

Wiley Series in Probability and Statistics

MULTIVARIATE TIME SERIES ANALYSIS



With R and Financial Applications

RUEY S. TSAY

Multivariate Time Series Analysis

With R and Financial Applications

RUEY S. TSAY

Booth School of Business
University of Chicago
Chicago, IL



WILEY

Copyright © 2014 by John Wiley & Sons, Inc. All rights reserved

Published by John Wiley & Sons, Inc., Hoboken, New Jersey
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 Section 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, Inc., 222 Rosewood Drive, Danvers, MA 01923, (978) 750-8400, fax (978) 750-4470, or on the web at www.copyright.com. 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, or online at <http://www.wiley.com/go/permission>.

Limit of Liability/Disclaimer of Warranty: While the publisher and author have used their best efforts in preparing this book, they make no representations or warranties with respect to the accuracy or completeness of the contents of this book and specifically disclaim any implied warranties of merchantability or fitness for a particular purpose. No warranty may be created or extended by sales representatives or written sales materials. The advice and strategies contained herein may not be suitable for your situation. You should consult with a professional where appropriate. Neither the publisher nor author shall be liable for any loss of profit or any other commercial damages, including but not limited to special, incidental, consequential, or other damages.

For general information on our other products and services or for technical support, please contact our Customer Care Department within the United States at (800) 762-2974, outside the United States at (317) 572-3993 or fax (317) 572-4002.

Wiley also publishes its books in a variety of electronic formats. Some content that appears in print may not be available in electronic formats. For more information about Wiley products, visit our web site at www.wiley.com.

Library of Congress Cataloging-in-Publication Data:

Tsay, Ruey S., 1951–

Multivariate time series analysis: with R and financial applications / Ruey S. Tsay, Booth School of Business, University of Chicago, Chicago, IL.

pages cm

Includes bibliographical references and index.

ISBN 978-1-118-61790-8 (hardback)

1. Time-series analysis. 2. R (Computer program language) 3. Econometric models. I. Title. QA280.T73 2014 519.5'5–dc23

2014009453

Printed in the United States of America

10 9 8 7 6 5 4 3 2 1

Multivariate Time Series Analysis

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, Harvey Goldstein, Iain M. Johnstone, Geert Molenberghs, David W. Scott, Adrian F. M. Smith, Ruey S. Tsay, Sanford Weisberg*

Editors Emeriti: *Vic Barnett, J. Stuart Hunter, Joseph B. Kadane, Jozef L. Teugels*

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

To my teacher and friend
George

Preface

This book is based on my experience in teaching and research on multivariate time series analysis over the past 30 years. It summarizes the basic concepts and ideas of analyzing multivariate dependent data, provides econometric and statistical models useful for describing the dynamic dependence between variables, discusses the identifiability problem when the models become too flexible, introduces ways to search for simplifying structure hidden in high-dimensional time series, addresses the applicabilities and limitations of multivariate time series methods, and, equally important, develops a software package for readers to apply the methods and models discussed in the book.

Multivariate time series analysis provides useful tools and methods for processing information embedded in multiple measurements that have temporal and cross-sectional dependence. The goal of the analysis is to provide a better understanding of the dynamic relationship between variables and to improve the accuracy in forecasting. The models built can also be used in policy simulation or in making inference. The book focuses mainly on linear models as they are easier to comprehend and widely applicable. I tried to draw a balance between theory and applications and kept the notation as consistent as possible. I also tried to make the book self-contained. However, given the complexity of the subject, the level of coverage on selected topics may vary throughout the book. This reflects in part my own preference and understanding of the topics and in part my desire to keep the book at a reasonable length.

The field of high-dimensional data analysis is still under rapid developments, especially for dependent data. Omission of some important topics or methods is not avoidable for a book like this one. For instance, nonlinear models are not discussed, nor the categorical time series. Readers are advised to consult recent articles or journals for further development.

The book starts with some general concepts of multivariate time series in Chapter 1, including assessing and quantifying temporal and cross-sectional dependence. As the dimension increases, the difficulty in presenting multivariate data quickly becomes evident. I tried to keep the presentation in a compact form if possible. In some cases, scalar summary statistics are given. Chapter 2 focuses on vector

autoregressive (VAR) models as they are, arguably, the most widely used multivariate time series models. My goal is to make the chapter as comprehensive as possible for readers who are interested in VAR models. Both Bayesian and classical analyses of VAR models are included. Chapter 3 studies stationary vector autoregressive moving-average (VARMA) models. It begins with properties and estimation of vector moving-average (VMA) models. The issue of identifiability of VARMA models is investigated and properties of the models are given. Chapter 4 investigates the structural specification of a multivariate time series. Two methods are introduced to seek the simplifying structure hidden in a vector time series. These methods enable users to discover the *skeleton* of a linear multivariate time series. Chapter 5 deals with unit-root nonstationarity and cointegration. It includes the basic theory for understanding unit-root time series and some applications. In Chapter 6, I discuss factor models and some selected topics of multivariate time series. Both the classical and approximate factor models are studied. My goal is to cover all factor models currently available in the literature and to provide the relationship between them. Chapter 7 focuses on multivariate volatility modeling. It covers volatility models that are relatively easy to use and produce positive-definite volatility matrices. The chapter also discusses ways to detect conditional heteroscedasticity in a vector time series and methods for checking a fitted multivariate volatility model. Throughout the book, real examples are used to demonstrate the analysis. Every chapter contains some exercises that analyze empirical vector time series.

Software is an integral part of multivariate time series analysis. Without software packages, multivariate time series becomes a pure theoretical exercise. I have tried my best to write R programs that enable readers to apply all methods and models discussed in the book. These programs are included in the `MTS` package available in R. Readers can duplicate all the analyses shown in the book with the package and some existing R packages. Not a professional programmer, I am certain that many of the codes in `MTS` are not as efficient as they can be and are likely to have bugs. I would appreciate any suggestions and/or corrections to both the package and the book.

RUEY S. TSAY

Chicago, Illinois
September 2014

Acknowledgements

This book would not have been written without the great teachers I have had. In particular, I would like to express my sincere thanks to Professor George C. Tiao who taught me time series analysis and statistical research. His insightful view of empirical time series and his continual encouragements are invaluable. I would like to thank Professor Tea-Yuan Hwang who introduced me to statistics and has remained a close friend over the past four decades. I would also like to thank Mr. Sung-Nan Chen, my junior high school teacher. Without his foresight, I would not have pursued my college education. I would like to thank many other teachers, including late Professor George E. P. Box and late Professor Gregory Reinsel of University of Wisconsin, and friends, including Dr. David F. Findley, Professors Daniel Peña, Manny Parzen, Buddy Gray, and Howell Tong, and late Professor Hirotugu Akaike, for their support of my research in time series analysis. Dr. David Matteson and Mr. Yongning Wang kindly allowed me to use their programs and Yongning has read over the draft carefully. I appreciate their help. I would also like to thank many students who asked informative questions both in and outside the classrooms. I wish to express my sincere thanks to Stephen Quigley and Sari Friedman for their support in preparing this book. I also wish to acknowledge the financial support of Chicago Booth. Finally, I would like to thank my parents who sacrificed so much to support me and for their unconditional love. As always, my children are my inspiration and sources of energy. Finally, I would like to express my sincere thanks to my wife for her love and constant encouragement. In particular, she has always put my career ahead of her own.

The web page of the book is <http://faculty.chicagobooth.edu/ruey.tsay/teaching/mtsbk>.

R. S. T.

Contents

Preface	xv
Acknowledgements	xvii
1 Multivariate Linear Time Series	1
1.1 Introduction, 1	
1.2 Some Basic Concepts, 5	
1.2.1 Stationarity, 5	
1.2.2 Linearity, 6	
1.2.3 Invertibility, 7	
1.3 Cross-Covariance and Correlation Matrices, 8	
1.4 Sample CCM, 9	
1.5 Testing Zero Cross-Correlations, 12	
1.6 Forecasting, 16	
1.7 Model Representations, 18	
1.8 Outline of the Book, 22	
1.9 Software, 23	
Exercises, 23	
2 Stationary Vector Autoregressive Time Series	27
2.1 Introduction, 27	
2.2 VAR(1) Models, 28	
2.2.1 Model Structure and Granger Causality, 28	
2.2.2 Relation to Transfer Function Model, 30	
2.2.3 Stationarity Condition, 31	
2.2.4 Invertibility, 32	
2.2.5 Moment Equations, 32	
2.2.6 Implied Models for the Components, 35	
2.2.7 Moving-Average Representation, 36	

- 2.3 VAR(2) Models, 37
 - 2.3.1 Stationarity Condition, 37
 - 2.3.2 Moment Equations, 38
 - 2.3.3 Implied Marginal Component Models, 40
 - 2.3.4 Moving-Average Representation, 40
- 2.4 VAR(p) Models, 41
 - 2.4.1 A VAR(1) Representation, 41
 - 2.4.2 Stationarity Condition, 42
 - 2.4.3 Moment Equations, 42
 - 2.4.4 Implied Component Models, 43
 - 2.4.5 Moving-Average Representation, 43
- 2.5 Estimation, 44
 - 2.5.1 Least-Squares Methods, 44
 - 2.5.2 Maximum Likelihood Estimate, 47
 - 2.5.3 Limiting Properties of LS Estimate, 49
 - 2.5.4 Bayesian Estimation, 55
- 2.6 Order Selection, 61
 - 2.6.1 Sequential Likelihood Ratio Tests, 61
 - 2.6.2 Information Criteria, 63
- 2.7 Model Checking, 66
 - 2.7.1 Residual Cross-Correlations, 66
 - 2.7.2 Multivariate Portmanteau Statistics, 71
 - 2.7.3 Model Simplification, 72
- 2.8 Linear Constraints, 80
- 2.9 Forecasting, 82
 - 2.9.1 Forecasts of a Given Model, 82
 - 2.9.2 Forecasts of an Estimated Model, 84
- 2.10 Impulse Response Functions, 89
 - 2.10.1 Orthogonal Innovations, 92
- 2.11 Forecast Error Variance Decomposition, 96
- 2.12 Proofs, 98
- Exercises, 100

3 Vector Autoregressive Moving-Average Time Series

105

- 3.1 Vector MA Models, 106
 - 3.1.1 VMA(1) Model, 106
 - 3.1.2 Properties of VMA(q) Models, 110

- 3.2 Specifying VMA Order, 112
- 3.3 Estimation of VMA Models, 113
 - 3.3.1 Conditional Likelihood Estimation, 113
 - 3.3.2 Exact Likelihood Estimation, 116
 - 3.3.3 Initial Parameter Estimation, 126
- 3.4 Forecasting of VMA Models, 126
- 3.5 VARMA Models, 127
 - 3.5.1 Identifiability, 128
 - 3.5.2 VARMA(1,1) Models, 130
 - 3.5.3 Some Properties of VARMA Models, 133
- 3.6 Implications of VARMA Models, 139
 - 3.6.1 Granger Causality, 139
 - 3.6.2 Impulse Response Functions, 141
- 3.7 Linear Transforms of VARMA Processes, 141
- 3.8 Temporal Aggregation of VARMA Processes, 144
- 3.9 Likelihood Function of a VARMA Model, 146
 - 3.9.1 Conditional Likelihood Function, 146
 - 3.9.2 Exact Likelihood Function, 150
 - 3.9.3 Interpreting the Likelihood Function, 152
 - 3.9.4 Computation of Likelihood Function, 154
- 3.10 Innovations Approach to Exact Likelihood Function, 155
 - 3.10.1 Block Cholesky Decomposition, 157
- 3.11 Asymptotic Distribution of Maximum Likelihood Estimates, 160
 - 3.11.1 Linear Parameter Constraints, 162
- 3.12 Model Checking of Fitted VARMA Models, 163
- 3.13 Forecasting of VARMA Models, 164
 - 3.13.1 Forecasting Updating, 166
- 3.14 Tentative Order Identification, 166
 - 3.14.1 Consistent AR Estimates, 166
 - 3.14.2 Extended Cross-Correlation Matrices, 169
 - 3.14.3 A Summary Two-Way Table, 171
- 3.15 Empirical Analysis of VARMA Models, 176
 - 3.15.1 Personal Income and Expenditures, 176
 - 3.15.2 Housing Starts and Mortgage Rate, 184
- 3.16 Appendix, 192
- Exercises, 194

4 Structural Specification of VARMA Models 199

- 4.1 The Kronecker Index Approach, 200
 - 4.1.1 A Predictive Interpretation, 205
 - 4.1.2 A VARMA Specification, 207
 - 4.1.3 An Illustrative Example, 208
 - 4.1.4 The Echelon Form, 211
 - 4.1.5 The Example Continued, 212
- 4.2 The Scalar Component Approach, 212
 - 4.2.1 Scalar Component Models, 213
 - 4.2.2 Model Specification Via Scalar Component Models, 215
 - 4.2.3 Redundant Parameters, 216
 - 4.2.4 VARMA Model Specification, 218
 - 4.2.5 The Transformation Matrix, 218
- 4.3 Statistics for Order Specification, 220
 - 4.3.1 Reduced Rank Tests, 220
- 4.4 Finding Kronecker Indices, 222
 - 4.4.1 Application, 224
- 4.5 Finding Scalar Component Models, 226
 - 4.5.1 Implication of Scalar Component Models, 227
 - 4.5.2 Exchangeable Scalar Component Models, 229
 - 4.5.3 Searching for Scalar Components, 232
 - 4.5.4 Application, 233
- 4.6 Estimation, 237
 - 4.6.1 Illustration of the Kronecker Index Approach, 238
 - 4.6.2 Illustration of the SCM Approach, 241
- 4.7 An Example, 245
 - 4.7.1 The SCM Approach, 245
 - 4.7.2 The Kronecker Index Approach, 252
 - 4.7.3 Discussion and Comparison, 257
- 4.8 Appendix: Canonical Correlation Analysis, 259
- Exercises, 262

5 Unit-Root Nonstationary Processes 265

- 5.1 Univariate Unit-Root Processes, 266
 - 5.1.1 Motivation, 267
 - 5.1.2 Unit Root with Stationary Innovations, 269
 - 5.1.3 AR(1) Case, 274

- 5.1.4 AR(p) Case, 274
- 5.1.5 MA(1) Case, 276
- 5.1.6 Unit-Root Tests, 276
- 5.1.7 Example, 277
- 5.2 Multivariate Unit-Root Processes, 279
 - 5.2.1 An Alternative Model Representation, 282
 - 5.2.2 Unit-Root VAR Processes, 285
- 5.3 Spurious Regressions, 290
- 5.4 Multivariate Exponential Smoothing, 291
- 5.5 Cointegration, 294
 - 5.5.1 An Example of Cointegration, 295
 - 5.5.2 Some Justifications of Cointegration, 297
- 5.6 An Error-Correction Form, 297
- 5.7 Implications of Cointegrating Vectors, 300
 - 5.7.1 Implications of the Deterministic Term, 300
 - 5.7.2 Implications for Moving-Average Representation, 301
- 5.8 Parameterization of Cointegrating Vectors, 302
- 5.9 Cointegration Tests, 303
 - 5.9.1 The Case of VAR Models, 303
 - 5.9.2 Specification of Deterministic Terms, 304
 - 5.9.3 Review of Likelihood Ratio Tests, 305
 - 5.9.4 Cointegration Tests of VAR Models, 306
 - 5.9.5 An Illustration, 309
 - 5.9.6 Cointegration Tests of VARMA Models, 313
- 5.10 Estimation of Error-Correction Models, 313
 - 5.10.1 VAR Models, 313
 - 5.10.2 Reduced Regression Method, 317
 - 5.10.3 VARMA Models, 318
- 5.11 Applications, 319
- 5.12 Discussion, 326
- 5.13 Appendix, 327
- Exercises, 328

6 Factor Models and Selected Topics

333

- 6.1 Seasonal Models, 333
- 6.2 Principal Component Analysis, 341
- 6.3 Use of Exogenous Variables, 345

- 6.3.1 VARX Models, 346
- 6.3.2 Regression Model, 352
- 6.4 Missing Values, 357
 - 6.4.1 Completely Missing, 358
 - 6.4.2 Partially Missing, 361
- 6.5 Factor Models, 364
 - 6.5.1 Orthogonal Factor Models, 364
 - 6.5.2 Approximate Factor Models, 370
 - 6.5.3 Diffusion Index Models, 371
 - 6.5.4 Dynamic Factor Models, 375
 - 6.5.5 Constrained Factor Models, 376
 - 6.5.6 Asymptotic Principal Component Analysis, 380
- 6.6 Classification and Clustering Analysis, 386
 - 6.6.1 Clustering Analysis, 386
 - 6.6.2 Bayesian Estimation, 387
 - 6.6.3 An MCMC Procedure, 390
- Exercises, 394

7 Multivariate Volatility Models

399

- 7.1 Testing Conditional Heteroscedasticity, 401
 - 7.1.1 Portmanteau Test, 401
 - 7.1.2 Rank-Based Test, 402
 - 7.1.3 Simulation, 403
 - 7.1.4 Application, 405
- 7.2 Estimation of Multivariate Volatility Models, 407
- 7.3 Diagnostic Checks of Volatility Models, 409
 - 7.3.1 Ling and Li Statistics, 409
 - 7.3.2 Tse Statistics, 412
- 7.4 Exponentially Weighted Moving Average, 414
- 7.5 BEKK Models, 417
 - 7.5.1 Discussion, 420
- 7.6 Cholesky Decomposition and Volatility Modeling, 420
 - 7.6.1 Volatility Modeling, 422
 - 7.6.2 Application, 423
- 7.7 Dynamic Conditional Correlation Models, 428
 - 7.7.1 A Procedure for Building DCC Models, 430
 - 7.7.2 An Example, 430
- 7.8 Orthogonal Transformation, 434

- 7.8.1 The Go-GARCH Model, 434
- 7.8.2 Dynamic Orthogonal Components, 438
- 7.8.3 Testing the Existence of DOC, 440
- 7.9 Copula-Based Models, 443
 - 7.9.1 Copulas, 444
 - 7.9.2 Gaussian and t -Copulas, 445
 - 7.9.3 Multivariate Volatility Modeling, 449
- 7.10 Principal Volatility Components, 454
 - 7.10.1 Sample Principal Volatility Components, 458
- Exercises, 461

Appendix A Review of Mathematics and Statistics

465

- A.1 Review of Vectors and Matrices, 465
 - A.1.1 Basic Operations, 466
 - A.1.2 Inverse, Trace, Eigenvalue, and Eigenvector, 467
 - A.1.3 Positive-Definite Matrix, 467
 - A.1.4 Comparison of Two Symmetric Matrices, 469
 - A.1.5 Idempotent Matrix, 469
 - A.1.6 Cholesky Decomposition, 469
 - A.1.7 Partition of a Matrix, 470
 - A.1.8 Vectorization and Kronecker Product, 470
 - A.1.9 Vector and Matrix Differentiation, 472
- A.2 Least-Squares Estimation, 477
- A.3 Multivariate Normal Distributions, 478
- A.4 Multivariate Student- t Distribution, 479
- A.5 Wishart and Inverted Wishart Distributions, 480
- A.6 Vector and Matrix Differentials, 481
 - A.6.1 Scalar Function, 481
 - A.6.2 Vector Function, 484
 - A.6.3 Matrix Function, 485
 - A.6.4 Properties of Matrix Differential, 485
 - A.6.5 Application, 486

Index

489

Multivariate Linear Time Series

1.1 INTRODUCTION

Multivariate time series analysis considers simultaneously multiple time series. It is a branch of multivariate statistical analysis but deals specifically with dependent data. It is, in general, much more complicated than the univariate time series analysis, especially when the number of series considered is large. We study this more complicated statistical analysis in this book because in real life decisions often involve multiple inter-related factors or variables. Understanding the relationships between those factors and providing accurate predictions of those variables are valuable in decision making. The objectives of multivariate time series analysis thus include

1. To study the dynamic relationships between variables
2. To improve the accuracy of prediction

Let $\mathbf{z}_t = (z_{1t}, \dots, z_{kt})'$ be a k -dimensional time series observed at equally spaced time points. For example, let z_{1t} be the quarterly U.S. real gross domestic product (GDP) and z_{2t} the quarterly U.S. civilian unemployment rate. By studying z_{1t} and z_{2t} jointly, we can assess the temporal and contemporaneous dependence between GDP and unemployment rate. In this particular case, $k = 2$ and the two variables are known to be instantaneously negatively correlated. Figure 1.1 shows the time plots of quarterly U.S. real GDP (in logarithm of billions of chained 2005 dollars) and unemployment rate, obtained via monthly data with averaging, from 1948 to 2011. Both series are seasonally adjusted. Figure 1.2 shows the time plots of the real GDP growth rate and the changes in unemployment rate from the second quarter of 1948 to the fourth quarter of 2011. Figure 1.3 shows the scatter plot of the two time series given in Figure 1.2. From these figures, we can see that the GDP