | 298 |
| | References | 298 |
| 9 | Multivariate spectral analysis of time series | 301 |
| 9.1 | Introduction | 301 |
| 9.2 | Spectral representations of multivariate time series processes | 304 |
| 9.3 | The estimation of the spectral density matrix | 309 |
| 9.3.1 | The smoothed spectrum matrix | 309 |
| 9.3.2 | Multitaper smoothing | 313 |
| 9.3.3 | Smoothing spline | 315 |
| 9.3.4 | Bayesian method | 316 |
| 9.3.5 | Penalized Whittle likelihood | 317 |
| 9.3.6 | VARMA spectral estimation | 318 |
| 9.4 | Empirical examples of stationary vector time series | 320 |
| 9.4.1 | Sample spectrum | 320 |
| 9.4.2 | Bayesian method | 325 |
| 9.4.3 | Penalized Whittle likelihood method | 327 |
| 9.4.4 | Example of VAR spectrum estimation | 327 |
| 9.5 | Spectrum analysis of a nonstationary vector time series | 329 |
| 9.5.1 | Introduction | 329 |
| 9.5.2 | Spectrum representations of a nonstationary multivariate process | 331 |
| 9.5.2.1 | <i>Time-varying autoregressive model</i> | 332 |

| | | |
|------------|---|-----|
| 9.5.2.2 | <i>Smoothing spline ANOVA model</i> | 333 |
| 9.5.2.3 | <i>Piecewise vector autoregressive model</i> | 334 |
| 9.5.2.4 | <i>Bayesian methods</i> | 336 |
| 9.6 | Empirical spectrum example of nonstationary vector time series | 337 |
| | Software code | 341 |
| | Projects | 434 |
| | References | 435 |
| 10 | Dimension reduction in high-dimensional multivariate time series analysis | 437 |
| 10.1 | Introduction | 437 |
| 10.2 | Existing methods | 438 |
| 10.2.1 | Regularization methods | 439 |
| 10.2.1.1 | <i>The lasso method</i> | 439 |
| 10.2.1.2 | <i>The lag-weighted lasso method</i> | 440 |
| 10.2.1.3 | <i>The hierarchical vector autoregression (HVAR) method</i> | 440 |
| 10.2.2 | The space–time AR (STAR) model | 442 |
| 10.2.3 | The model-based cluster method | 443 |
| 10.2.4 | The factor analysis | 443 |
| 10.3 | The proposed method for high-dimension reduction | 444 |
| 10.4 | Simulation studies | 446 |
| 10.4.1 | Scenario 1 | 446 |
| 10.4.2 | Scenario 2 | 449 |
| 10.4.3 | Scenario 3 | 449 |
| 10.5 | Empirical examples | 452 |
| 10.5.1 | The macroeconomic time series | 452 |
| 10.5.2 | The yearly U.S. STD data | 457 |
| 10.6 | Further discussions and remarks | 459 |
| 10.6.1 | More on clustering | 459 |
| 10.6.2 | Forming aggregate data through both time domain and frequency domain clustering | 461 |
| 10.6.2.1 | <i>Example of time domain clustering</i> | 461 |
| 10.6.2.2 | <i>Example of frequency domain clustering</i> | 461 |
| 10.6.2.2.1 | <i>Clustering using similarity measures</i> | 463 |
| 10.6.2.2.2 | <i>Clustering by subjective observation</i> | 463 |
| 10.6.2.2.3 | <i>Hierarchical clustering</i> | 463 |
| 10.6.2.2.4 | <i>Nonhierarchical clustering using the K-means method</i> | 463 |
| 10.6.3 | The specification of aggregate matrix and its associated aggregate dimension | 466 |
| 10.6.4 | Be aware of other forms of aggregation | 466 |
| 10.A | Appendix: Parameter estimation results of various procedures | 467 |
| 10.A.1 | Further details of the macroeconomic time series | 467 |
| 10.A.1.1 | <i>VAR(1)</i> | 467 |
| 10.A.1.2 | <i>Lasso</i> | 468 |
| 10.A.1.3 | <i>Componentwise</i> | 470 |
| 10.A.1.4 | <i>Own-other</i> | 471 |
| 10.A.1.5 | <i>Elementwise</i> | 473 |

| | | |
|----------|--|------------|
| 10.A.1.6 | <i>The factor model</i> | 475 |
| 10.A.1.7 | <i>The model-based cluster</i> | 475 |
| 10.A.1.8 | <i>The proposed method</i> | 477 |
| 10.A.2 | Further details of the STD time series | 478 |
| 10.A.2.1 | VAR | 478 |
| 10.A.2.2 | Lasso | 478 |
| 10.A.2.3 | Componentwise | 479 |
| 10.A.2.4 | Own-other | 481 |
| 10.A.2.5 | Elementwise | 482 |
| 10.A.2.6 | <i>The STAR model</i> | 484 |
| 10.A.2.7 | <i>The factor model</i> | 486 |
| 10.A.2.8 | <i>The model-based cluster</i> | 487 |
| 10.A.2.9 | <i>The proposed method</i> | 488 |
| | Software code | 490 |
| | Projects | 505 |
| | References | 506 |
| | <i>Author index</i> | 509 |
| | <i>Subject index</i> | 515 |

About the author

William W.S. Wei is a Professor of Statistical Science at Temple University in Philadelphia, Pennsylvania, United States of America. He earned his BA in Economics from the National Taiwan University (1966), BA in Mathematics from the University of Oregon (1969), and M.S. (1972) and Ph.D. (1974) in Statistics from the University of Wisconsin, Madison. From 1982 to 1987, he was the Chair of the Department of Statistics at Temple University. He has been a Visiting Professor at many universities including the Nankai University in China, National University of Colombia in Colombia, Korea University in Korea, National Sun Yat-Sen University, National Chiao Tung University, and National Taiwan University in Taiwan, and Middle East Technical University in Turkey. His research interests include time series analysis, forecasting methods, high dimensional problems, statistical modeling, and their applications. He has developed new methodologies in seasonal adjustment, aggregation and disaggregation, outlier detection, robust estimation, and vector time series analysis. Some of his most significant contributions include extensive research on the effects of aggregation, methods of measuring information loss due to aggregation, new stochastic procedures of performing data disaggregation, model-free outlier detection techniques, robust methods of estimating autocorrelations, statistics for analyzing multivariate time series, and dimension reduction for high-dimensional time series. His first book, *Time Series Analysis – Univariate and Multivariate Methods*, the first edition published in 1990 and the second edition published in 2006, has been translated into several languages and heavily cited by researchers worldwide. He has just completed his second book, *Multivariate Time Series Analysis and Applications*. He is an active educator and researcher. He has successfully supervised many Ph.D. students, who hold teaching positions at universities or leadership positions in government and industry throughout the world. He is a Fellow of the American Statistical Association, a Fellow of the Royal Statistical Society, and an Elected Member of the International Statistical Institute. He was the 2002 President of ICSA (International Chinese Statistical Association). He is currently an Associate Editor of the *Journal of Forecasting* and the *Journal of Applied Statistical Science*. In addition to teaching and research, he is also active in community service. He served on the educational advisory committee of his local school district, as the chair of the selection committee for a community high school scholarship program, and as the president of several community organizations, including the Taiwanese Hakka Associations of America. Among the many awards he has received are the 2014 Lifetime Achievement Award and the 2016 Musser Award for Excellence in Research from the Temple University Fox School of Business.

Preface

My main research area is time series analysis and forecasting. I have written a book, *Time Series Analysis – Univariate and Multivariate Methods*. Since the first edition was published in 1990 and the second edition in 2006, the book has been used by many researchers and universities worldwide, and I have received many encouraging letters and emails from researchers, instructors, and students about the usefulness of the book in their research and studies. It has been translated into several languages including Chinese, Spanish, and Portuguese. With the development of computers and the internet, we have had a data explosion, and many new theories and methods have been developed in high-dimensional time series analysis. Many publishers have contacted me asking for a new edition. Because of the development of so much new material, it would be impractical to include it all in a new edition of the book. Therefore, I decided to write a new book and call it *Multivariate Time Series Analysis and Applications*. Due to the enthusiasm of multiple publishing editors, I had a difficult time deciding which publisher to choose. After consulting with Dr. Sarkar, our department chair, I decided to choose Wiley. After four stages of rigorous reviews, the Wiley directors unanimously approved the publication of this book.

Many research studies involve multivariate time series. For example, a study of monthly cancer rates in the United States during the past 10 years can involve 50, many hundreds, or even thousands of time series depending on whether we investigate the cancer rates for states, cities, or counties, and a study of the quarterly sales performance of one company's different branches in the country or the world from 2010 to 2018 may involve many hundreds or thousands of time series, depending on the number of products and the number of branches within the company. Multivariate time series analysis methods are needed to properly analyze the data in these studies, which are different from standard statistical theory and methods based on random samples that assume independence. Dependence is the fundamental nature of time series. The use of highly correlated high-dimensional time series data introduces many complications and challenges. The methods and theory to solve these issues will make up the contents of this book. After introducing the fundamental concepts and reviewing the standard vector time series models, we will explore many important issues, including multivariate time series regression, dimension reduction and model simplification, multivariate GARCH (generalized autoregressive conditional heteroskedasticity) models, repeated measurement phenomenon, space–time series modeling, multivariate spectral analysis for both stationary and nonstationary vector time series, and the high-dimension problem in multivariate time series analysis.

In this book, I follow the same fundamental themes of my research with a balanced emphasis on both theory and applications. Methodologies are introduced with proper theoretical justifications and illustrated with simulated and empirical data sets. It should be pointed out that this book is designed for a research-oriented second time series analysis course and therefore standard exercises normally found in the first course will not be provided. Instead, research-oriented projects will be suggested.

I am grateful to the Department of Statistical Science, Fox School of Business, and Temple University for granting me a sabbatical research leave to finish this book. Fox School is celebrating its 100-year anniversary in 2018, and this book is my way of joining in its centennial celebration. I am also grateful to my wife, Susanna, a professor at Saint Joseph's University, sons and daughter, Stephen, Stanley, and Jessica, who are all married physicians, for their help in proofreading my manuscripts even as they are so busy with their own careers and families. My very sincere thanks go to Zeda Li, Andrew Gehman, Nandi Shinjini, and Kaijun Wang, who are either my Ph.D. dissertation students, research/teaching assistants, or both, and Kevin Liu, who was one of our excellent Masters students in the department, for their outstanding assistance in collecting data sets and developing software programs used in the book. Finally, I would like to thank Ms. Alison Oliver, Publications Manager, and Ms. Jemima Kingsly, Project Editor; Ms. Francesca McMahon, Editorial Assistant, Ms. Blesy Regulas, Project Editor; Mr. Mustaq Ahamed Noorullah, Production Editor, all at John Wiley & Sons, and Ms. Lynette Woodward, Editorial & Academic Services, UK, who have been invaluable in completing this important book project.

William W.S. Wei
March 2018

About the companion website

This book is accompanied by a companion website:

www.wiley.com/go/wei/datasets



The website includes:

- Data Appendix (Bookdata)

Scan this QR code to visit the companion website.

