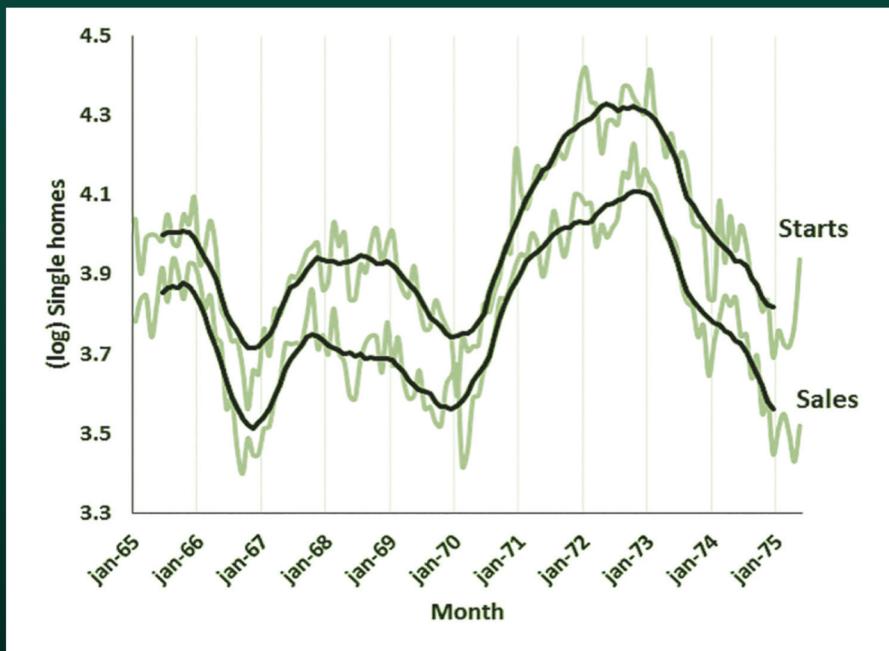


State-Space Methods for Time Series Analysis

Theory, Applications and Software



Jose Casals
Alfredo Garcia-Hiernaux
Miguel Jerez
Sonia Sotoca
A. Alexandre Trindade



CRC Press

Taylor & Francis Group

A CHAPMAN & HALL BOOK

State-Space Methods for Time Series Analysis

Theory, Applications and Software

MONOGRAPHS ON STATISTICS AND APPLIED PROBABILITY

General Editors

F. Bunea, V. Isham, N. Keiding, T. Louis, R. L. Smith, and H. Tong

1. Stochastic Population Models in Ecology and Epidemiology *M.S. Barlett* (1960)
2. Queues *D.R. Cox and W.L. Smith* (1961)
3. Monte Carlo Methods *J.M. Hammersley and D.C. Handscomb* (1964)
4. The Statistical Analysis of Series of Events *D.R. Cox and P.A.W. Lewis* (1966)
5. Population Genetics *W.J. Ewens* (1969)
6. Probability, Statistics and Time *M.S. Barlett* (1975)
7. Statistical Inference *S.D. Silvey* (1975)
8. The Analysis of Contingency Tables *B.S. Everitt* (1977)
9. Multivariate Analysis in Behavioural Research *A.E. Maxwell* (1977)
10. Stochastic Abundance Models *S. Engen* (1978)
11. Some Basic Theory for Statistical Inference *E.J.G. Pitman* (1979)
12. Point Processes *D.R. Cox and V. Isham* (1980)
13. Identification of Outliers *D.M. Hawkins* (1980)
14. Optimal Design *S.D. Silvey* (1980)
15. Finite Mixture Distributions *B.S. Everitt and D.J. Hand* (1981)
16. Classification *A.D. Gordon* (1981)
17. Distribution-Free Statistical Methods, 2nd edition *J.S. Maritz* (1995)
18. Residuals and Influence in Regression *R.D. Cook and S. Weisberg* (1982)
19. Applications of Queueing Theory, 2nd edition *G.F. Newell* (1982)
20. Risk Theory, 3rd edition *R.E. Beard, T. Pentikäinen and E. Pesonen* (1984)
21. Analysis of Survival Data *D.R. Cox and D. Oakes* (1984)
22. An Introduction to Latent Variable Models *B.S. Everitt* (1984)
23. Bandit Problems *D.A. Berry and B. Fristedt* (1985)
24. Stochastic Modelling and Control *M.H.A. Davis and R. Vinter* (1985)
25. The Statistical Analysis of Composition Data *J. Aitchison* (1986)
26. Density Estimation for Statistics and Data Analysis *B.W. Silverman* (1986)
27. Regression Analysis with Applications *G.B. Wetherill* (1986)
28. Sequential Methods in Statistics, 3rd edition *G.B. Wetherill and K.D. Glazebrook* (1986)
29. Tensor Methods in Statistics *P. McCullagh* (1987)
30. Transformation and Weighting in Regression *R.J. Carroll and D. Ruppert* (1988)
31. Asymptotic Techniques for Use in Statistics *O.E. Bandorff-Nielsen and D.R. Cox* (1989)
32. Analysis of Binary Data, 2nd edition *D.R. Cox and E.J. Snell* (1989)
33. Analysis of Infectious Disease Data *N.G. Becker* (1989)
34. Design and Analysis of Cross-Over Trials *B. Jones and M.G. Kenward* (1989)
35. Empirical Bayes Methods, 2nd edition *J.S. Maritz and T. Lwin* (1989)
36. Symmetric Multivariate and Related Distributions *K.T. Fang, S. Kotz and K.W. Ng* (1990)
37. Generalized Linear Models, 2nd edition *P. McCullagh and J.A. Nelder* (1989)
38. Cyclic and Computer Generated Designs, 2nd edition *J.A. John and E.R. Williams* (1995)
39. Analog Estimation Methods in Econometrics *C.F. Manski* (1988)
40. Subset Selection in Regression *A.J. Miller* (1990)
41. Analysis of Repeated Measures *M.J. Crowder and D.J. Hand* (1990)
42. Statistical Reasoning with Imprecise Probabilities *P. Walley* (1991)
43. Generalized Additive Models *T.J. Hastie and R.J. Tibshirani* (1990)
44. Inspection Errors for Attributes in Quality Control *N.L. Johnson, S. Kotz and X. Wu* (1991)
45. The Analysis of Contingency Tables, 2nd edition *B.S. Everitt* (1992)
46. The Analysis of Quantal Response Data *B.J.T. Morgan* (1992)
47. Longitudinal Data with Serial Correlation—A State-Space Approach *R.H. Jones* (1993)

48. Differential Geometry and Statistics *M.K. Murray and J.W. Rice* (1993)
49. Markov Models and Optimization *M.H.A. Davis* (1993)
50. Networks and Chaos—Statistical and Probabilistic Aspects
O.E. Barndorff-Nielsen, J.L. Jensen and W.S. Kendall (1993)
51. Number-Theoretic Methods in Statistics *K.-T. Fang and Y. Wang* (1994)
52. Inference and Asymptotics *O.E. Barndorff-Nielsen and D.R. Cox* (1994)
53. Practical Risk Theory for Actuaries *C.D. Daykin, T. Pentikäinen and M. Pesonen* (1994)
54. Biplots *J.C. Gower and D.J. Hand* (1996)
55. Predictive Inference—An Introduction *S. Geisser* (1993)
56. Model-Free Curve Estimation *M.E. Tarter and M.D. Lock* (1993)
57. An Introduction to the Bootstrap *B. Efron and R.J. Tibshirani* (1993)
58. Nonparametric Regression and Generalized Linear Models *P.J. Green and B.W. Silverman* (1994)
59. Multidimensional Scaling *T.F. Cox and M.A.A. Cox* (1994)
60. Kernel Smoothing *M.P. Wand and M.C. Jones* (1995)
61. Statistics for Long Memory Processes *J. Beran* (1995)
62. Nonlinear Models for Repeated Measurement Data *M. Davidian and D.M. Giltinan* (1995)
63. Measurement Error in Nonlinear Models *R.J. Carroll, D. Rupert and L.A. Stefanski* (1995)
64. Analyzing and Modeling Rank Data *J.J. Marden* (1995)
65. Time Series Models—In Econometrics, Finance and Other Fields
D.R. Cox, D.V. Hinkley and O.E. Barndorff-Nielsen (1996)
66. Local Polynomial Modeling and its Applications *J. Fan and I. Gijbels* (1996)
67. Multivariate Dependencies—Models, Analysis and Interpretation *D.R. Cox and N. Wermuth* (1996)
68. Statistical Inference—Based on the Likelihood *A. Azzalini* (1996)
69. Bayes and Empirical Bayes Methods for Data Analysis *B.P. Carlin and T.A. Louis* (1996)
70. Hidden Markov and Other Models for Discrete-Valued Time Series *I.L. MacDonald and W. Zucchini* (1997)
71. Statistical Evidence—A Likelihood Paradigm *R. Royall* (1997)
72. Analysis of Incomplete Multivariate Data *J.L. Schafer* (1997)
73. Multivariate Models and Dependence Concepts *H. Joe* (1997)
74. Theory of Sample Surveys *M.E. Thompson* (1997)
75. Retrial Queues *G. Falin and J.G.C. Templeton* (1997)
76. Theory of Dispersion Models *B. Jørgensen* (1997)
77. Mixed Poisson Processes *J. Grandell* (1997)
78. Variance Components Estimation—Mixed Models, Methodologies and Applications *P.S.R.S. Rao* (1997)
79. Bayesian Methods for Finite Population Sampling *G. Meeden and M. Ghosh* (1997)
80. Stochastic Geometry—Likelihood and computation
O.E. Barndorff-Nielsen, W.S. Kendall and M.N.M. van Lieshout (1998)
81. Computer-Assisted Analysis of Mixtures and Applications—Meta-Analysis, Disease Mapping and Others
D. Böhning (1999)
82. Classification, 2nd edition *A.D. Gordon* (1999)
83. Semimartingales and their Statistical Inference *B.L.S. Prakasa Rao* (1999)
84. Statistical Aspects of BSE and vCJD—Models for Epidemics *C.A. Donnelly and N.M. Ferguson* (1999)
85. Set-Indexed Martingales *G. Ivanoff and E. Merzbach* (2000)
86. The Theory of the Design of Experiments *D.R. Cox and N. Reid* (2000)
87. Complex Stochastic Systems *O.E. Barndorff-Nielsen, D.R. Cox and C. Klüppelberg* (2001)
88. Multidimensional Scaling, 2nd edition *T.F. Cox and M.A.A. Cox* (2001)
89. Algebraic Statistics—Computational Commutative Algebra in Statistics
G. Pistone, E. Riccomagno and H.P. Wynn (2001)
90. Analysis of Time Series Structure—SSA and Related Techniques
N. Golyandina, V. Nekrutkin and A.A. Zhigljavsky (2001)
91. Subjective Probability Models for Lifetimes *Fabio Spizzichino* (2001)
92. Empirical Likelihood *Art B. Owen* (2001)
93. Statistics in the 21st Century *Adrian E. Raftery, Martin A. Tanner, and Martin T. Wells* (2001)
94. Accelerated Life Models: Modeling and Statistical Analysis
Vilijandas Bagdonavicius and Mikhail Nikulin (2001)

95. Subset Selection in Regression, Second Edition *Alan Miller* (2002)
96. Topics in Modelling of Clustered Data *Marc Aerts, Helena Geys, Geert Molenberghs, and Louise M. Ryan* (2002)
97. Components of Variance *D.R. Cox and P.J. Solomon* (2002)
98. Design and Analysis of Cross-Over Trials, 2nd Edition *Byron Jones and Michael G. Kenward* (2003)
99. Extreme Values in Finance, Telecommunications, and the Environment
Bärbel Finkenstädt and Holger Rootzén (2003)
100. Statistical Inference and Simulation for Spatial Point Processes
Jesper Møller and Rasmus Plenge Waagepetersen (2004)
101. Hierarchical Modeling and Analysis for Spatial Data
Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand (2004)
102. Diagnostic Checks in Time Series *Wai Keung Li* (2004)
103. Stereology for Statisticians *Adrian Baddeley and Eva B. Vedel Jensen* (2004)
104. Gaussian Markov Random Fields: Theory and Applications *Håvard Rue and Leonhard Held* (2005)
105. Measurement Error in Nonlinear Models: A Modern Perspective, Second Edition
Raymond J. Carroll, David Ruppert, Leonard A. Stefanski, and Ciprian M. Crainiceanu (2006)
106. Generalized Linear Models with Random Effects: Unified Analysis via H-likelihood
Youngjo Lee, John A. Nelder, and Yudi Pawitan (2006)
107. Statistical Methods for Spatio-Temporal Systems
Bärbel Finkenstädt, Leonhard Held, and Valerie Isham (2007)
108. Nonlinear Time Series: Semiparametric and Nonparametric Methods *Jiti Gao* (2007)
109. Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis
Michael J. Daniels and Joseph W. Hogan (2008)
110. Hidden Markov Models for Time Series: An Introduction Using R
Walter Zucchini and Iain L. MacDonald (2009)
111. ROC Curves for Continuous Data *Wojtek J. Krzanowski and David J. Hand* (2009)
112. Antedependence Models for Longitudinal Data *Dale L. Zimmerman and Vicente A. Niñez-Antón* (2009)
113. Mixed Effects Models for Complex Data *Lang Wu* (2010)
114. Introduction to Time Series Modeling *Genshiro Kitagawa* (2010)
115. Expansions and Asymptotics for Statistics *Christopher G. Small* (2010)
116. Statistical Inference: An Integrated Bayesian/Likelihood Approach *Murray Aitkin* (2010)
117. Circular and Linear Regression: Fitting Circles and Lines by Least Squares *Nikolai Chernov* (2010)
118. Simultaneous Inference in Regression *Wei Liu* (2010)
119. Robust Nonparametric Statistical Methods, Second Edition
Thomas P. Hettmansperger and Joseph W. McKean (2011)
120. Statistical Inference: The Minimum Distance Approach
Ayanendranath Basu, Hiroyuki Shioya, and Chanseok Park (2011)
121. Smoothing Splines: Methods and Applications *Yuedong Wang* (2011)
122. Extreme Value Methods with Applications to Finance *Serguei Y. Novak* (2012)
123. Dynamic Prediction in Clinical Survival Analysis *Hans C. van Houwelingen and Hein Putter* (2012)
124. Statistical Methods for Stochastic Differential Equations
Mathieu Kessler, Alexander Lindner, and Michael Sørensen (2012)
125. Maximum Likelihood Estimation for Sample Surveys
R. L. Chambers, D. G. Steel, Suojin Wang, and A. H. Welsh (2012)
126. Mean Field Simulation for Monte Carlo Integration *Pierre Del Moral* (2013)
127. Analysis of Variance for Functional Data *Jin-Ting Zhang* (2013)
128. Statistical Analysis of Spatial and Spatio-Temporal Point Patterns, Third Edition *Peter J. Diggle* (2013)
129. Constrained Principal Component Analysis and Related Techniques *Yoshio Takane* (2014)
130. Randomised Response-Adaptive Designs in Clinical Trials *Anthony C. Atkinson and Atanu Biswas* (2014)
131. Theory of Factorial Design: Single- and Multi-Stratum Experiments *Ching-Shui Cheng* (2014)
132. Quasi-Least Squares Regression *Justine Shults and Joseph M. Hilbe* (2014)
133. Data Analysis and Approximate Models: Model Choice, Location-Scale, Analysis of Variance, Nonparametric Regression and Image Analysis *Laurie Davies* (2014)
134. Dependence Modeling with Copulas *Harry Joe* (2014)
135. Hierarchical Modeling and Analysis for Spatial Data, Second Edition *Sudipto Banerjee, Bradley P. Carlin, and Alan E. Gelfand* (2014)

136. Sequential Analysis: Hypothesis Testing and Changepoint Detection *Alexander Tartakovsky, Igor Nikiforov, and Michèle Basseville* (2015)
137. Robust Cluster Analysis and Variable Selection *Gunter Ritter* (2015)
138. Design and Analysis of Cross-Over Trials, Third Edition *Byron Jones and Michael G. Kenward* (2015)
139. Introduction to High-Dimensional Statistics *Christophe Giraud* (2015)
140. Pareto Distributions: Second Edition *Barry C. Arnold* (2015)
141. Bayesian Inference for Partially Identified Models: Exploring the Limits of Limited Data *Paul Gustafson* (2015)
142. Models for Dependent Time Series *Granville Tunnicliffe Wilson, Marco Reale, John Haywood* (2015)
143. Statistical Learning with Sparsity: The Lasso and Generalizations *Trevor Hastie, Robert Tibshirani, and Martin Wainwright* (2015)
144. Measuring Statistical Evidence Using Relative Belief *Michael Evans* (2015)
145. Stochastic Analysis for Gaussian Random Processes and Fields: With Applications *Vidyadhar S. Mandrekar and Leszek Gawarecki* (2015)
146. Semialgebraic Statistics and Latent Tree Models *Piotr Zwiernik* (2015)
147. Inferential Models: Reasoning with Uncertainty *Ryan Martin and Chuanhai Liu* (2016)
148. Perfect Simulation *Mark L. Huber* (2016)
149. State-Space Methods for Time Series Analysis: Theory, Applications and Software
Jose Casals, Alfredo Garcia-Hiernaux, Miguel Jerez, Sonia Sotoca, and A. Alexandre Trindade (2016)

This page intentionally left blank

Monographs on Statistics and Applied Probability 149

State-Space Methods for Time Series Analysis

Theory, Applications and Software

Jose Casals

Universidad Complutense de Madrid, Spain

Alfredo Garcia-Hiernaux

Universidad Complutense de Madrid, Spain

Miguel Jerez

Universidad Complutense de Madrid, Spain

Sonia Sotoca

Universidad Complutense de Madrid, Spain

A. Alexandre Trindade

Texas Tech University, Lubbock, U.S.A.



CRC Press

Taylor & Francis Group

Boca Raton London New York

CRC Press is an imprint of the
Taylor & Francis Group, an **informa** business
A CHAPMAN & HALL BOOK

CRC Press
Taylor & Francis Group
6000 Broken Sound Parkway NW, Suite 300
Boca Raton, FL 33487-2742

© 2016 by Taylor & Francis Group, LLC
CRC Press is an imprint of Taylor & Francis Group, an Informa business

No claim to original U.S. Government works
Version Date: 20160211

International Standard Book Number-13: 978-1-4822-1960-9 (eBook - PDF)

This book contains information obtained from authentic and highly regarded sources. Reasonable efforts have been made to publish reliable data and information, but the author and publisher cannot assume responsibility for the validity of all materials or the consequences of their use. The authors and publishers have attempted to trace the copyright holders of all material reproduced in this publication and apologize to copyright holders if permission to publish in this form has not been obtained. If any copyright material has not been acknowledged please write and let us know so we may rectify in any future reprint.

Except as permitted under U.S. Copyright Law, no part of this book may be reprinted, reproduced, transmitted, or utilized in any form by any electronic, mechanical, or other means, now known or hereafter invented, including photocopying, microfilming, and recording, or in any information storage or retrieval system, without written permission from the publishers.

For permission to photocopy or use material electronically from this work, please access www.copyright.com (<http://www.copyright.com/>) or contact the Copyright Clearance Center, Inc. (CCC), 222 Rosewood Drive, Danvers, MA 01923, 978-750-8400. CCC is a not-for-profit organization that provides licenses and registration for a variety of users. For organizations that have been granted a photocopy license by the CCC, a separate system of payment has been arranged.

Trademark Notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Visit the Taylor & Francis Web site at
<http://www.taylorandfrancis.com>

and the CRC Press Web site at
<http://www.crcpress.com>

A wise man once said that books by young writers should be dedicated to their parents, while mature authors must bestow this honor on their children.

Sadly this book is dedicated, with love, to our children.

This page intentionally left blank

Contents

List of Figures	xvii
List of Tables	xix
Preface	xxiii
About the Authors	xxvii
1 Introduction	1
2 Linear State-Space models	5
2.1 The multiple error model	6
2.1.1 Model formulation	6
2.1.2 Similar transformations	8
2.1.3 Properties of the State-Space model	8
2.2 Single error models	12
2.2.1 Model formulation	12
2.2.2 State estimation in the SEM	12
2.2.3 Obtaining the SEM equivalent to a MEM	13
2.2.4 Obtaining the SEM equivalent to a general linear process	15
3 Model transformations	19
3.1 Model decomposition	19
3.1.1 Deterministic and stochastic subsystems	19
3.1.2 Implied univariate models	21
3.1.3 Block-diagonal forms	22
3.2 Model combination	24
3.2.1 Endogeneization of stochastic inputs	24
3.2.2 Colored errors	26
3.3 Change of variables in the output	28
3.3.1 Observation errors	28
3.3.2 Missing values and aggregated data	29
3.3.3 Temporal aggregation	30
3.4 Uses of these transformations	31

4 Filtering and Smoothing	33
4.1 The conditional moments of a State-Space model	33
4.2 The Kalman Filter	34
4.3 Decomposition of the smoothed moments	36
4.4 Smoothing for a general State-Space model	37
4.5 Smoothing for fixed-coefficients and single-error models	39
4.6 Uncertainty of the smoothed estimates in a SEM	40
4.7 Examples	42
4.7.1 Recursive Least Squares	42
4.7.2 Cleaning the Wolf sunspot series	42
4.7.3 Extracting the Hodrick-Prescott trend	44
5 Likelihood computation for fixed-coefficients models	45
5.1 Maximum likelihood estimation	45
5.1.1 Problem statement	45
5.1.2 Prediction error decomposition	46
5.1.3 Initialization of the Kalman filter in the stationary case	47
5.2 The likelihood for a nonstationary model	48
5.2.1 Diffuse likelihood	49
5.2.2 Minimally conditioned likelihood	50
5.2.3 Likelihood computation for a fixed-parameter SEM	52
5.2.4 Initialization of the Kalman filter in the nonstationary case	52
5.3 The likelihood for a model with inputs	54
5.3.1 Models with deterministic inputs	54
5.3.2 Models with stochastic inputs	55
5.4 Examples	57
5.4.1 Models for the Airline Passenger series	57
5.4.2 Modeling the series of Housing Starts and Sales	60
6 The likelihood of models with varying parameters	65
6.1 Regression with time-varying parameters	66
6.1.1 SS formulation	66
6.1.2 Maximum likelihood estimation	67
6.2 Periodic models	68
6.2.1 All the seasons have the same dynamic structure	69
6.2.2 The s models do not have the same dynamic structure	70
6.2.3 Stationarity and invertibility	71
6.2.4 Maximum likelihood estimation	72
6.3 The likelihood of models with GARCH errors	74
6.4 Examples	76
6.4.1 A time-varying CAPM regression	76
6.4.2 A periodic model for West German Consumption	78
6.4.3 A model with vector-GARCH errors for two exchange rate series	79

Contents	xiii
7 Subspace methods	83
7.1 Theoretical foundations	83
7.1.1 Subspace structure and notation	84
7.1.2 Assumptions, projections and model reduction	86
7.1.3 Estimating the system matrices	87
7.2 System order estimation	89
7.2.1 Preliminary data analysis methods	89
7.2.2 Model comparison methods	91
7.3 Constrained estimation	92
7.3.1 State sequence structure	92
7.3.2 Subspace-based likelihood	93
7.4 Multiplicative seasonal models	94
7.5 Examples	95
7.5.1 Univariate models	95
7.5.2 A multivariate model for the interest rates	99
8 Signal extraction	107
8.1 Input and error-related components	108
8.1.1 The deterministic and stochastic subsystems	108
8.1.2 Enforcing minimality	109
8.2 Estimation of the deterministic components	110
8.2.1 Estimating the total effect of the inputs	111
8.2.2 Estimating the individual effect of each input	113
8.3 Decomposition of the stochastic component	114
8.3.1 Characterization of the structural components	114
8.3.2 Estimation of the structural components	116
8.4 Structure of the method	116
8.5 Examples	116
8.5.1 Comparing different methods with simulated data	116
8.5.2 Common features in wheat prices	120
8.5.3 The effect of advertising on sales	123
9 The VARMAX representation of a State-Space model	129
9.1 Notation and previous results	130
9.2 Obtaining the VARMAX form of a State-Space model	131
9.2.1 From State-Space to standard VARMAX	132
9.2.2 From State-Space to canonical VARMAX	133
9.3 Practical applications and examples	135
9.3.1 The VARMAX form of some common State-Space models	135
9.3.2 Identifiability and conditioning of the estimates	135
9.3.3 Fitting an errors-in-variables model to Wolf's sunspot series	139
9.3.4 "Bottom-up" modeling of quarterly US GDP trend	140

10 Aggregation and disaggregation of time series	145
10.1 The effect of aggregation on an SS model	146
10.1.1 The high-frequency model in stacked form	146
10.1.2 Aggregation relationships	148
10.1.3 Relationships between the models for high, low, and mixed-frequency data	149
10.1.4 The effect of aggregation on predictive accuracy	150
10.2 Observability in the aggregated model	150
10.2.1 Unobservable modes	150
10.2.2 Observability and fixed-interval smoothing	152
10.2.3 An algorithm to aggregate a linear model: theory and examples	153
10.3 Specification of the high-frequency model	154
10.3.1 Enforcing approximate consistency	154
10.3.2 “Bottom-up” determination of the quarterly model	158
10.4 Empirical example	158
10.4.1 Annual model	159
10.4.2 Decomposition of the quarterly indicator	159
10.4.3 Specification and estimation of the quarterly model	160
10.4.4 Diagnostics	160
10.4.5 Forecast accuracy and non-conformable samples	162
10.4.6 Comparison with alternative methods	163
11 Cross-sectional extension: longitudinal and panel data	167
11.1 Model formulation	168
11.2 The Kalman Filter	169
11.2.1 Case of uncorrelated state and observational errors	170
11.2.2 Case of correlated state and observational errors	171
11.3 The linear mixed model in SS form	171
11.4 Maximum likelihood estimation	174
11.5 Missing data modifications	176
11.5.1 Missingness in responses only	176
11.5.2 Missingness in both responses and covariates: method 1	178
11.5.3 Missingness in both responses and covariates: method 2	180
11.6 Real data examples	182
11.6.1 A LMM for the mare ovarian follicles data	182
11.6.2 Smoothing and prediction of missing values for the beluga whales data	183

Contents	xv
Appendices	191
A Some results in numerical algebra and linear systems	193
A.1 QR Decomposition	193
A.2 Schur decomposition	195
A.3 The Hessenberg Form	196
A.4 SVD Decomposition	197
A.5 Canonical Correlations	198
A.6 Algebraic Lyapunov and Sylvester equations	198
A.7 Numerical solution of a Sylvester equation	200
A.8 Block-diagonalization of a matrix	201
A.9 Reduced rank least squares	202
A.10 Riccati equations	203
A.10.1 Definition	204
A.10.2 Solving the ARE in the general case	205
A.10.3 Solving the ARE for GARCH models	207
A.11 Kalman filter	208
B Asymptotic properties of maximum likelihood estimates	211
B.1 Preliminaries	211
B.2 Basic likelihood results for the State-Space model	214
B.2.1 The information matrix	214
B.2.2 Regularity conditions	216
B.2.3 Choice of estimation method	217
B.3 The State-Space model with cross-sectional extension	218
C Software (E^4)	223
C.1 Models supported in E^4	224
C.1.1 State-Space model	224
C.1.2 The THD format	224
C.1.3 Basic models	226
C.1.3.1 Mathematical definition	226
C.1.3.2 Definition in THD format	227
C.1.4 Models with GARCH errors	228
C.1.4.1 Mathematical definition of the GARCH process	228
C.1.4.2 Defining a model with GARCH errors in THD format	229
C.1.5 Nested models	230
C.2 Overview of computational procedures	233
C.2.1 Standard procedures for time series analysis	233
C.2.2 Signal extraction methods	235
C.2.3 Likelihood and model estimation	238
C.3 Who can benefit from E^4 ?	241

D Downloading E^4 and the examples in this book	243
D.1 The E^4 website	243
D.2 Downloading and installing E^4	243
D.3 Downloading the code for the examples in this book	245
Bibliography	247

List of Figures

4.1	Percent relative corrections on the original Wolf numbers: $(z_t^*/z_t - 1) \times 100$.	43
4.2	Changes in the HP trend.	44
5.1	International Airline Passengers (thousands) from 1949 to 1960. Source: see Box, Jenkins and Reinsel [30].	58
5.2	Standardized residuals of the transfer function (5.67).	60
5.3	Standardized residuals of the transfer function (5.68).	61
5.4	Monthly U.S. single-family Housing Starts and Housing Sales, in thousands of units, from January 1965 through May 1975.	61
5.5	Additive seasonal component of (log) Housing Starts (thick line) and Housing Sales (thin line), from January 1965 through May 1975.	63
5.6	Seasonality and calendar-adjusted (log) Housing Starts and Housing Sales series, from January 1965 through May 1975.	64
6.1	Fixed-interval smoothed estimates for β_t (solid line) in the time-varying parameter CAPM regression with $\pm 2\sigma$ limits (dashed line).	78
6.2	Quarterly unadjusted West German Consumption. The data is in billions of Deutsch Marks.	78
6.3	Conditional correlation between Deutsch Mark and Japanese Yen returns.	81
7.1	Data in examples 7.5.1 and 7.5.2. Top: Airline passengers time series from Box, Jenkins and Reinsel [30]. Bottom: US short-term interest rates.	96
8.1	Comparison between the model components obtained from the STSM (left) and the equivalent SEM (right).	119
8.2	Trace of the smoothed covariance matrices corresponding to the STSM and the equivalent SEM.	120
8.3	Trace of the smoother covariance of Model (8.54).	123
8.4	Common features of (log) wheat prices.	124

8.5	Monthly series of sales and advertising of the Lydia Pinkham vegetable compound from January 1954 to June 1960. Source: Palda [175].	125
8.6	Sales versus the exponential of the stochastic component for the Lydia Pinkham data. The distance between both series (gray area) is an estimate of the effect of advertising over sales.	126
8.7	Estimated value added by advertising, computed in each period as the estimate of sales generated by advertising minus the advertising investment (Lydia Pinkham data).	127
8.8	Multiplicative effect of advertising (thick line) versus multiplicative seasonality of sales (thin line) for the Lydia Pinkham data. The values between November 1957 and February 1958 are smoothed interpolations.	127
9.1	Parameters of the ARMA model (9.24) and smallest singular value (SV) of the Jacobian for several values of ϕ in the AR(1) plus error model.	139
10.1	Decomposition of the quarterly indicator. The adjusted indicator includes the trend, irregular, and level change components, and excludes seasonality and calendar affects.	161
10.2	Standardized plot of quarterly estimates of VAI (thick line) and adjusted indicator (thin line). The last values are forecasts computed from 2002.1 to 2002.4 quarters using model (10.27).	162
11.1	Observed and predicted time series of ovarian follicles for the 11th mare in dataset Ovary of Section 11.6.1. The 12-step ahead predictions (open circles) and corresponding 95% prediction bands (dashed lines), are obtained by running the Kalman recursions for the SS formulation of the LMM described for the series.	184
11.2	Kalman smoothing and prediction for beluga whale Casey in the whale.dat dataset of Subsection 11.6.2. Shown are smoothed values (solid lines) and 5-step ahead predictions (open circles) for nursing time (top panel), number of bouts (middle panel), and number of lockons (bottom panel), from period 100 onward. Corresponding 95% confidence bands for the smoothed and predicted values are shown as dotted and dashed lines, respectively.	187

List of Tables

2.1	Some common specifications for the structural time series components, see [118]. The cycle parameters are such that $0 \leq \rho < 1$ and $0 \leq \lambda_c < \pi$. The white noise errors η_t , ζ_t , κ_t , κ_t^* , and ω_t are often constrained to be independent. In the models for the seasonal component, the letter S denotes the seasonal period.	7
4.1	Estimates and standard errors for the AR(2)Autoregression(s) plus noise model fitted to the Wolf sunspot series.	43
5.1	Estimation results for the airline models (5.64) and (5.65).	59
6.1	Estimation results for model (6.77)–(6.79).	77
6.2	Estimation results for the model (6.80). The values on convergence of the (log) likelihood function and the information criteria are, respectively, $\ell^*(\hat{\theta}) = -313.766$; AIC = -5.715, SBC = -5.515	79
7.1	Definition of the weight matrices \mathbf{W}_1 and \mathbf{W}_2 in equation (7.13). Note that when the CVA approach is used, then the singular values obtained from the SVD coincide with the canonical correlation coefficients between Z_f and V_p .	87
7.2	Regular system order estimation for z_t .	97
7.3	Seasonal system order estimation for z_t with seasonal structure ($s = 12$) Block Hankel matrices.	97
7.4	Regular system order estimation for \hat{a}_{2t} .	98
7.5	Seasonal system order estimation for \hat{a}_{2t} with the seasonal Block Hankel matrix and $s = 12$.	98
7.6	Datasets used to compare the SM, SUBML and ML methods in Tables 7.7, 7.8, and 7.9.	100
7.7	Univariate models. Comparison of models estimated for the data described in Table 7.6 - Part 1. Standard errors are in parenthesis. $\hat{\sigma}_a^2$ denotes the residual variance. SM is a subspace methods estimation with weights corresponding to the CVA algorithm (see Table 7.1) and using an extended observability matrix approach (see Section 7.1).	101

7.8	Univariate models. Comparison of models estimated for the data described in Table 7.6 - Part 2. Standard errors are in parenthesis. $\hat{\sigma}_a^2$ denotes the residual variance. SUBML values were obtained with the algorithm described in Section 7.3, by pruning the non-significant parameters found in Table 7.7.	102
7.9	Univariate models. Comparison of models estimated for the data described in Table 7.6 - Part 3. Standard errors are in parenthesis. $\hat{\sigma}_a^2$ denotes the residual variance. ML estimates were initiated with the parameters obtained by SUBML shown in Table 7.8.	103
7.10	System order estimation for \mathbf{Y}_{1t} .	104
7.11	Sequence of identified identifiability models for the process \mathbf{Y}_{1t} . The figures in parentheses are the standard errors of the parameters and the symbol “-” denotes that the corresponding parameter has been constrained to its current value, so it does not have a standard error.	105
8.1	Association between state variables and structural components according to the eigenvalues of the transition matrix and their spectral properties.	115
8.2	Definition of the data generating process in STSM and ARIMA forms.	117
8.3	Structure of the block-diagonal SEM representation of the data generating process and association of the state variables to structural components.	118
8.4	Comparison between the ARIMA representations of the STSM and the block-diagonal SEM form.	118
8.5	Structure of the block-diagonal SEM equivalent to model (8.54) and association of the state variables to structural components.	122
9.1	Structural and ARIMAX forms for some common univariate specifications. Both representations are observationally equivalent.	136
9.2	Structural and VARMAX forms for some common bivariate specifications. Both representations are observationally equivalent.	137
9.3	Characterization of the identifiability of an SS model using the rank of the Jacobian (J) measuring how a change in the structural model parameters affects the parameters in the reduced form.	138
9.4	Step-by-step results from a bottom-up modeling of the US GDP series.	143

10.1 Aggregation of several univariate models. The columns labelled “states” show the number of dynamic components in the minimal SS representation of the corresponding model. Therefore, the difference between the number of states in the quarterly and annual models is the number of dynamic components that become unobservable after aggregation. The quarterly variables z_t , z_{1t} and z_{2t} are assumed to be flows, so their annual aggregate is the sum of the corresponding quarterly values.	155
10.2 Aggregation of several bivariate models. All the quarterly variables z_{1t} and z_{2t} are assumed to be flows, so their annual aggregate is the sum of the corresponding quarterly values. If a variable were to be aggregated as an average of the quarterly values, the model would require an appropriate re-scaling.	156
10.3 Original data and interpolations obtained with full and non-conformable samples. Underlined values correspond to interpolations, extrapolations or forecasts. The column “Indicator” refers to the calendar and seasonally adjusted values computed with model (10.26).	164
10.4 Ranking of RMSEs for end-of year forecasts of annual VAI from 1997 to 2001. RMSEs are computed for forecasts of both, VAI levels and growth rates. The columns RMSE% show the corresponding RMSEs normalized so that the RMSE of Model (10.29) is 100%.	165
11.1 Final parameter estimates for the SS model fitted to the beluga whale data of Subsection 11.6.2. Estimates were obtained via the EM algorithm initialized with the values in square brackets. The final estimate of the initial state mean was $\hat{\mu}^T = [3.296, 1.009, 1.390, -8.848 \times 10^{-3}, 6.653 \times 10^{-1}]$, starting from the zero vector.	186
11.2 Five-step ahead predictions and 95% confidence bounds for Casey in the beluga whale data of Subsection 11.6.2.	188
C.1 Standard statistics procedures to specify and simulate a time series model.	234
C.2 Signal extraction functions. These are the main procedures offered for time series disaggregation, structural decomposition, smoothing and forecasting.	236
C.3 Model estimation functions. These are the main procedures which can be applied to compute the Gaussian likelihood, its first-order derivatives, and information matrix. These functions can be fed to the standard E^4 optimizer (<code>e4min</code>) to compute ML estimates. Initial estimates can be computed using a specialized function (<code>e4preest</code>).	239

This page intentionally left blank

Preface

The story of how any book comes into being is typically complex. Even moreso in the case of this one...

Some of us started this project (at least in our minds) during our undergraduate studies, when we discovered the awesome Box and Jenkins [29, 30] book on time series analysis. This text, together with the influence of some excellent instructors, was key to our decision to pursue an academic career in Statistics/Econometrics and seeded an early ambition to write a book on this matter.

Some or many years later, depending on each particular case, we also discovered the elegance and clarity of the State-Space approach, which was a revelation comparable to the previous one.

In the 1980s, State-Space methods were applied almost exclusively to problems in engineering but some excellent books, such as those by Jazwinsky [130] or Anderson and Moore [6], sprouted spontaneously on the shelves of many Departments of Mathematics and Economics, frequently with some volumes of the dominant journals on systems engineering. This progressive input from the engineering field led to the increased interest of time series analysts to adopt these methods. As a consequence, some important books on the State-Space approach were written by authors with a mainstream time series background, such as Hannan and Deistler [116], Harvey [118] or West and Harrison [224, 225]. Also, it has become customary to include a chapter on State-Space models in time series textbooks such as those by Hamilton [115], Wei [221], or Lutkepohl [158].

State-Space methods also appear frequently in the foundations of some well-known statistics and econometrics packages. This is the case of the STATESPACE, UCM, KALDFS or RANTWO procedures in SAS/ETS and SAS/IML, see [193], or the default likelihood evaluation method for ARIMA models implemented in GRETL, see [13]. In R [186], the popular free software environment for statistical computing and graphics, the main State-Space modeling packages are dlm, dse, sspir, KFAS, and FKF, but there is a proliferation of many others; see the review articles by Petris and Petrone [182] and Tusell [213]. Last, an important software package specifically devoted to State-Space modeling is SsfPack, a library of routines written in C and linked to Ox, see Koopman, Shephard and Doornik [141] and Pelagatti [179].

Returning to our personal history, some of us discovered this literature in the 1980s and soon learned that the State-Space model was a simple and comprehensive formulation that could accommodate any of the standard linear processes that we used in our work.

This idea was awesome. Bear in mind that in those years, time series software was scarce and underpowered, so we had to work very hard at coding our own procedures. Moreover, the software that we developed for an ARIMA process required painful adaptations before it could be applied to other formulations, such as transfer functions or VARMAX. After doing so, the different versions of the basic algorithm would require their specific documentation, debugging, and updating cycles.

In comparison with this messy situation, the State-Space approach offered us a formal framework where any result or procedure developed for the basic model could be seamlessly applied to any standard formulation written in State-Space form. In short: you only needed a single version of each basic procedure, no matter if it was related to likelihood computation, interpolation, or extrapolation. Moreover, you could accommodate with a reasonable effort nonstandard situations such as observation errors, aggregation constraints or missing in-sample values, also taking advantage of the excellent “off-the-shelf” numerical procedures and code that our colleagues in the systems engineering field had been developing and sharing over the years.

These compelling advantages convinced us that the State-Space approach would be productive for our academic work. The first results exceeded our expectations and, after many articles and PhD Theses (including some of our own!) we want to synthesize and present these advantages to a wider audience by means of a book including theory, examples, and software.

Therefore, this monograph is the result of an extended process of maturation and discovery. It is addressed to postgraduate students, professional statisticians and academics. Our goal is to share with its readers the reasons why we are so enthusiastic about the State-Space approach. Because our potential audience is relatively wide, most of the concepts and methods presented are explained on an intuitive basis. Despite our effort to favor intuition we recognize that the subject discussed here is inherently difficult, so grasping 100% of the content will require an advanced level in linear algebra, dynamical systems, and time series analysis.

Concerning the scope of the book, we opted to concentrate on discussing the computational processes that can be applied to a specified linear model that is being entertained. This specific focus lead to omitting some substantive topics which would deserve specialized monographs in their own right. In particular, we do not cover the nonlinear/non-Gaussian case. Also, we barely discuss how one might arrive at an initial model specification, except in Chapter 7. Last, we do not provide a systematic coverage of the structural time series modeling approach, initially due to Harvey [118] and later developed by many impressive works, such as those of Koopman [139], Koopman and Shephard [140] or Harvey, Ruiz and Shephard [119]¹.

On the other hand, we do provide a broad presentation of what can be done after a tentative specification has been determined. In particular, we discuss two classical topics in the State-Space literature: model estimation and signal extraction. Besides, we pay special attention to the manipulation of the model and, accordingly, present many procedures to combine, decompose, aggregate and disaggregate a State-Space

¹For a comprehensive overview of the foundations of this methodology, the reader should check the books by Commandeur and Koopman [55] and Durbin and Koopman [82].

form. Finally, we also cover the connection between mainstream time series models and the State-Space representation. We do so by showing how one can obtain the State-Space formulation of a general linear process (see Subsection 2.2.4) and, conversely, provide procedures to transform a State-Space model into an equivalent VARMAX representation (see Chapter 9).

An important feature of this book is that it comes with a substantial software package.

As we explained above, our original interest in State-Space methods had much to do with developing flexible and powerful software tools for time series analysis. Our first attempts were coded in standard high-level languages such as FORTRAN or Pascal. They were a huge advance in comparison with the previous situation, but software development and maintenance was hard and time-consuming. Also, it was difficult to share this code with other researchers. For these reasons, in the early 1990s we launched the E^4 project, which was specifically devised to solve these issues.

The name E^4 refers to the Spanish: “Estimación de modelos Econométricos en Espacio de los Estados” (State-Space Estimation of Econometric Models). Under this name, we started developing a MATLAB Toolbox which collects all the computational procedures resulting from our research, together with the administrative and analytical functions required to conform a fully operational time series analysis package. This toolbox is freely distributed under the terms of the GNU General Public License through the website www.ucm.es/e-4/ which, besides source code, provides a complete user manual and other reference materials.

The connection between E^4 and this book is very clear: all the procedures presented here are implemented as E^4 functions, so the reader will be able to apply them to his/her own work. To facilitate this process, appendices C and D provide additional information about E^4 and all the information needed to download it and to replicate all the practical examples in this book.

Additional material is available at www.crcpress.com/9781482219593 and the aforementioned E^4 website, www.ucm.es/e-4/.

We do not wish to end this Preface without thanking our editor, Rob Calver, for planting the initial “seed” in our minds by means of a timely e-mail in late 2012; a message which hung in the balance between “deletion” and “response” for several weeks before succumbing to the latter. We are especially grateful to his editorial assistants, Sarah Gelson, Sarfraz Khan, and Alexander Edwards, for their responsive and knowledgeable help in all the phases of this project. Their continual requests for progress reports in particular, helped us avoid delving into the classical vices of any author: unproductive procrastination and frolicking. Last, we are deeply indebted to Laurie Oknowsky, Michael W. Davidson and the Taylor & Francis production team for their excellent work in the final phases of this project. Shashi Kumar provided fast and flawless support to solve our \LaTeX issues.

The first four authors gratefully acknowledge financial support from *Ministerio de Economía y Competitividad*, through Grant ECO2011-23972. The last author wishes to thank Texas Tech University for teaching-load reductions in the final two

years leading up to publication. Finally, four anonymous reviewers provided constructive criticism and suggestions which improved this book in many ways.

Somosaguas, Madrid, Spain

Lubbock, Texas, USA

October 2015

About the Authors

Jose Casals is the Head of Global Risk Management at Bankia. He also works part-time as an Associate Professor of Econometrics at Universidad Complutense de Madrid.

Alfredo Garcia-Hiernaux is an Associate Professor of Econometrics at Universidad Complutense and freelance consultant.

Miguel Jerez is an Associate Professor of Econometrics at Universidad Complutense and a freelance consultant. For six years he was Executive Vice-President at Caja de Madrid.

Sonia Sotoca is an Associate Professor of Econometrics at Universidad Complutense.

These authors are engaged in a long-term research project to apply State-Space techniques to standard econometric problems. Their common interests include State-Space methods and time series econometrics. A list of their publications on these topics can be found at: www.ucm.es/grupoee/publicaciones

Many algorithms derived from their State-Space research are implemented in *E*⁴ (www.ucm.es/info/icae/e4) a MATLAB toolbox freely offered under the terms of the GNU General Public License. They can be reached at the same address: Departamento de Fundamentos del Análisis Económico II, Universidad Complutense de Madrid, Campus de Somosaguas s/n, 28223 Madrid (SPAIN).

A. Alexandre (Alex) Trindade is Professor of Statistics in the Department of Mathematics & Statistics at Texas Tech University, and Adjunct Professor in the Graduate School of Biomedical Sciences, Texas Tech University Health Sciences Center. His research spans a broad swath of theoretical and computational statistics. Prior to Texas Tech, he held appointments at the University of Florida, Colorado State University, the University of New Mexico, and the University of Oklahoma.

Alex can be reached at the Department of Mathematics & Statistics, Texas Tech University, Broadway and Boston, Lubbock, TX 79409-1042, USA.

This page intentionally left blank

Chapter 1

Introduction

This book provides a comprehensive approach to the many computational procedures that can be applied to a previously specified linear model in State-Space (hereafter, SS) form. Our interest in this field is motivated by the important advantages of the SS approach. Some of them are:

1. Any linear time series model can be written in an equivalent SS form¹. In particular, ARIMAX, VARMAX or transfer functions can be represented in this way, see Subsection 2.2.4. In other cases, such as that of structural time series models (see Harvey [118]) the model is directly defined in SS form, so no translation is required.
2. Conversely, any linear SS model can be written in an equivalent VARMAX form, see Chapter 9.
3. The flexibility of the SS representation allows one to accommodate nonstandard situations like time-varying parameters (Pagan [173] and Watson and Engle [219]), unobserved variables (Harvey [118]), observation errors (Terceiro [206]), aggregation constraints (Casals, Sotoca and Jerez [46]) or missing in-sample values (Ansley and Kohn [7], Kohn and Ansley [137] and Naranjo, Trindade and Casella [165]).
4. System engineering provides a wealth of excellent “off-the-shelf” procedures and ideas to solve many relevant problems such as the specification, estimation and optimal control of SS models.

These ideas have deep implications. Consider for example points 1 and 2 above. Taken together they imply that the specific form of a model is a matter of choice as: (a) one can work with a SS representation and finally translate it to a VARMAX form, if this is convenient, or (b) build on a VARMAX model and run chosen parts of the analysis using the equivalent SS model.

¹Throughout this book, we will consider that two models are equivalent if they result in the same probability distribution of the sample. This concept is also known as “observational equivalence” in the literature.

The ability to translate a VARMAX model to the SS form and vice versa also implies that the flexibility and power of the SS procedures, highlighted in points 3 and 4 above, can be immediately put into practice in mainstream time series analysis. Specifically, any identification, estimation, decomposition or aggregation method devised for the linear SS model can be applied to mainstream time series specifications, such as VAR, VARMAX, structural econometric models or transfer functions.

Finally, a model is a mathematical representation of a system. A procedure is a tool to manipulate a model. Models and procedures are different things and, therefore, a desirable feature for any procedure is to be independent of particular model formulations. Accordingly, it makes sense to concentrate the development of computational procedures on the SS model, using the generality of this formulation as an “abstraction layer” to separate them from specific model families. In our experience, this brings flexibility, reliability and consistency to statistical software.

The structure of this book is as follows. Chapter 2 discusses the formulation of the SS model. To this end, we define the general form of the multiple-error linear SS model and describe how one can obtain equivalent formulations by applying a similar transformation to the state variables. We also present the single-error SS model, which is an alternative formulation with clear advantages when one wants to estimate the state variables. The last section explains how to obtain the SS model matrices corresponding to a generalized VARMAX process.

Chapter 3 illustrates the flexibility of the SS representation by presenting three classes of SS model transformations: decomposition, combination, and change of output variables. Decomposing an SS model consists in breaking it into several meaningful sub-models, each one with its own SS representation. Some examples are: (a) the deterministic-stochastic decomposition, which breaks the model into additive components affected either by the inputs or the errors of the system, (b) the decomposition of an SS model into the univariate models implied by its structure and (c) the block-diagonal SS form, which breaks the model structure into trend, seasonal, cyclical, and irregular components.

On the other hand, we use the term “model combination” or “model nesting” to describe the construction of a complex SS representation by combining several simpler SS models. The link between the different sub-models may be given either by their inputs (“combination in inputs” or “endogeneization”) or their errors (“combination in errors”). Building on these procedures one can tackle nontrivial problems such as defining a complex model as the product of several multiplicative factors or linking the model for the system output with those of the inputs into a unique representation where all the variables (inputs and outputs) are endogenous.

Finally, the change of output variables in the SS model consists of replacing the endogenous variable with another one which takes into account nonstandard data features such as, e.g., observation errors, missing values, or aggregated data.

Chapter 4 covers the main state estimation algorithms considered in the SS literature: filtering and smoothing. Both focus on estimating the sequence of state variables with different information sets. In particular, filtering algorithms provide the expected value and error covariance of the states at time $t + 1$, conditional on all the

sample information available up to time t^2 . On the other hand, smoothing procedures provide the conditional first and second-order moments of the states with flexible information sets. Anderson and Moore [6] distinguish between fixed-point, fixed-lag and fixed-interval smoothing. For the purposes of this book we will concentrate on the latter, which provides the first and second-order moments of the states conditional on all the sample information.

Filtering and fixed-interval smoothing procedures are applied in this book with many purposes in mind. In particular, we use the Kalman filter to compute the innovations required to evaluate the likelihood for a SS model, see Terceiro [206], while fixed-interval smoothing methods are applied to estimate unobserved components in structural time series models, and to deal with samples containing missing values or aggregated data. The last Section in Chapter 4 includes some examples of these applications.

Chapters 5 and 6 describe in detail how one can compute the Gaussian likelihood for the unknown coefficients in the SS matrices of a given model. In particular, Chapter 5 concentrates on models with fixed coefficients, while Chapter 6 covers the case where some coefficients or covariances are time-varying, including periodic models, regression equations with time-varying coefficients or models with GARCH errors. Combining these algorithms with a standard iterative optimization procedure allows one to obtain maximum-likelihood (ML) estimates for any unknown parameter in the SS model matrices. Likelihood computation also provides a basis from which to approximate the information matrix, see Terceiro [206], and for inference based on the likelihood ratio, see Casella and Berger [48].

Chapter 7 provides an introduction to the so-called “subspace methods.” This term refers to a dominant trend in modern system identification, see e.g. Ljung [152] or Qin [185]. In the context of this book, we focus on their application to compute rank-deficient Generalized Least Squares (hereafter, GLS) estimates for the model parameters, see Ljung [152]. The fast and stable estimates provided by these procedures can be used either as an alternative to ML, or as preliminary estimates, useful to initiate an iterative estimation procedure.

Chapter 8 deals with a classical subject in the SS literature, generally described as “signal extraction.” This term refers to the estimation of signals buried in a vector of time series. The time series literature often reduces it to breaking a single variable into the sum of its “structural components,” which are the trend, cycle, seasonal, and irregular components. In this chapter we present our approach to this matter which has some particular features, as it focuses on the case of multiple time series, allows for exogenous variables, and produces “revision-free” components, that is, estimates which are invariant to changes in the sample.

As is well known, any linear time series model can be written in SS form, but a not so well known fact is that this is a two-way trip, as any linear SS model can also be written in VARMAX form, see Casals, Garcia-Hieraux, and Jerez [39]. Chapter 9 describes two algorithms used to obtain the VARMAX matrices corresponding to

²In books using a slightly different SS formulation the filter provides estimates of the contemporary states and covariances, i.e. the first and second-order moment at time t conditional on the information available up to time t . We will further discuss this issue in Chapter 4.

any linear SS model and showcases some applications in three areas: analyzing the identifiability of the parameters in a SS model, specifying a model from noisy data, and estimating the parameters of a structural SS specification so that it realizes a given VARMAX reduced form.

Chapter 10 deals with several issues relating to the aggregation and disaggregation of time series and, in particular, discusses how to specify an observable high-frequency model for a vector of time series sampled at high and low frequencies. To this end, we first study how aggregation over time affects both the dynamic components of a time series and their observability. The main conclusion is that aggregation does not affect the state variables driving the system but some of them, mainly those related to the seasonal structure, become unobservable. Building on these results, one can devise a structured specification process built on the idea that the models relating the variables in high and low sampling frequencies should be mutually consistent. After specifying such a model, the methods discussed in previous chapters provide an adequate framework for estimation, diagnostic checking, data interpolation and forecasting. An example with national accounting data illustrates the practical application of this methodology.

Finally, Chapter 11 discusses a cross-sectional extension to the classical SS formulation, in order to accommodate longitudinal or panel data. This is important if one wants to globally analyze several time series, each one originating from a different “unit.” Common examples are biological variables for different subjects, or econometric variables for different firms. The assumption of independence among the units allows one to easily combine the separate likelihoods. Appropriate smoothing, filtering, and prediction recursions in this context will be discussed, as well as maximum likelihood estimates and the quantification of their uncertainty through asymptotic results. A particularly attractive and powerful feature of this methodology is the inherent ability to deal with missing or irregularly-spaced time series data, together with the ease with which smoothing and forecasting is accomplished.

The appendices provide some supplementary material. In particular: Appendix A discusses some specialized tools in numerical algebra and linear systems. Appendix B presents the asymptotic theory underlying the ML estimates for simplified versions of the SS model. Appendix C describes E^4 , a free MATLAB Toolbox for time series analysis using the SS methods discussed in this book. Finally, Appendix D provides detailed instructions about how to obtain the code and data required to replicate the practical examples in this book with E^4 .

Chapter 2

Linear State-Space models

2.1	The multiple error model	6
2.1.1	Model formulation	6
2.1.2	Similar transformations	8
2.1.3	Properties of the State-Space model	8
2.2	Single error models	12
2.2.1	Model formulation	12
2.2.2	State estimation in the SEM	12
2.2.3	Obtaining the SEM equivalent to a MEM	13
2.2.4	Obtaining the SEM equivalent to a general linear process	15

This chapter describes the two basic SS formulations that will be used in the remaining chapters. We will refer to the first one as a “multiple error model,” or “MEM,” because it allows for different errors in the state and observation equations of the model. On the other hand, the “single error model,” or “SEM” has a unique vector of random disturbances, affecting both equations.

The relationship between both formulations is very rich, and can be summarized in two ideas. First, under nonrestrictive assumptions a MEM model can be written in an observationally equivalent SEM form. Second, the SEM has specific advantages for applications requiring precise estimates of the state variables. The combination of both ideas allows one to devise powerful computational procedures, as we will see e.g. in Chapters 5, 6 or 8.

The main properties of the SS model are presented using the MEM representation, bearing in mind that their extension to the SEM case is trivial. Last, we show how to write a standard linear stochastic process in SEM form.

2.1 The multiple error model

2.1.1 Model formulation

The basic fixed-coefficients SS formulation is given by a *state equation*¹, characterizing the system dynamics:

$$\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{E} \mathbf{w}_t, \quad (2.1)$$

and an *observation equation*, describing how the system output is realized as a linear combination of different dynamic components:

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_t + \mathbf{D} \mathbf{u}_t + \mathbf{C} \mathbf{v}_t, \quad (2.2)$$

where $\mathbf{z}_t \in \mathbb{R}^m$ is a random vector of *endogenous variables* or *outputs* at time t , $\mathbf{x}_t \in \mathbb{R}^n$ is vector of *state variables* or simply *states*, $\mathbf{u}_t \in \mathbb{R}^k$ is a vector of *exogenous variables* or *inputs*, and \mathbf{w}_t and \mathbf{v}_t are conformable vectors of zero-mean white noise errors, such that:

$$\mathbb{E} \left(\begin{bmatrix} \mathbf{w}_t \\ \mathbf{v}_t \end{bmatrix} \begin{bmatrix} \mathbf{w}_t & \mathbf{v}_t \end{bmatrix} \right) = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix}. \quad (2.3)$$

Last, Φ , Γ , \mathbf{E} , \mathbf{H} , \mathbf{D} , \mathbf{C} , \mathbf{Q} , \mathbf{S} and \mathbf{R} are time-invariant coefficient matrices and the covariance matrix defined in (2.3) is positive semi-definite.

Traditionally, the parameter and covariance matrices in the SS model depend on time. In (2.1)–(2.3) we dropped the time subindex. We did this to simplify the notation, because most of the theory and applications considered in this book are covered by the time-invariant framework. Occasionally we will consider some types of parameter variation, such as e.g. Example 2.1.2, Subsection 3.3.2 or Chapter 6. Whenever suitable, the text will make clear that we are using a time-varying SS model.

We will also assume that all the errors in model (2.1)–(2.2) are independent of the exogenous variables \mathbf{u}_t and the initial state vector \mathbf{x}_1 ². Last, we will consider that the initial state, \mathbf{x}_1 , is a random vector such that its mean and covariance, conditional to the input values, are μ_1 and P_1 , respectively.

Example 2.1.1 Structural time series models

The SS representation can be used to formulate time series models directly. For example, Harvey [118] proposes a wide family of *structural time series models*, building on the following decomposition for \mathbf{z}_t :

$$\mathbf{z}_t = \mu_t + \psi_t + \gamma_t + \varepsilon_t, \quad (2.4)$$

¹There are other ways to define the state equation. For example, two common alternatives are: $\mathbf{x}_t = \Phi \mathbf{x}_{t-1} + \Gamma \mathbf{u}_t + \mathbf{E} \mathbf{w}_t$ and $\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \Gamma \mathbf{u}_{t+1} + \mathbf{E} \mathbf{w}_{t+1}$. These specific representations are as general as the one we use here because they can be derived from Equation (2.1) by means of a suitable change of variables.

²This assumption is not restrictive because, if there were any linear relationship between the model inputs and errors, we could endogeneize the former as shown in Section 3.2.

Table 2.1 Some common specifications for the structural time series components, see [118]. The cycle parameters are such that $0 \leq \rho < 1$ and $0 \leq \lambda_c < \pi$. The white noise errors η_t , ζ_t , κ_t , κ_t^* , and ω_t are often constrained to be independent. In the models for the seasonal component, the letter S denotes the seasonal period.

Structural component	Model	Specification				
Trend (μ_t)	Random walk	$\mu_{t+1} = \mu_t + \eta_t$				
	Stochastic trend	$\begin{bmatrix} \mu_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} \eta_t \\ \zeta_t \end{bmatrix}$				
	Random walk with drift	$\begin{bmatrix} \mu_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} \eta_t \\ 0 \end{bmatrix}$				
	Integrated random walk	$\begin{bmatrix} \mu_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} 0 \\ \zeta_t \end{bmatrix}$				
Cycle (ψ_t)	Stochastic cycle	$\begin{bmatrix} \psi_{t+1} \\ \psi_{t+1}^* \end{bmatrix} = \rho \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}$				
	Nonstationary cycle	$\begin{bmatrix} \psi_{t+1} \\ \psi_{t+1}^* \end{bmatrix} = \begin{bmatrix} \cos \lambda_c & \sin \lambda_c \\ -\sin \lambda_c & \cos \lambda_c \end{bmatrix} \begin{bmatrix} \psi_t \\ \psi_t^* \end{bmatrix} + \begin{bmatrix} \kappa_t \\ \kappa_t^* \end{bmatrix}$				
Seasonality (γ_t)	Dummy variable seasonality	$\gamma_{t+1} = \sum_{j=0}^{S-2} \gamma_{-j} + \omega_t$				
	Trigonometric seasonality	$\gamma_{t+1} = \sum_{j=1}^{[S/2]} (\alpha_j \cos \lambda_j t + \beta_j \sin \lambda_j t) + \omega_t; \quad \lambda_j = 2\pi j/S$				

where μ_t , ψ_t , γ_t , and ε_t are, respectively, the trend, cycle, seasonal, and irregular components of the time series. Each of these *structural components* is specified by means of a state equation. Some common specifications used for this purpose are summarized in Table 2.1. Harvey [118] and related subsequent literature extend this basic model to accommodate exogenous variables, multiple time series, and cointegration constraints.

Example 2.1.2 Linear regression models

The SS representation is so flexible that it can be used either to formulate models, as in the previous example, or to accommodate as particular cases most mainstream specifications, allowing in this case for interesting variations of the basic model. Consider for example the multiple regression model:

$$z_t = \mathbf{u}_t^T \boldsymbol{\beta} + v_t, \quad (2.5)$$

which could be written as a particular case of (2.1)–(2.2) just by setting $\mathbf{H} = \mathbf{0}$. Alternatively, it could be formulated as:

$$\begin{aligned} \boldsymbol{\beta}_{t+1} &= \boldsymbol{\beta}_t \\ z_t &= \mathbf{u}_t^T \boldsymbol{\beta}_t + v_t, \end{aligned} \quad (2.6)$$

where the state equation characterizes a fixed parameter, whose value at time $t + 1$ is equal to its value at time t . An immediate extension of model (2.6) would allow for a time-varying parameter. For example, we could define a regression model with random walk parameters:

$$\begin{aligned}\beta_{t+1} &= \beta_t + w_t \\ z_t &= u_t^T \beta_t + v_t,\end{aligned}\tag{2.7}$$

with $\text{cov}(w_t, v_t) = 0$. Building on equation (2.7), one can immediately define a wide family of time-varying regression formulations, see Chapter 6 and the references therein.

2.1.2 Similar transformations

There are infinite numbers of SS models realizing the same output. To see this, consider the linear transformation of the state variables $\bar{x}_t = T x_t$, where T is any arbitrary matrix such that $|T| \neq 0$, and $|\cdot|$ denotes the matrix determinant. It is therefore possible to compute the inverse transformation:

$$x_t = T^{-1} \bar{x}_t.\tag{2.8}$$

Substituting (2.8) in (2.1)–(2.2) we obtain:

$$T^{-1} \bar{x}_{t+1} = \Phi T^{-1} \bar{x}_t + \Gamma u_t + E w_t\tag{2.9}$$

$$z_t = H T^{-1} \bar{x}_t + D u_t + C v_t,\tag{2.10}$$

which immediately yields an SS model equivalent to (2.1)–(2.2):

$$\bar{x}_{t+1} = T \Phi T^{-1} \bar{x}_t + T \Gamma u_t + T E w_t\tag{2.11}$$

$$z_t = H T^{-1} \bar{x}_t + D u_t + C v_t.\tag{2.12}$$

This change of variables in the SS model is known as a *similar transformation* because the transition matrices in (2.1)–(2.2) and (2.11)–(2.12) are similar to each other. By selecting an adequate similar transformation, one can obtain specific forms of the SS model with convenient properties, see e.g. Subsection 3.1.3.

2.1.3 Properties of the State-Space model

We will now define some properties of the SS model (2.1)–(2.2) that will be used throughout the rest of the book. Some of them rely on the basic concept of the *system mode*.

A *system mode* (henceforth *mode*) is an eigenvector of the transition matrix Φ . By definition, it could be interpreted as a value of the state vector which would not be altered (except for a multiplicative constant) upon propagation from t to $t + 1$. The multiplicative constant is the corresponding eigenvalue, which therefore acts as an attenuation or amplification factor of the mode.

For example, consider a situation where the state transition matrix is given by

$$\Phi = \begin{bmatrix} 1.0 & 1.0 & 0 & 0 \\ 0 & 1.0 & 0 & 0 \\ 0 & 0 & 0.7 & 0 \\ 0 & 0 & 0 & 0.5 \end{bmatrix},$$

with the eigenvalue-eigenvector pairs $\{\lambda_i, \nu_i\}$, where $\lambda_1 = 1.0 = \lambda_2$ (the eigenvalue 1.0 has multiplicity 2), $\lambda_3 = 0.7$, and $\lambda_4 = 0.5$. The corresponding eigenvectors are

$$\nu_1 = \begin{bmatrix} 1.0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \nu_2 = \begin{bmatrix} -1.0 \\ 0 \\ 0 \\ 0 \end{bmatrix}, \quad \nu_3 = \begin{bmatrix} 0 \\ 0 \\ 1.0 \\ 0 \end{bmatrix}, \quad \nu_4 = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1.0 \end{bmatrix}.$$

This system would have three modes (which coincides with the number of independent eigenvectors, or geometric multiplicity of the transition matrix), two of them stationary, driven by λ_3 and λ_4 , and one of them nonstationary, driven by λ_1 along the direction specified by ν_1 . Thus: the first mode is ν_1 , the second mode is ν_3 , and the third mode is ν_4 .

With this in place, we are now ready to define the following fundamental properties.

Definition 2.1 (Stability). The SS model is *stable* if all the roots of the characteristic equation $|\mathbf{I} - \Phi \lambda| = 0$ lie outside the unit circle.

Note that: (a) the condition above holds if and only if all the eigenvalues of the transition matrix Φ lie inside the unit circle, and (b) stability implies that, if the system inputs and errors are set to constant values, the state vector will converge to a steady-state equilibrium.

Definition 2.2 (Reachability / Controllability). The SS model is *reachable* if choosing adequately the value of the system inputs, \mathbf{u}_t , one can reach any arbitrary value for the state vector, starting from a null initial condition $\mathbf{x}_1 = \mathbf{0}$.

This property is also known as *controllability from the origin*. Hereafter, we will call it *controllability*, without further specification. Controllability is also linked to a dual property known as *observability*.

Definition 2.3 (Observability). The SS model is *observable* if, given the information about the inputs and outputs, one can uniquely determine the initial state.

The concepts of controllability and observability can be formalized as follows. First (and recalling that n is the dimension of the state vector) it is easy to see that,

when $t = n$, the state vector can be written as:

$$\mathbf{x}_n = \Phi^n \mathbf{x}_1 + [\Phi^{n-1}(\Gamma E) \quad \Phi^{n-2}(\Gamma E) \quad \dots \quad \Phi(\Gamma E) \quad \Gamma E] \begin{bmatrix} u_1 \\ w_1 \\ u_2 \\ w_2 \\ \vdots \\ u_{n-2} \\ w_{n-2} \\ u_{n-1} \\ w_{n-1} \end{bmatrix} \quad (2.13)$$

and, from (2.13), a necessary and sufficient condition for controllability is $\text{rank}(\mathbf{C}_n) = n$, where \mathbf{C}_n is the controllability matrix:

$$\mathbf{C}_n = [\Phi^{n-1}(\Gamma E Q^{1/2}) \quad \Phi^{n-2}(\Gamma E Q^{1/2}) \quad \dots \quad \Phi(\Gamma E Q^{1/2}) \quad \Gamma E Q^{1/2}] \quad (2.14)$$

so, if the rank of \mathbf{C}_n coincides with the number of states, the image set of \mathbf{C}_n is \mathbb{R}^n and one can determine the value of \mathbf{x}_n by choosing the sequences $\{u_t, w_t\}_{t=1}^n$. Similarly, it is easy to see that the observations up to $t = n$ can be written as:

$$\begin{aligned} \begin{bmatrix} z_1 \\ z_2 \\ \vdots \\ z_n \end{bmatrix} &= \begin{bmatrix} H \\ H\Phi \\ \vdots \\ H\Phi^{n-1} \end{bmatrix} \mathbf{x}_1 + \begin{bmatrix} D & 0 & \dots & 0 \\ H\Gamma & D & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ H\Phi^{n-2}\Gamma & H\Phi^{n-3}\Gamma & \dots & D \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix} \\ &+ \begin{bmatrix} 0 & 0 & \dots & 0 \\ HE & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ H\Phi^{n-2}E & H\Phi^{n-3}E & \dots & 0 \end{bmatrix} \begin{bmatrix} w_1 \\ w_2 \\ \vdots \\ w_n \end{bmatrix} + \begin{bmatrix} C & 0 & \dots & 0 \\ 0 & C & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & C \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ \vdots \\ v_n \end{bmatrix} \end{aligned} \quad (2.15)$$

Therefore, a necessary and sufficient condition for observability is that $\text{rank}(\mathbf{O}_n) = n$, where \mathbf{O}_n is the observability matrix:

$$\mathbf{O}_n = [H^T \quad (H\Phi)^T \dots \quad (H\Phi^{n-1})^T]^T. \quad (2.16)$$

Definition 2.4 (Minimality). The SS model is *minimal* if there is no alternative representation that realizes the same output, \mathbf{z}_t , with fewer states.

It can be proved that a model is minimal if and only if it is observable and controllable. Throughout this book we will assume that the SS models employed are minimal. This assumption is made without loss of generality because any given system can be transformed into an equivalent minimal representation by deleting its unobservable/non-controllable modes using the so-called *staircase algorithm*, see Rosenbrock [190].

Definition 2.5 (Canonical representation). An SS model is said to be in *canonical representation* if its matrices have a particular structure such that there is no similar transformation matrix, other than identity or trivial permutation matrices, that preserves this structure.

Canonical models assure identifiability and reduce the dimension of the parametric space, as they have the minimal number of parameters required to realize the output, z_t . A well known canonical SS representation is the *Luenberger Canonical Form*, which can be obtained from any SS model by means of a similar transformation, see Petkov, Christov and Konstantinov [181]. Chapter 9 provides further details about this form and its usefulness in deriving the VARMAX representation of an SS model.

Definition 2.6 (Detectability). The SS model is *detectable* if any unobservable mode is stable.

If a given SS model has some unobservable modes, there always exists a similar transformation that allows one to separate the observable and unobservable modes. That is:

$$\Phi^{**} = \begin{bmatrix} \Phi_{11}^{**} & \mathbf{0} \\ \Phi_{21}^{**} & \Phi_{22}^{**} \end{bmatrix} = \mathbf{T} \Phi \mathbf{T}^{-1}; \quad \mathbf{H}^{**} = \begin{bmatrix} \mathbf{H}_1^{**} & \mathbf{0} \end{bmatrix} = \mathbf{H} \mathbf{T}^{-1} \quad (2.17)$$

where the pair $[\mathbf{H}_1^{**} \quad \Phi_{11}^{**}]$ is observable and \mathbf{T} is the matrix characterizing the similar transformation. If the system is detectable, all the eigenvalues of Φ_{22}^{**} would then lie within the unit circle. Therefore, any unobservable states are assured to be stable.

This transformation allows us to understand the concept of observability because only the states associated to Φ_{11}^{**} affect the system output, so we cannot estimate the unobservable states conditioned to the observed variables.

Similarly, if an SS model has some non-controllable modes, there exists a similar transformation such that:

$$\Phi^{**} = \begin{bmatrix} \Phi_{11}^{**} & \mathbf{0} \\ \Phi_{21}^{**} & \Phi_{22}^{**} \end{bmatrix}; \quad \Gamma^{**} = \begin{bmatrix} \mathbf{0} \\ \Gamma_2^{**} \end{bmatrix}; \quad \mathbf{E} \mathbf{Q}^{1/2} = \begin{bmatrix} \mathbf{0} \\ \mathbf{E}_2^{**} \end{bmatrix} \quad (2.18)$$

where the non-controllable modes are associated to Φ_{11}^{**} and are not affected by any source of excitation. As a consequence, the non-controllable subsystem affects the system as a deterministic function of the initial state. If the eigenvalues of Φ_{11}^{**} are less than one, we can cancel this effect by assuming that the initial state of the non-controllable subsystem is zero, see the immemorial time assumption in Definition 2.8 below.

Definition 2.7 (Stabilizability). The SS model is *stabilizable* if any non-controllable mode is stable.

Definition 2.8 (Immemorial time assumption): The SS model fulfills the immemorial time assumption if: (a) the state transition equation holds for $t = 0, -1, -2, \dots$ and (b) $\mathbf{x}_r = \mathbf{0}$ with $r \rightarrow -\infty$.

Definition 2.9 (Weak stationarity): The SS model is weakly stationary (hereafter, stationary) if the first and second-order moments of its output, z_t , are time-invariant.

The following results describe the relationship between the concepts of stationarity and stability.

Result 2.1: Given a stable system without inputs, initialized with $E(\mathbf{x}_1)$ and $\text{cov}(\mathbf{x}_1)$, the process $\{\mathbf{x}_t\}$ is asymptotically weakly stationary with $\lim_{t \rightarrow \infty} E(\mathbf{x}_t) = \mathbf{0}$ and $\lim_{t \rightarrow \infty} \text{cov}(\mathbf{x}_t) = \mathbf{P}$, where the covariance \mathbf{P} satisfies the Lyapunov algebraic equation $\mathbf{P} = \Phi \mathbf{P} \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T$.

Result 2.2: If a stable system is initialized with $E(\mathbf{x}_1) = \mathbf{0}$ and $\mathbf{P}_1 = \Phi \mathbf{P}_1 \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T$, then the process $\{\mathbf{x}_t\}$ is stationary.

Result 2.3: If a stable system without inputs was started in immemorial time, then the processes $\{\mathbf{x}_t\}$ and $\{\mathbf{z}_t\}$ are stationary within the time span of the sample.

Therefore, the concepts of stationarity and stability are different *stricto sensu*, but closely related if the system has no inputs³. For a formal proof of these results see Anderson and Moore [6].

2.2 Single error models

2.2.1 Model formulation

The SS representation defined in Section 2.1 is a *multiple-error model*, or MEM, because it has different errors affecting the state and observation equations. An alternative is a representation with a single error affecting both equations. This *single-error model*, or SEM, is given by:

$$\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{K} \mathbf{a}_t \quad (2.19)$$

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_t + \mathbf{D} \mathbf{u}_t + \mathbf{a}_t \quad (2.20)$$

where the error term is such that $\mathbf{a}_t \sim \text{IID}(\mathbf{0}, \mathbf{B})$, that is, \mathbf{a}_t is a sequence of *independent and identically distributed* (IID) zero-mean random vectors with common covariance matrix \mathbf{B} .

2.2.2 State estimation in the SEM

The importance of the SEM representation lies in the fact that, under nonrestrictive assumptions, its states can be estimated with null uncertainty. To see this, note that equation (2.20) implies that $\mathbf{a}_t = \mathbf{z}_t - \mathbf{H} \mathbf{x}_t - \mathbf{D} \mathbf{u}_t$. Substituting this expression in (2.19) and rearranging the resulting terms yields:

$$\mathbf{x}_{t+1} = (\Phi - \mathbf{K} \mathbf{H}) \mathbf{x}_t + (\Gamma - \mathbf{K} \mathbf{D}) \mathbf{u}_t + \mathbf{K} \mathbf{z}_t. \quad (2.21)$$

Therefore, if the initial state \mathbf{x}_1 were known, then the entire sequence of states would be determined by (2.21) because all the terms on the right-hand-side of this expression would be known exactly.

³The assumption that the system has no inputs avoids the interference of an exogenous component that could be nonstationary. Equivalent results can be formulated under the assumption that all the inputs are stationary.

However, usually \mathbf{x}_1 is not known, but is treated as a random vector. Therefore, the propagation of (2.21) will converge exponentially fast to the exact state values as the number of samples processed increases, under the following *invertibility* condition.

Definition 2.10 (Invertibility): Model (2.19)–(2.20) is said to be invertible if all the roots of $|I - (\Phi - \mathbf{K} \mathbf{H}) \lambda| = 0$ lie outside the unit circle, where $\mathbf{K} = (\Phi \mathbf{P} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}^T) \mathbf{B}^{-1}$.

This requirement is not restrictive since the SEM representation was obtained assuming that the system is detectable. It can be proved that detectability ensures that the algebraic Riccati equation (2.25) has a strong solution, which in turn guarantees invertibility of the SEM; see Theorem 3.2. in De Souza, Gevers, and Goodwin [71].

Note that Definition 2.10 above is limited to models in SEM form because it uses the error distribution matrix \mathbf{K} in (2.19). On the other hand, it can be extended to a MEM model just by obtaining the equivalent SEM formulation by means of the conversion procedure described in the next subsection.

2.2.3 Obtaining the SEM equivalent to a MEM

At first sight, the SEM (2.19)–(2.20) may seem to be a particular case of (2.1)–(2.2), constrained so that $\mathbf{w}_t = \mathbf{v}_t = \mathbf{a}_t$ and $\mathbf{C} = \mathbf{I}$. However, under weak assumptions model (2.1)–(2.2) can also be written as an observationally equivalent SEM, and thus both representations are equally general in practice.

Consider the following MEM for \mathbf{z}_t :

$$\bar{\mathbf{x}}_{t+1} = \Phi \bar{\mathbf{x}}_t + \Gamma \mathbf{u}_t + \mathbf{E} \mathbf{w}_t \quad (2.22)$$

$$\mathbf{z}_t = \mathbf{H} \bar{\mathbf{x}}_t + \mathbf{D} \mathbf{u}_t + \mathbf{C} \mathbf{v}_t, \quad (2.23)$$

where the mean and covariance of the initial state are denoted, respectively, by $\bar{\mu}_1$ and $\bar{\mathbf{P}}_1$. Under these conditions, the time series \mathbf{z}_t in (2.23) is also the output of the SEM (2.19)–(2.20) with:

$$\mathbf{x}_1 \sim (\bar{\mu}_1, \bar{\mathbf{P}}_1 - \mathbf{P}) \quad (2.24)$$

$$\mathbf{P} = \Phi \mathbf{P} \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T - \mathbf{K} \mathbf{B} \mathbf{K}^T \quad (2.25)$$

$$\mathbf{K} = (\Phi \mathbf{P} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}^T) \mathbf{B}^{-1} \quad (2.26)$$

$$\mathbf{B} = \mathbf{H} \mathbf{P} \mathbf{H}^T + \mathbf{C} \mathbf{R} \mathbf{C}^T. \quad (2.27)$$

This result needs only two nonrestrictive assumptions: (a) that the system (2.22)–(2.23) be detectable (see Definition 2.6), and (b) that the difference between the covariances of the initial states $\bar{\mathbf{P}}_1 - \mathbf{P}$ be positive semi-definite⁴. The proof of this result can be found in Casals, Sotoca and Jerez [45]. Note that:

1. We use the same notation for the coefficient matrices Φ , Γ , \mathbf{H} and \mathbf{D} in the MEM and SEM forms because they coincide when the observational equivalence between both formulations is enforced according to equations (2.24)–(2.27).

⁴This condition is sometimes justified using the immemorial time hypothesis; see Definition 2.8 and De Jong and Chu-Chun-Lin [69].

2. Expressions (2.25)–(2.27) can be interpreted as the steady-state of the Kalman filter equations, see Chapter 4.
3. The MEM (2.22)–(2.23) includes two error terms whereas the SEM (2.19)–(2.20) has only one. This means that, despite the fact that both are observationally equivalent, some information which is explicit in the covariance matrices of the MEM (\mathbf{Q} , \mathbf{R} and \mathbf{S}) is lost when they are combined in the single error covariance matrix \mathbf{B} of the SEM.
4. If the MEM parameter matrices are known, equations (2.24)–(2.27) can be solved for the SEM parameters and vice versa. The key to this process lies in solving equations (2.25)–(2.27) using the efficient and stable procedures described by Ionescu, Oara and Weiss [128].
5. Choosing the strong solution of the Riccati equation (2.25) has an important advantage, as it suffices to ensure that the eigenvalues of $(\Phi - \mathbf{E}\mathbf{H})$ will lie in or within the unit circle, thus resulting in an invertible VARMAX model.

Example 2.2.3.1 Hodrick-Prescott filter

Consider a trend plus error decomposition:

$$z_t = \mu_t + \varepsilon_t \quad (2.28)$$

where the unobserved trend, μ_t , is governed by an integrated random walk process:

$$\begin{bmatrix} \mu_{t+1} \\ \beta_{t+1} \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t \\ \beta_t \end{bmatrix} + \begin{bmatrix} 0 \\ \zeta_t \end{bmatrix} \quad (2.29)$$

with the error covariance matrix:

$$cov \begin{bmatrix} \zeta_t \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} 1/1600 & 0 \\ 0 & 1 \end{bmatrix}. \quad (2.30)$$

Note that the noise-variance ratio of 1/1600 in (2.30) is chosen to make this model compatible with the Hodrick and Prescott [126] filter for quarterly data, see Harvey and Trimbur [120]. Solving equations (2.24)–(2.27) for this model yields the equivalent SEM:

$$z_t = \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + a_t ; \sigma_a^2 = 1.2509 \quad (2.31)$$

$$\begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \begin{bmatrix} 0.2229 \\ 0.0224 \end{bmatrix} a_t. \quad (2.32)$$

The observational equivalence of models (2.28)–(2.30) and (2.31)–(2.32) can be tested by checking that they yield the same innovations, likelihood values, and out-of-sample point forecasts.

2.2.4 Obtaining the SEM equivalent to a general linear process

The procedure used to obtain the SEM representation described in Subsection 2.2.3 is very general, as it supports any model written as a linear fixed coefficient SS model. However, it requires solving the nontrivial equations (2.25)–(2.27). This can be avoided when the model for the time series is written as a general linear process or in VARMAX form because, in this case, the matrices in (2.19)–(2.20) can be obtained directly from the model parameters.

Assume that z_t follows the model:

$$\bar{\mathbf{F}}(\mathbf{B}) \mathbf{z}_t = \bar{\mathbf{G}}(\mathbf{B}) \mathbf{u}_t + \bar{\mathbf{L}}(\mathbf{B}) \mathbf{a}_t, \quad (2.33)$$

where \mathbf{u}_t and \mathbf{a}_t are, as in previous sections, vectors of exogenous variables and white noise errors, respectively, \mathbf{B} denotes the backshift operator, such that for any ω_t : $\mathbf{B}^i \omega_t = \omega_{t-i}$, $i = 0, \pm 1, \pm 2, \dots, I$ and the polynomial matrices $\bar{\mathbf{F}}(\mathbf{B})$, $\bar{\mathbf{G}}(\mathbf{B})$ and $\bar{\mathbf{L}}(\mathbf{B})$ are defined by:

$$\bar{\mathbf{F}}(\mathbf{B}) = \sum_{i=0}^p \bar{F}_i \mathbf{B}^i \quad (2.34)$$

$$\bar{\mathbf{G}}(\mathbf{B}) = \sum_{i=0}^s \bar{G}_i \mathbf{B}^i \quad (2.35)$$

$$\bar{\mathbf{L}}(\mathbf{B}) = \sum_{i=0}^q \bar{L}_i \mathbf{B}^i. \quad (2.36)$$

Note that model (2.33)–(2.36) relaxes the standard VARMAX form by not requiring the matrices $\bar{\mathbf{F}}_0$ and $\bar{\mathbf{L}}_0$ to be the identity. Now, following Terceiro [206], (2.33) can be written as an the equivalent SEM by pre-multiplying (2.33) by $\bar{\mathbf{F}}_0^{-1}$, which yields:

$$\mathbf{F}(\mathbf{B}) \mathbf{z}_t = \mathbf{G}(\mathbf{B}) \mathbf{u}_t + \mathbf{L}(\mathbf{B}) \mathbf{a}_t \quad (2.37)$$

with $\mathbf{F}(\mathbf{B}) = \bar{\mathbf{F}}_0^{-1} \bar{\mathbf{F}}(\mathbf{B})$, $\mathbf{G}(\mathbf{B}) = \bar{\mathbf{F}}_0^{-1} \bar{\mathbf{G}}(\mathbf{B})$, and $\mathbf{L}(\mathbf{B}) = \bar{\mathbf{F}}_0^{-1} \bar{\mathbf{L}}(\mathbf{B})$. The matrices of the SEM equivalent to (2.37) are then:

$$\Phi = \begin{bmatrix} -\mathbf{F}_1 & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ -\mathbf{F}_2 & \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\mathbf{F}_{k-1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} \\ -\mathbf{F}_k & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix} \quad (2.38)$$

$$\Gamma = \begin{bmatrix} \mathbf{G}_1 - \mathbf{F}_1 \mathbf{G}_0 \\ \mathbf{G}_2 - \mathbf{F}_2 \mathbf{G}_0 \\ \vdots \\ \mathbf{G}_{k-1} - \mathbf{F}_{k-1} \mathbf{G}_0 \\ \mathbf{G}_k - \mathbf{F}_k \mathbf{G}_0 \end{bmatrix} \quad (2.39)$$

$$\mathbf{E} = \begin{bmatrix} \mathbf{L}_1 - \mathbf{F}_1 \\ \mathbf{L}_2 - \mathbf{F}_2 \\ \vdots \\ \mathbf{L}_{k-1} - \mathbf{F}_{k-1} \\ \mathbf{L}_k - \mathbf{F}_k \end{bmatrix} \quad (2.40)$$

$$\mathbf{H} = [\mathbf{I} \ 0 \ \dots \ 0] \quad (2.41)$$

$$\mathbf{D} = \mathbf{G}_0, \quad (2.42)$$

where the error term coincides with the innovations in model (2.33), so that $\mathbf{B} = \text{cov}(\mathbf{a}_t)$. The dimensions of the matrices (2.38)–(2.42) are governed by the dynamic order of the system $k = \max \{p, s, q\}$ and, according to this order, the dimension of the state vector is $n = mk$, being m the dimension of \mathbf{z}_t .

Example 2.2.1 ARIMA models

Consider the following MA(1) and ARMA(1,1) models:

$$z_{t+1} = a_{t+1} - \theta a_t \quad (2.43)$$

$$z_{t+1} = \phi z_t + a_{t+1} - \theta a_t. \quad (2.44)$$

Applying the general transformation defined by equations (2.38)–(2.42) to these models we obtain:

$$x_{t+1} = 0 \cdot x_t - \theta a_t \quad (2.45)$$

$$z_t = x_t + a_t \quad (2.46)$$

$$x_{t+1} = \phi x_t + (\phi - \theta) a_t \quad (2.47)$$

$$z_t = x_t + a_t, \quad (2.48)$$

where (2.45)–(2.46) is equivalent to (2.43), and (2.47)–(2.48) is equivalent to (2.44).

Example 2.2.2 Transfer functions

Consider also the single-input single-output (SISO) transfer function model:

$$z_t = \frac{\omega_0 - \omega_1 B}{1 - \delta_1 B} u_t + (1 - \theta B) a_t. \quad (2.49)$$

The equivalent SS formulation in this case would be:

$$\begin{aligned} \begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \end{bmatrix} &= \begin{bmatrix} \delta_1 & 1 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \begin{bmatrix} -\omega_1 + \delta_1 \omega_0 \\ 0 \end{bmatrix} u_t + \begin{bmatrix} \theta \\ -\theta \delta_1 \end{bmatrix} a_t \\ z_t &= \begin{bmatrix} 1 & 0 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \omega_0 u_t + a_t. \end{aligned} \quad (2.50)$$

Example 2.2.3 VARMAX model

As a final example, consider the VAR(2) model:

$$\begin{bmatrix} \phi_{11} & \phi_{12} \\ 0 & \phi_{22} \end{bmatrix} \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} = \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix}, \quad (2.51)$$

whose equivalent SS representation is:

$$\begin{aligned} \begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \end{bmatrix} &= \begin{bmatrix} -\phi_{11} & -\phi_{12} \\ 0 & -\phi_{22} \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \begin{bmatrix} -\phi_{11} & -\phi_{12} \\ 0 & -\phi_{22} \end{bmatrix} \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix} \\ \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} &= \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix}. \end{aligned} \quad (2.52)$$

This page intentionally left blank

Chapter 3

Model transformations

3.1	Model decomposition	19
3.1.1	Deterministic and stochastic subsystems	19
3.1.2	Implied univariate models	21
3.1.3	Block-diagonal forms	22
3.2	Model combination	24
3.2.1	Endogeneization of stochastic inputs	24
3.2.2	Colored errors	26
3.3	Change of variables in the output	28
3.3.1	Observation errors	28
3.3.2	Missing values and aggregated data	29
3.3.3	Temporal aggregation	30
3.4	Uses of these transformations	31

In this chapter we discuss different model transformations. These can be classified into three groups: model decomposition, model combination, and changes in the output variable. The last section discusses their practical applications.

3.1 Model decomposition

The expression “model decomposition” refers to the extraction of meaningful submodels from a given initial formulation.

3.1.1 Deterministic and stochastic subsystems

Assume that \mathbf{z}_t is the observable output of the MEM (2.1)–(2.2). Consider also the additive decompositions:

$$\mathbf{z}_t = \mathbf{z}_t^d + \mathbf{z}_t^s \quad (3.1)$$

$$\mathbf{x}_t = \mathbf{x}_t^d + \mathbf{x}_t^s \quad (3.2)$$

where z_t^d denotes an additive “deterministic” component of the time series, which depends exclusively on the value of the inputs, while z_t^s refers to the “stochastic” component, which only depends on the errors. Accordingly, x_t^d and x_t^s are the deterministic and stochastic components of the state vector, respectively. Substituting (3.2) into (2.1)–(2.2) and (3.1) into (2.2) yields:

$$\mathbf{x}_{t+1}^d + \mathbf{x}_{t+1}^s = \Phi(\mathbf{x}_t^d + \mathbf{x}_t^s) + \Gamma \mathbf{u}_t + \mathbf{E} \mathbf{w}_t \quad (3.3)$$

$$\mathbf{z}_t^d + \mathbf{z}_t^s = \mathbf{H}(\mathbf{x}_t^d + \mathbf{x}_t^s) + \mathbf{D} \mathbf{u}_t + \mathbf{C} \mathbf{v}_t, \quad (3.4)$$

and (3.3)–(3.4) can be immediately decomposed into two different SS models. The deterministic subsystem, which realizes z_t^d , is:

$$\mathbf{x}_{t+1}^d = \Phi \mathbf{x}_t^d + \Gamma \mathbf{u}_t \quad (3.5)$$

$$\mathbf{z}_t^d = \mathbf{H} \mathbf{x}_t^d + \mathbf{D} \mathbf{u}_t, \quad (3.6)$$

and the stochastic subsystem, whose output is z_t^s , is given by:

$$\mathbf{x}_{t+1}^s = \Phi \mathbf{x}_t^s + \mathbf{E} \mathbf{w}_t \quad (3.7)$$

$$\mathbf{z}_t^s = \mathbf{H} \mathbf{x}_t^s + \mathbf{C} \mathbf{v}_t. \quad (3.8)$$

This deterministic/stochastic decomposition is a standard result of linear systems theory, see e.g., Van Overschee and De Moor [215] or Casals, Jerez, and Sotoca [42]. The terms “deterministic” and “stochastic” may be confusing for some readers, who would probably find the terms “input-driven” and “error-driven” to be more accurate. Taking this remark into account, we will maintain the original terminology. In Chapter 8 we will show how this decomposition can be applied to extract the deterministic components from a time series.

The following example illustrates this decomposition on the basis of a simple transfer function model.

Example 3.1.1 Obtaining the deterministic and stochastic subsystems implied by a transfer function

Consider the following SISO transfer function:

$$z_t = \frac{0.5}{1 - 0.7B} u_t + \frac{1 - 0.8B}{1 - B} a_t. \quad (3.9)$$

With this notation, writing the models for the deterministic and stochastic components is trivial, since taking into account expression (3.1), it is obvious that:

$$z_t^d = \frac{0.5}{1 - 0.7B} u_t \quad (3.10)$$

$$z_t^s = \frac{1 - 0.8B}{1 - B} a_t. \quad (3.11)$$

On the other hand, following the procedure described in Subsection 2.2.4, it is easy to obtain the SS representation of model (3.9):

$$\begin{aligned} \begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \end{bmatrix} &= \begin{bmatrix} 1.7 & 1 \\ -0.7 & 0 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \begin{bmatrix} 0.35 \\ -0.35 \end{bmatrix} u_t + \begin{bmatrix} 0.2 \\ -0.14 \end{bmatrix} a_t \\ z_t = [1 & 0] \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} &+ 0.5 u_t + a_t, \end{aligned} \quad (3.12)$$

which can be decomposed into the following additive deterministic and stochastic subsystems:

$$\begin{bmatrix} x_{t+1}^{1,d} \\ x_{t+1}^{2,d} \end{bmatrix} = \begin{bmatrix} 1.7 & 1 \\ -0.7 & 0 \end{bmatrix} \begin{bmatrix} x_t^{1,d} \\ x_t^{2,d} \end{bmatrix} + \begin{bmatrix} 0.35 \\ -0.35 \end{bmatrix} u_t \quad (3.13)$$

$$\begin{aligned} z_t^d &= [1 & 0] \begin{bmatrix} x_t^{1,d} \\ x_t^{2,d} \end{bmatrix} + 0.5 u_t \\ \begin{bmatrix} x_{t+1}^{1,s} \\ x_{t+1}^{2,s} \end{bmatrix} &= \begin{bmatrix} 1.7 & 1 \\ -0.7 & 0 \end{bmatrix} \begin{bmatrix} x_t^{1,s} \\ x_t^{2,s} \end{bmatrix} + \begin{bmatrix} 0.2 \\ -0.14 \end{bmatrix} a_t \end{aligned} \quad (3.14)$$

$$z_t^s = [1 & 0] \begin{bmatrix} x_t^{1,s} \\ x_t^{2,s} \end{bmatrix} + a_t.$$

3.1.2 Implied univariate models

Consider a minimal system written in the SEM form (2.19)–(2.20). If we combine the state equation (2.19) with the i -th row of the observation equation (2.20), we obtain the implied univariate model for the i -th variable in z_t :

$$\mathbf{x}_{t+1} = \boldsymbol{\Phi} \mathbf{x}_t + \boldsymbol{\Gamma} \mathbf{u}_t + \boldsymbol{K} \mathbf{a}_t \quad (3.15)$$

$$z_t^i = \mathbf{H}^i \mathbf{x}_t + \mathbf{D}^i \mathbf{u}_t + \mathbf{a}_t^i, \quad (3.16)$$

where z_t^i , \mathbf{H}^i , \mathbf{D}^i , and \mathbf{a}_t^i denote the i -th row of the corresponding vectors and matrices. Note that the i -th variable z_t^i may be unaffected by some states in \mathbf{x}_t , so that minimality of (2.1)–(2.2) does not ensure that model (3.15)–(3.16) is minimal, see Definition 2.4. To obtain the minimal representation we could apply the aforementioned staircase algorithm to the implied univariate model, see Rosenbrock [190], to delete any unobservable/non-controllable modes and, accordingly, to obtain a minimal representation for z_t^i . Note also that Chapter 9 presents two procedures to obtain the VARMAX representation equivalent to an SS model. Applying them to model (3.15)–(3.16) would result in the equivalent ARIMAX form.

The following example illustrates these ideas by decomposing a simple bivariate model.

Example 3.1.2 Obtaining the ARIMA models implied by a VARMAX process

Consider the VARMA model:

$$\begin{bmatrix} 1 - 0.8B & 0.2B \\ -0.3B & 1 \end{bmatrix} \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 - 0.5B \end{bmatrix} \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix}, \quad (3.17)$$

whose equivalent SEM representation is given by:

$$\begin{aligned} \begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \end{bmatrix} &= \begin{bmatrix} 0.8 & -0.2 \\ 0.3 & 0 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \begin{bmatrix} 0.8 & -0.2 \\ 0.3 & -0.5 \end{bmatrix} \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix}, \\ \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} &= \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix}. \end{aligned} \quad (3.18)$$

In this case, the implied SS model for the first output would be given by the state equation in (3.18) and the first row of its observation equation, that is:

$$z_t^1 = x_t^1 + a_t^1. \quad (3.19)$$

According to (3.18), the implied model for the second output would be given by the same state equation and the second row of its observation equation:

$$z_t^2 = x_t^2 + a_t^2. \quad (3.20)$$

Finally, applying the procedures discussed in Chapter 9, we would obtain the ARIMA processes implied by models (3.17) or (3.18):

$$(1 - 0.800B + 0.060B^2) z_t^1 = (1 - 0.111B + 0.048B^2) \varepsilon_t^1, \quad \sigma_{\varepsilon^1}^2 = 1.035, \quad (3.21)$$

$$(1 - 0.800B + 0.060B^2) z_t^2 = (1 - 1.004B + 0.332B^2) \varepsilon_t^2, \quad \sigma_{\varepsilon^2}^2 = 1.204, \quad (3.22)$$

where the errors ε_t^1 and ε_t^2 are functions of a_t^1 and a_t^2 , see Chapter 9.

The code required to replicate this example can be found in the file `Example0312.m`, see Appendix D.

3.1.3 Block-diagonal forms

As we saw in Section 2.1, if \mathbf{z}_t is realized by (2.1)–(2.2), then it is also realized by the model:

$$\bar{\mathbf{x}}_{t+1} = \bar{\Phi} \bar{\mathbf{x}}_t + \bar{\Gamma} \mathbf{u}_t + \bar{\mathbf{E}} \mathbf{w}_t \quad (3.23)$$

$$\mathbf{z}_t = \bar{\mathbf{H}} \bar{\mathbf{x}}_t + \mathbf{D} \mathbf{u}_t + \mathbf{C} \mathbf{v}_t, \quad (3.24)$$

where the states in (3.23)–(3.24) are related to those in (2.1)–(2.2) by the similar transformation $\mathbf{T} \mathbf{x}_t = \bar{\mathbf{x}}_t$, see Subsection 2.1.2.

The transformation matrix \mathbf{T} is typically chosen so that the resulting model has some specific advantage. In particular, we could determine \mathbf{T} so that the transformed

transition matrix $\bar{\Phi}$ is *block-diagonal*¹. The advantage of this specific decomposition lies in the fact that the states of the transformed model are uniquely associated with the eigenvalues of the transition matrix which characterize the dynamic structure of the model, as well as its properties in the frequency domain.

Specifically, if $\lambda_{j,k} = a_j \pm b_j i$ is a pair of complex conjugate eigenvalues of Φ , the corresponding states will generate a peak in the pseudo-spectrum of the series at the frequency $f_{j,k} = (2\pi)^{-1} \arctan(b_j/a_j)$ cycles per time unit, and the corresponding values of the period (p) and damping factor (df) would then be $p = 1/f_{j,k}$ and $df = \|\lambda_{j,k}\|$, where $\|\cdot\|$ denotes modulus. In this way, a block-diagonal model decomposes the response to shocks in u_t and w_t into the basic movements corresponding to the different eigenvalues of the transition matrix, which represent independent reactions of the state variables to an excitation in any input of the state equation. The following example clarifies these ideas.

Example 3.1.3 An ARIMA model in block-diagonal form

Consider the ARIMA(2,1,1) model:

$$(1 + 0.75B + 0.25B^2)(1 - B)z_t = (1 - 0.5B)a_t, \quad (3.25)$$

which has a unit root and two stationary AR roots: $-1.5 \pm 1.323i$. Applying the method described in Subsection 2.2.4, we can obtain the equivalent SEM:

$$\begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \\ x_{t+1}^3 \end{bmatrix} = \begin{bmatrix} 0.25 & 1 & 0 \\ 0.50 & 0 & 1 \\ 0.25 & 0 & 0 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \\ x_t^3 \end{bmatrix} + \begin{bmatrix} -0.25 \\ 0.50 \\ 0.25 \end{bmatrix} a_t \quad (3.26)$$

$$z_t = [1 \ 0 \ 0] \begin{bmatrix} x_t^1 \\ x_t^2 \\ x_t^3 \end{bmatrix} + a_t, \quad (3.27)$$

in which the states do not have a clear interpretation. However, the previously described transformation provides the equivalent block-diagonal form:

$$\begin{bmatrix} \bar{x}_{t+1}^1 \\ \bar{x}_{t+1}^2 \\ \bar{x}_{t+1}^3 \end{bmatrix} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.375 & 0.909 \\ 0 & -0.120 & -0.375 \end{bmatrix} \begin{bmatrix} \bar{x}_t^1 \\ \bar{x}_t^2 \\ \bar{x}_t^3 \end{bmatrix} + \begin{bmatrix} -0.319 \\ 0.594 \\ 0.022 \end{bmatrix} a_t \quad (3.28)$$

$$z_t = [-0.785 \ -0.833 \ -0.238] \begin{bmatrix} \bar{x}_t^1 \\ \bar{x}_t^2 \\ \bar{x}_t^3 \end{bmatrix} + a_t, \quad (3.29)$$

¹The most famous block-diagonalizing algorithm is the Jordan decomposition which, on the other hand, is very unstable, see Golub and Van Loan [109]. We find more convenient a simpler and more stable method which consists of applying the real Schur decomposition to Φ and diagonalizing the resulting real Schur matrix by solving a system of Sylvester equations, see Petkov, Christov and Konstantinov [181]; see Appendix A.

which clearly shows that the first state corresponds to the unit root, while the second and third states describe the harmonic cycle generated by the complex AR roots. Therefore, block-diagonalizing an SS model yields an equivalent form in which the states can be interpreted as unobserved structural components of the time series, see Casals, Jerez, and Sotoca [41].

The code required to replicate this example can be found in the file `Example0313.m`, see Appendix D.

3.2 Model combination

The expression “model combination” refers to the composition of different models into a single SS formulation. In practice, model combination is a useful technique for implementing a complex model by defining simpler pieces and then joining them together.

3.2.1 Endogeneization of stochastic inputs

Following Terceiro [206], Chapter 3, consider the MEM formulation:

$$\mathbf{x}_{t+1}^z = \Phi^z \mathbf{x}_t^z + \Gamma \mathbf{u}_t + \mathbf{E}^z \mathbf{w}_t^z \quad (3.30)$$

$$\mathbf{z}_t = \mathbf{H}^z \mathbf{x}_t^z + \mathbf{D} \mathbf{u}_t + \mathbf{C}^z \mathbf{v}_t^z, \quad (3.31)$$

and an SS model for the inputs in (3.30)–(3.31):

$$\mathbf{x}_{t+1}^u = \Phi^u \mathbf{x}_t^u + \mathbf{E}^u \mathbf{w}_t^u \quad (3.32)$$

$$\mathbf{u}_t = \mathbf{H}^u \mathbf{x}_t^u + \mathbf{C}^u \mathbf{v}_t^u, \quad (3.33)$$

where the errors in (3.30)–(3.31) and (3.32)–(3.33) are assumed to be mutually independent.

Building on this notation, our goal consists of formulating a single model for the output vector $[\mathbf{z}_t^T \mathbf{u}_t^T]^T$. We will refer to this representation as an “endogeneized” or “nested in inputs” model. To this end, we substitute (3.33) in (3.30)–(3.31), obtaining:

$$\mathbf{x}_{t+1}^z = \Phi^z \mathbf{x}_t^z + \Gamma (\mathbf{H}^u \mathbf{x}_t^u + \mathbf{C}^u \mathbf{v}_t^u) + \mathbf{E}^z \mathbf{w}_t^z \quad (3.34)$$

$$\mathbf{z}_t = \mathbf{H}^z \mathbf{x}_t^z + \mathbf{D} (\mathbf{H}^u \mathbf{x}_t^u + \mathbf{C}^u \mathbf{v}_t^u) + \mathbf{C}^z \mathbf{v}_t^z. \quad (3.35)$$

Finally, the state equation of the endogeneized model results from combining expressions (3.34) and (3.32) into a single state equation:

$$\begin{bmatrix} \mathbf{x}_{t+1}^z \\ \mathbf{x}_{t+1}^u \end{bmatrix} = \begin{bmatrix} \Phi^z & \Gamma \mathbf{H}^u \\ 0 & \Phi^u \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^z \\ \mathbf{x}_t^u \end{bmatrix} + \begin{bmatrix} \mathbf{E}^z & \Gamma \mathbf{C}^u & 0 \\ 0 & 0 & \mathbf{E}^u \end{bmatrix} \begin{bmatrix} \mathbf{w}_t^z \\ \mathbf{v}_t^u \\ \mathbf{w}_t^u \end{bmatrix}, \quad (3.36)$$

whose observation equation is obtained by combining equations (3.35) and (3.33):

$$\begin{bmatrix} \mathbf{z}_t \\ \mathbf{u}_t \end{bmatrix} = \begin{bmatrix} \mathbf{H}^z & \mathbf{D} \mathbf{H}^u \\ 0 & \mathbf{H}^u \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^z \\ \mathbf{x}_t^u \end{bmatrix} + \begin{bmatrix} \mathbf{C}^z & \mathbf{D} \mathbf{C}^u \\ 0 & \mathbf{C}^u \end{bmatrix} \begin{bmatrix} \mathbf{v}_t^z \\ \mathbf{v}_t^u \end{bmatrix}. \quad (3.37)$$

This model combination procedure, together with the deterministic-stochastic decomposition presented in Subsection 3.1.1, has a clear implication: there is no loss of generality in defining an SS model without exogenous inputs, \mathbf{u}_t . If the inputs are deterministic or purely exogenous, they can be “deleted” by decomposing the model into its input and error driven terms; see Subsection 3.1.1. On the other hand, if we have models for both \mathbf{z}_t and \mathbf{u}_t , they can be combined into a single formulation without exogenous variables as just described.

Therefore, in the remainder of this book we will avoid unnecessary complexity by considering a purely stochastic SS formulation most often:

$$\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \mathbf{E} \mathbf{w}_t \quad (3.38)$$

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_t + \mathbf{C} \mathbf{v}_t. \quad (3.39)$$

Example 3.2.1 Combining a SISO model with the model for the input

Consider a SISO transfer function:

$$z_t = \frac{0.3 + 0.6B}{1 - 0.7B} u_t + \frac{1 - 0.8B}{1 - B} a_t^z \quad (3.40)$$

whose input follows the ARIMA(1,1) model:

$$u_t = \frac{1 - 0.9B}{1 - B} a_t^u \quad (3.41)$$

and such that:

$$\text{cov} \left(\begin{bmatrix} a_t^z \\ a_t^u \end{bmatrix} \right) = \begin{bmatrix} 1.0 & 0 \\ 0 & 0.5 \end{bmatrix}. \quad (3.42)$$

Following the procedure described above, the first step needed to combine these two models would be to formulate them in SS form. Accordingly, the SEM representation for model (3.40) is:

$$\begin{bmatrix} x_{t+1}^{1,z} \\ x_{t+1}^{2,z} \end{bmatrix} = \begin{bmatrix} 1.70 & 1 \\ -0.70 & 0 \end{bmatrix} \begin{bmatrix} x_t^{1,z} \\ x_t^{2,z} \end{bmatrix} + \begin{bmatrix} 0.81 \\ -0.81 \end{bmatrix} u_t + \begin{bmatrix} 0.20 \\ -0.14 \end{bmatrix} a_t^z \quad (3.43)$$

$$z_t = [1 \ 0] \begin{bmatrix} x_t^{1,z} \\ x_t^{2,z} \end{bmatrix} + 0.30 u_t + a_t^z,$$

while the SS form for the input model (3.41) is:

$$\begin{aligned} x_{t+1}^{1,u} &= x_t^{1,u} + 0.1 a_t^u \\ u_t &= x_t^{1,u} + a_t. \end{aligned} \quad (3.44)$$

Finally, combining (3.43) and (3.44) according to the procedure explained above results in a joint SS model for both the input and output variables:

$$\begin{aligned} \begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \\ x_{t+1}^3 \end{bmatrix} &= \begin{bmatrix} 1.70 & 1 & 0.81 \\ -0.70 & 0 & -0.81 \\ 0 & 0 & 0.1 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \\ x_t^3 \end{bmatrix} + \begin{bmatrix} 0.20 & 0.81 \\ -0.14 & -0.81 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_t^z \\ a_t^u \end{bmatrix} \\ \begin{bmatrix} z_t \\ u_t \end{bmatrix} &= \begin{bmatrix} 1 & 0 & 0.30 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \\ x_t^3 \end{bmatrix} + \begin{bmatrix} 1 & 0.30 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_t^z \\ a_t^u \end{bmatrix}. \end{aligned} \quad (3.45)$$

This can also be written in an equivalent VARMA form by applying the procedures described in Chapter 9:

$$\begin{bmatrix} 1 - 1.70B + 0.70B^2 & 0 \\ 0 & 1 - B \end{bmatrix} \begin{bmatrix} z_t \\ u_t \end{bmatrix} = \begin{bmatrix} 1 - 1.50B + 0.56B^2 & 0.78B - 0.708B^2 \\ 0 & 1 - 0.9B \end{bmatrix} \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix} \quad (3.46)$$

with:

$$\text{cov} \left(\begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix} \right) = \begin{bmatrix} 1.05 & 0.15 \\ 0.15 & 0.50 \end{bmatrix}. \quad (3.47)$$

Note that the error covariance (3.42) is diagonal, while that of the endogenized model, (3.47), is not. This happens because the latter corresponds to the reduced-form VARMAX model (3.46) which captures the instantaneous relationship between z_t and u_t as a nonzero error covariance. On the other hand, models (3.40) and (3.45) are structural forms, which represent this relationship by means of a parameter in the model structure. Accordingly, the errors in the structural form of this model are uncorrelated.

The code required to replicate this example can be found in the file Example0321.m, see Appendix D.

3.2.2 Colored errors

Consider an SS model in SEM form:

$$x_{t+1}^z = \Phi^z x_t^z + E^z a_t^a \quad (3.48)$$

$$z_t = H^z x_t^z + a_t^a, \quad (3.49)$$

where the errors a_t^a are colored noise realized by the model:

$$x_{t+1}^a = \Phi^a x_t^a + E^a a_t \quad (3.50)$$

$$a_t^a = H^a x_t^a + a_t, \quad (3.51)$$

and a_t is a vector of zero-mean white noise errors. Once again, we will formulate a single model combining the dynamic structure of z_t with that of the colored errors a_t^a . We refer to this formulation as a “nested in errors” model. Substituting (3.51) in (3.48) and (3.49) yields:

$$x_{t+1}^z = \Phi^z x_t^z + E^z (H^a x_t^a + a_t) \quad (3.52)$$

$$\mathbf{z}_t = \mathbf{H}^z \mathbf{x}_t^z + \mathbf{H}^a \mathbf{x}_t^a + \mathbf{a}_t. \quad (3.53)$$

Last, the state equation of the nested model results from combining equations (3.52) and (3.50) into a single state equation:

$$\begin{bmatrix} \mathbf{x}_{t+1}^z \\ \mathbf{x}_{t+1}^a \end{bmatrix} = \begin{bmatrix} \Phi^z & \mathbf{E}^z \mathbf{H}^a \\ \mathbf{0} & \Phi^a \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^z \\ \mathbf{x}_t^a \end{bmatrix} + \begin{bmatrix} \mathbf{E}^z \\ \mathbf{E}^a \end{bmatrix} \mathbf{a}_t, \quad (3.54)$$

and the nested observation equation results from rearranging terms in (3.53):

$$\mathbf{z}_t = [\mathbf{H}^z \quad \mathbf{H}^a] \begin{bmatrix} \mathbf{x}_t^z \\ \mathbf{x}_t^a \end{bmatrix} + \mathbf{a}_t. \quad (3.55)$$

This procedure is useful in defining complex models by combining different multiplicative “building blocks.” The following example clarifies this idea.

Example 3.2.2 Defining a model as the product of several dynamic factors

Consider the AR(4) model:

$$(1 - 2.10B + 1.58B^2 - 0.62B^3 + 0.14B^4) z_t = a_t, \quad (3.56)$$

where the roots of the polynomial on the left-hand side are $1 \pm 2i$, $1/0.7$ and 1 . Assume now that we want to define model (3.56) in the equivalent factor form²:

$$(1 - 0.4B + 0.2B^2)(1 - 0.7B)(1 - B) z_t = a_t. \quad (3.57)$$

The starting point for this is to obtain the SS representations for the factors. Accordingly, the model for the factors $(1 - B)(1 - 0.7B) z_t = a_t^a$ is:

$$\begin{bmatrix} x_{t+1}^{1,z} \\ x_{t+1}^{2,z} \end{bmatrix} = \begin{bmatrix} 1.70 & 1 \\ -0.70 & 0 \end{bmatrix} \begin{bmatrix} x_t^{1,z} \\ x_t^{2,z} \end{bmatrix} + \begin{bmatrix} 1.7 \\ -0.7 \end{bmatrix} a_t^a \quad (3.58)$$

$$z_t = [1 \quad 0] \begin{bmatrix} x_t^{1,z} \\ x_t^{2,z} \end{bmatrix} + a_t^a,$$

and the model for the factor $(1 - 0.4B + 0.2B^2) a_t^a = a_t$ is:

$$\begin{bmatrix} x_{t+1}^{1,a} \\ x_{t+1}^{2,a} \end{bmatrix} = \begin{bmatrix} 0.4 & 1 \\ -0.2 & 0 \end{bmatrix} \begin{bmatrix} x_t^{1,a} \\ x_t^{2,a} \end{bmatrix} + \begin{bmatrix} 0.4 \\ -0.2 \end{bmatrix} a_t \quad (3.59)$$

$$a_t^a = [1 \quad 0] \begin{bmatrix} x_t^{1,a} \\ x_t^{2,a} \end{bmatrix} + a_t.$$

²There are many reasons why one may want to work with a model in factor form. For example, one may want to estimate its real roots and standard errors separately, in order to infer if this could be a unit root by means of a Dickey and Fuller [77] type statistic. It would also be useful to define a model with multiple seasonal factors such as, e.g., the ARIMA process describing an hourly time series with daily and weekly seasonal cycles, whose periods would then be 24 and 24×7 observations per cycle respectively.

Under these conditions, the SS representation of (3.57) is obtained by combining (3.58) and (3.59) into the composite formulation:

$$\begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \\ x_{t+1}^3 \\ x_{t+1}^4 \end{bmatrix} = \begin{bmatrix} 1.7 & 1 & 1.7 & 0 \\ -0.7 & 0 & -0.7 & 0 \\ 0 & 0 & 0.4 & 1 \\ 0 & 0 & -0.2 & 0 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \\ x_t^3 \\ x_t^4 \end{bmatrix} + \begin{bmatrix} 1.7 \\ -0.7 \\ 0.4 \\ -0.2 \end{bmatrix} a_t \quad (3.60)$$

$$z_t = [1 \ 0 \ 1 \ 0] \begin{bmatrix} x_t^1 \\ x_t^2 \\ x_t^3 \\ x_t^4 \end{bmatrix} + a_t.$$

The example in Subsection 5.4.2 illustrates how these formulations allow one to capture a complex dynamic structure by means of a nonstandard formulation.

The code required to replicate this example can be found in the file Example0322.m, see Appendix D.

3.3 Change of variables in the output

The expression “change of variables in the output” refers to a situation where the model output cannot be observed directly, usually because of certain imperfections in the data, such as observation errors, missing values, or aggregation constraints.

3.3.1 Observation errors

Assume that the output of a MEM model in the form (3.38)–(3.39), z_t , is affected by a white noise error, so that the observed variable is \bar{z}_t such that:

$$\bar{z}_t = z_t + \tilde{z}_t, \quad (3.61)$$

where \tilde{z}_t is an uncorrelated observation error. In this case, the derivation of a model for \bar{z}_t follows immediately by adding \tilde{z}_t to both sides of the observation equation (3.39):

$$z_t + \tilde{z}_t = \mathbf{H} \mathbf{x}_t + \mathbf{C} \mathbf{v}_t + \tilde{z}_t, \quad (3.62)$$

which simplifies to:

$$\bar{z}_t = \mathbf{H} \mathbf{x}_t + [\mathbf{C} \ \mathbf{I}] \begin{bmatrix} \mathbf{v}_t \\ \tilde{z}_t \end{bmatrix}. \quad (3.63)$$

Note that the case of colored observational noise can be treated by combining the previous change of variable with the procedure described in Subsection 3.2.2.

Example 3.3.1 Defining an AR(2) model with white noise observation errors

Consider the AR(2) model:

$$(1 - 0.4B + 0.2B^2) z_t = a_t, \quad (3.64)$$

whose output is affected by an observation error, such that $\bar{z}_t = z_t + \tilde{z}_t$. The errors a_t and \tilde{z}_t are zero-mean white noise sequences such that:

$$\text{cov} \left(\begin{bmatrix} a_t \\ \tilde{z}_t \end{bmatrix} \right) = \begin{bmatrix} 1.0 & 0 \\ 0 & 0.5 \end{bmatrix}. \quad (3.65)$$

Following the procedure described in Subsection 2.2.4, the SS model for the unobserved variable z_t would be:

$$\begin{aligned} \begin{bmatrix} x_{t+1}^{1,z} \\ x_{t+1}^{2,z} \end{bmatrix} &= \begin{bmatrix} 0.4 & 1 \\ -0.2 & 0 \end{bmatrix} \begin{bmatrix} x_t^{1,z} \\ x_t^{2,z} \end{bmatrix} + \begin{bmatrix} 0.4 \\ -0.2 \end{bmatrix} a_t \\ z_t &= [1 \ 0] \begin{bmatrix} x_t^{1,z} \\ x_t^{2,z} \end{bmatrix} + a_t, \end{aligned} \quad (3.66)$$

and, following the procedure described above, the MEM model for the error-contaminated variable has the same state equation³, but the following observation equation:

$$\bar{z}_t = [1 \ 0] \begin{bmatrix} x_t^{1,z} \\ x_t^{2,z} \end{bmatrix} + [1 \ 1] \begin{bmatrix} a_t \\ \tilde{z}_t \end{bmatrix}, \quad (3.67)$$

which can also be written in an equivalent ARMA form by applying the procedures described in Chapter 9 as:

$$(1 - 0.400B + 0.200B^2) \bar{z}_t = (1 - 0.145B + 0.064B^2) \bar{a}_t \quad \sigma_{\bar{a}}^2 = 1.561. \quad (3.68)$$

The code required to replicate this example can be found in the file `Example0331.m`, see Appendix D.

3.3.2 Missing values and aggregated data

Consider a MEM in the form (3.38)–(3.39), and assume now that we do not observe the m components of z_t directly, but the \bar{m} components of \bar{z}_t instead, such that:

$$\bar{z}_t = \bar{H}_t z_t, \quad (3.69)$$

where \bar{H}_t is a $\bar{m} \times m$ matrix of known coefficients which defines how the vector z_t is observed at time t . Some particular cases of interest are:

1. $\bar{H}_t = I_m$, so that $\bar{z}_t = z_t$;
2. $\bar{H}_t = 0$, in which case the t -th observation is missing;
3. $\bar{H}_t = [I_{\bar{m}} \ 0]$, where the last $m - \bar{m}$ components of the output are not observed;

³This happens because a white noise observation error does not add new dynamics to the system. A colored error would add its own dynamics to those of the unobserved variable and change the state equation accordingly.

4. $\bar{H}_t = [1, 1, \dots, 1]$, which implies that we only observe the sum of the m components in z_t .

Building on model (3.38)–(3.39) and the time-varying observation restriction (3.69), it is easy to see that the model for \bar{z}_t will have the state equation (3.38), and an observation equation that is obtained by pre-multiplying (3.39) by \bar{H}_t :

$$\bar{z}_t = \bar{H}_t H x_t + \bar{H}_t C v_t. \quad (3.70)$$

The fact that the change in variables does not affect the state equation means that the system dynamics are not altered by any linear constraints on the observation of its output. On the other hand, if the observation equation is changed, some dynamic components of model (3.38)–(3.39) may lose observability in model (3.38)–(3.70).

Finally, note that equation (3.70) has time-varying parameters. As is well known, see e.g. Anderson and Moore [6], most SS methods can be applied to systems with time-varying coefficients, see e.g. Chapter 6. In this book we opted to simplify the notation by dropping the time subindex in the matrices of the SS model. Despite this concession to clarity, most of the procedures described here can be used in the time-varying parameter case (with the obvious modifications) and are therefore able to deal with missing and aggregated data.

3.3.3 Temporal aggregation

Consider the MEM (3.38)–(3.39), defined for $t = 1, 2, \dots, T$. Without loss of generality, we will assume that z_t is a quarterly time series observable only at times $t = 4, 8, 12, \dots$ with an aggregation constraint determined by its nature as either a flow or a stock variable. In the first case, the sample would include the values $z_t + z_{t-1} + z_{t-2} + z_{t-3}$ ($t = 4, 8, 12, \dots$). On the other hand, if z_t is a stock, the observed values would be z_t ($t = 4, 8, 12, \dots$).

Following Terceiro [206], the model for the aggregated values of a flow variable can be formulated by augmenting the state vector by as many lags of z_t as needed to fit the aggregation rule. By doing so, we obtain:

$$\begin{bmatrix} x_{t+1} \\ z_t \\ z_{t-1} \\ z_{t-2} \\ z_{t-3} \end{bmatrix} = \begin{bmatrix} \Phi & 0 & 0 & 0 & 0 \\ H & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \end{bmatrix} \begin{bmatrix} x_t \\ z_{t-1} \\ z_{t-2} \\ z_{t-3} \\ z_{t-4} \end{bmatrix} + \begin{bmatrix} E & 0 \\ 0 & C \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} w_t \\ v_t \end{bmatrix} \quad (3.71)$$

$$\begin{bmatrix} z_t \\ z_{t-1} \\ z_{t-2} \\ z_{t-3} \end{bmatrix} = \begin{bmatrix} H & 0 & 0 & 0 & 0 \\ 0 & I & 0 & 0 & 0 \\ 0 & 0 & I & 0 & 0 \\ 0 & 0 & 0 & I & 0 \end{bmatrix} \begin{bmatrix} x_t \\ z_{t-1} \\ z_{t-2} \\ z_{t-3} \\ z_{t-4} \end{bmatrix} + \begin{bmatrix} C \\ 0 \\ 0 \\ 0 \end{bmatrix} v_t, \quad (3.72)$$

where the time index runs in multiples of four, that is, $t = 4, 8, 12, \dots$. Building on equations (3.71)–(3.72), it is trivial to obtain the corresponding model for a flow

variable by multiplying the observation equation (3.72) by $[\ I \ I \ I \ I]$. The resulting observation equation would then be:

$$z_t + z_{t-1} + z_{t-2} + z_{t-3} = [\ H \ I \ I \ I \ 0] \begin{bmatrix} x_t \\ z_{t-1} \\ z_{t-2} \\ z_{t-3} \\ z_{t-4} \end{bmatrix} + C v_t. \quad (3.73)$$

On the other hand, the model for a stock variable is obtained by multiplying equation (3.72) by $[\ I \ 0 \ 0 \ 0]$, resulting in

$$z_t = [\ H \ 0 \ 0 \ 0 \ 0] \begin{bmatrix} x_t \\ z_{t-1} \\ z_{t-2} \\ z_{t-3} \\ z_{t-4} \end{bmatrix} + C v_t. \quad (3.74)$$

The treatment for temporally aggregated data just outlined is adequate for many uses, but has a clear inconvenience: it increases the size of the state vector. In Chapter 10 we will see an efficient alternative which does not increase the size of the state vector and, therefore, is more stable and suitable for computationally-intensive tasks.

3.4 Uses of these transformations

The model transformations discussed in this chapter illustrate the flexibility of the SS representation, but their practical use may be unclear at first sight. In the rest of this book, and throughout the SS literature, we find many examples of their applicability, such as the following:

1. The endogeneization procedure described in Subsection 3.2.1 provides an immediate treatment for estimating a model with: (a) stochastic inputs, see Subsection 5.3.2, or (b) missing covariates in panel data models, see Chapter 11.
2. The deterministic-stochastic decomposition and the block-diagonal representation presented in subsections 3.1.1 and 3.1.3, respectively, are the foundations of the signal extraction procedure described in Chapter 8.
3. SS methods for data aggregated over space and time, see subsections 3.3.2 and 3.3.3, are the basis of many benchmarking and nowcasting procedures routinely applied by statistical agencies, see e.g., Durbin and Quenneville [83] or Nunes [169].
4. Finally, the observation equation transformations discussed in subsections 3.3.1 and 3.3.2 are the basis of the methods for modeling samples with missing values or observation errors, see e.g., Jones [132].

These are just a few examples which illustrate the flexibility of the SS model in allowing one to address sophisticated problems that are difficult to solve using standard statistical approaches.

This page intentionally left blank

Chapter 4

Filtering and Smoothing

4.1	The conditional moments of a State-Space model	33
4.2	The Kalman Filter	34
4.3	Decomposition of the smoothed moments	36
4.4	Smoothing for a general State-Space model	37
4.5	Smoothing for fixed-coefficients and single-error models	39
4.6	Uncertainty of the smoothed estimates in a SEM	40
4.7	Examples	42
4.7.1	Recursive Least Squares	42
4.7.2	Cleaning the Wolf sunspot series	42
4.7.3	Extracting the Hodrick-Prescott trend	44

In this chapter we present the main state estimation algorithms which are known as “filtering” and “fixed-interval smoothing.” Both focus on estimating the sequence of states, but are based on different information sets, and constitute the foundation of most SS procedures.

4.1 The conditional moments of a State-Space model

Consider an SS model, such as the general MEM (2.1)–(2.2) or the SEM (2.19)–(2.20) and denote:

$$\boldsymbol{x}_{t|j} = \mathbb{E}(\boldsymbol{x}_t | \Omega_j) \quad (4.1)$$

$$\boldsymbol{P}_{t|j} = \mathbb{E}[(\boldsymbol{x}_t - \boldsymbol{x}_{t|j})(\boldsymbol{x}_t - \boldsymbol{x}_{t|j})^T | \Omega_j], \quad (4.2)$$

where Ω_j is the information set containing all the input and output measures up to time j . With this notation, the sequences of filtered and fixed-interval smoothed states are, respectively¹:

$$\boldsymbol{x}_{t+1|t}, \quad \boldsymbol{P}_{t+1|t}, \quad t = 1, \dots, N, \quad (4.3)$$

¹In many works the results of filtering are denoted as: $\boldsymbol{x}_{t|t}$ and $\boldsymbol{P}_{t|t}$ because they use the alternative state equation $\boldsymbol{x}_t = \Phi \boldsymbol{x}_{t-1} + \Gamma \boldsymbol{u}_t + \boldsymbol{E} \boldsymbol{w}_t$.

$$\boldsymbol{x}_{t|N}, \quad \boldsymbol{P}_{t|N}, \quad t = 1, \dots, N. \quad (4.4)$$

Note that extending these sequences out of sample would provide forecasts for the state values, as well as for the corresponding error covariances,

$$\boldsymbol{x}_{t|N}, \quad \boldsymbol{P}_{t|N}, \quad t = N+1, N+2, \dots \quad (4.5)$$

Building on these results and the observation equation of the model, it is a straightforward exercise to compute forecasts and error covariances for the outputs of the SS model, \boldsymbol{z}_t . For example, assuming \boldsymbol{u}_{N+h} is observed (and non-stochastic), the relevant equations for h -step ahead forecasting in the case of MEM (2.1)–(2.2) would be:

$$\mathbb{E}(\boldsymbol{z}_{N+h} | \Omega_j) = \boldsymbol{H}\boldsymbol{x}_{N+h|N} + \boldsymbol{D}\boldsymbol{u}_{N+h},$$

$$\mathbb{E}[(\boldsymbol{z}_{N+h} - \boldsymbol{z}_{N+h|N})(\boldsymbol{z}_{N+h} - \boldsymbol{x}_{N+h|N})^T | \Omega_j] = \boldsymbol{H}\boldsymbol{P}_{N+h|N}\boldsymbol{H}^T + \boldsymbol{C}\boldsymbol{R}\boldsymbol{C}^T,$$

where $\boldsymbol{x}_{N+h|N}$ and $\boldsymbol{P}_{N+h|N}$ are obtained by iterative application (for $h = 1, 2, \dots$, etc.) of Kalman filter equations (4.6) and (4.9) given below, starting from a filtered state that has been evaluated up to time $t = N$.

Filtering and smoothing procedures are applied in this book with different goals in mind. For example, we use the Kalman filter to compute the innovations required to evaluate the likelihood for an SS model, see Terceiro [206] and Chapter 5, while smoothing algorithms are applied to estimate unobserved components in structural time-series models, as in Harvey [118], and to deal with samples containing missing values or aggregated data, see e.g., Ansley and Kohn [7], Kohn and Ansley [137] or Naranjo, Trindade and Casella [165].

4.2 The Kalman Filter

Building on an SS model in the form (2.1)–(2.2), the Kalman Filter (KF), see Kalman [134], computes one-step-ahead estimates of the states as well as the corresponding error covariance matrix by propagating the following equations:

$$\boldsymbol{x}_{t+1|t} = \boldsymbol{\Phi}\boldsymbol{x}_{t|t-1} + \boldsymbol{\Gamma}\boldsymbol{u}_t + \boldsymbol{K}_t \tilde{\boldsymbol{z}}_{t|t-1} \quad (4.6)$$

$$\tilde{\boldsymbol{z}}_{t|t-1} = \boldsymbol{z}_t - \boldsymbol{H}\boldsymbol{x}_{t|t-1} - \boldsymbol{D}\boldsymbol{u}_t \quad (4.7)$$

$$\boldsymbol{B}_t = \boldsymbol{H}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^T + \boldsymbol{C}\boldsymbol{R}\boldsymbol{C}^T \quad (4.8)$$

$$\boldsymbol{P}_{t+1|t} = \boldsymbol{\Phi}\boldsymbol{P}_{t|t-1}\boldsymbol{\Phi}^T + \boldsymbol{E}\boldsymbol{Q}\boldsymbol{E}^T - \boldsymbol{K}_t \boldsymbol{B}_t \boldsymbol{K}_t^T \quad (4.9)$$

$$\boldsymbol{K}_t = (\boldsymbol{\Phi}\boldsymbol{P}_{t|t-1}\boldsymbol{H}^T + \boldsymbol{E}\boldsymbol{S}\boldsymbol{C}^T)\boldsymbol{B}_t^{-1}, \quad (4.10)$$

see Appendix A, starting from suitable initial conditions, \boldsymbol{x}_1 and \boldsymbol{P}_1 , see e.g., De Jong and Chu-Chun-Lin [69], or Casals and Sotoca [43]. Note that:

1. The updating equation (4.6) transforms the *ex-ante* estimate for the state at time t , $\boldsymbol{x}_{t|t-1}$, in the *ex-post* value, $\boldsymbol{x}_{t+1|t}$, according to a linear function of the innovation, $\tilde{\boldsymbol{z}}_{t|t-1}$, so the larger the value of this forecast error, the larger the update will be.

2. Equation (4.7) computes the one-step-ahead forecast error for \mathbf{z}_t , also known as “innovation,” on the basis of the *ex-ante* state estimate, $\mathbf{x}_{t|t-1}$. The corresponding error covariance is given by the matrix \mathbf{B}_t in equation (4.8).
3. Expression (4.9) is known as the Riccati equation of the KF. It provides the covariance matrix for the sequence of KF states and, as we shall see later, it has a stationary solution under nonrestrictive assumptions. This fact plays an important role in many SS computational procedures.
4. The Riccati equation (4.9) does not depend on the sample values, \mathbf{z}_t . This fact has two interesting implications. First, no observation helps more than any other in decreasing uncertainty about the states. Second, equations (4.8)–(4.9) and (4.6)–(4.7) can be propagated independently.
5. Last, the matrix \mathbf{K}_t defined in equation (4.10) is known as the KF gain. Its appearance in the KF can be justified in many ways. First, it can be shown that the sequence (4.10) minimizes the trace of the *ex-post* covariance $\mathbf{P}_{t+1|t}$, see Jazwinsky [130] and Appendix A, as well as that of any linear combination of the variances in the main diagonal of $\mathbf{P}_{t+1|t}$, see Demetry [72]. Under normality, it provides the maximum likelihood estimates of the states.

Note that equation (4.7) can be trivially rewritten as

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_{t|t-1} + \mathbf{D} \mathbf{u}_t + \tilde{\mathbf{z}}_{t|t-1}.$$

Therefore, expressions (4.6) and this restructured (4.7) can be interpreted, respectively, as the state and observation equations of the time-varying parameter MEM:

$$\begin{aligned} \mathbf{x}_{t+1|t} &= \Phi \mathbf{x}_{t|t-1} + \Gamma \mathbf{u}_t + \mathbf{K}_t \tilde{\mathbf{z}}_{t|t-1} \\ \mathbf{z}_t &= \mathbf{H} \mathbf{x}_{t|t-1} + \mathbf{D} \mathbf{u}_t + \tilde{\mathbf{z}}_{t|t-1}. \end{aligned} \tag{4.11}$$

The errors in (4.11) are the innovations of the KF and, because of this, it is known in the literature as the “innovations representation,” see e.g. Harvey [118]. The difference between this model and the previously defined SEM, (2.19)–(2.20), is that the error distribution matrix in (4.11), \mathbf{K}_t , is time-varying, while the corresponding matrix in the SEM, \mathbf{K} , is constant. It can be shown that the matrix \mathbf{K} in the SEM is the steady-state solution for the sequence \mathbf{K}_t . Because of this, the SEM is also known in the literature as the “steady-state innovations model,” see e.g. Anderson and Moore [6].

In a stationary system with no inputs, adequate initial conditions are given by the steady-state solution of equations (4.6)–(4.10):

$$\mathbf{x}_1 = \mathbf{0} \tag{4.12}$$

$$\mathbf{P}_1 = \Phi \mathbf{P}_1 \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T, \tag{4.13}$$

where the Lyapunov equation (4.13) has a unique positive semi-definite solution, see Akaike [3] and Anderson and Moore [6]. When the model has exogenous inputs, the

initialization of the KF is more complex and needs to take into account the deterministic or stochastic structure of \mathbf{u}_t . We will discuss this issue further in Section 5.2.

Finally, when the system is nonstationary the uncertainty in the prior distribution of the initial state is infinite. The literature refers to this case as “diffuse initial conditions,” and it has different consequences on filtering and fixed-interval smoothing algorithms. First, it is impossible to initialize the standard KF with exact diffuse initial conditions². Therefore, in the nonstationary case one can adopt two basic strategies: (a) using the information filter, see Anderson and Moore [6], which propagates the system states and the inverse of their covariances so the exact diffuse initial conditions are $\mathbf{x}_1 = \mathbf{0}$ and $\mathbf{P}_1^{-1} = \mathbf{0}$; and (b) developing specialized algorithms to cope with nonstationary states for each specific application. An example of the latter approach is the likelihood evaluation procedure presented in Subsection 5.1.3.

Fixed-interval smoothing, on the other hand, allows for diffuse initial conditions. To this end, in Section 4.3 we will characterize the effect of initial conditions on the smoothed moments. This result will then be used for the exact smoother described in Section 4.4.

4.3 Decomposition of the smoothed moments

Consider the SS model:

$$\bar{\mathbf{x}}_{t+1} = \Phi \bar{\mathbf{x}}_t + \Gamma \mathbf{u}_t + \mathbf{E} \mathbf{w}_t \quad (4.14)$$

$$\bar{\mathbf{z}}_t = \mathbf{H} \bar{\mathbf{x}}_t + \mathbf{D} \mathbf{u}_t + \mathbf{C} \mathbf{v}_t, \quad (4.15)$$

where the states and observations correspond to model (2.1)–(2.2) with the initial conditions $\bar{\mathbf{x}}_1 = \mathbf{0}$ and $\bar{\mathbf{P}}_1 = \mathbf{0}$. Note that $\bar{\mathbf{z}}_t$ is not observed. Propagating state equations (2.1) and (4.14), it follows that:

$$\mathbf{x}_t = \Phi^{t-1} \mathbf{x}_1 + \bar{\mathbf{x}}_t, \quad (4.16)$$

where $\bar{\mathbf{x}}_t$ is independent of \mathbf{x}_1 . Thus the expectation of (4.16) conditional on the information set Ω_N is:

$$\mathbf{x}_{t|N} = \Phi^{t-1} \mathbf{x}_{1|N} + \bar{\mathbf{x}}_{t|N}. \quad (4.17)$$

Also, from (2.1)–(2.2) and (4.14)–(4.15) it is easy to prove, see Rosenberg [189], that:

$$\tilde{\mathbf{z}}_t = \mathbf{H} \bar{\Phi}_{t-1} \mathbf{x}_1 + \tilde{\mathbf{z}}_t, \quad (4.18)$$

where $\tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{z}_{t|t-1}$ is the sequence of KF innovations corresponding to (2.1)–(2.2). The values $\tilde{\mathbf{z}}_t$, defined accordingly by $\tilde{\mathbf{z}}_t = \bar{\mathbf{z}}_t - \bar{\mathbf{z}}_{t|t-1}$, are uncorrelated innovations, given that $\bar{\mathbf{x}}_1 = \mathbf{0}$ and $\bar{\mathbf{P}}_1 = \mathbf{0}$ are the proper initial conditions for

²The literature often suggests initializing the KF with $\mathbf{x}_1 = \mathbf{0}$ and $\mathbf{P}_1 = k \mathbf{I}$, where k is an arbitrary scalar whose value is “large enough” to approximate an infinite covariance matrix. As Ansley and Kohn [8] point out, such an initialization yields substantial estimation errors and numerical instability.

(4.14)–(4.15). Finally, the matrix sequence $\bar{\Phi}_t$ is defined recursively via $\bar{\Phi}_t = (\Phi - \mathbf{K}_t \mathbf{H}) \bar{\Phi}_{t-1}$ with $\bar{\Phi}_1 = \mathbf{I}$.

Equation (4.18) can be written for the entire sample as:

$$\tilde{\mathbf{z}} = \mathbf{X} \mathbf{x}_1 + \tilde{\mathbf{z}}, \quad (4.19)$$

where \mathbf{X} is the block-diagonal matrix whose t -th block is $\mathbf{H} \bar{\Phi}_{t-1}$, and the mN -dimensional vectors $\tilde{\mathbf{z}}$ and $\tilde{\mathbf{z}}$ contain the respective innovations resulting from a KF initialized with null values for the mean and covariance of the initial state, hereafter KF($\mathbf{0}, \mathbf{0}$). Note that this initialization assures that $\tilde{\mathbf{z}}$ is independent of \mathbf{x}_1 . The problem reduces then to obtaining the conditional expectation $\mathbf{x}_{t|N}$ on the right-hand side of (4.17), as well as its covariance, $\mathbf{P}_{t|N}$, taking into account (4.19). The following theorem shows how to achieve this.

Theorem 4.1 *The smoothed moments of the state vector can be written as:*

$$\mathbf{x}_{t|N} = [\Phi^{t-1} - E(\bar{\mathbf{x}}_t \tilde{\mathbf{z}}^T) \mathbf{B}^{-1} \mathbf{X}] \mathbf{x}_{1|N} + E(\bar{\mathbf{x}}_t \tilde{\mathbf{z}}^T) \mathbf{B}^{-1} \tilde{\mathbf{z}} \quad (4.20)$$

$$\mathbf{P}_{t|N} = [\Phi^{t-1} - E(\bar{\mathbf{x}}_t \tilde{\mathbf{z}}^T) \mathbf{B}^{-1} \mathbf{X}] \mathbf{P}_{1|N} [\Phi^{t-1} - E(\bar{\mathbf{x}}_t \tilde{\mathbf{z}}^T) \mathbf{B}^{-1} \mathbf{X}]^T + \bar{\mathbf{P}}_{t|N}, \quad (4.21)$$

where \mathbf{B} is a block-diagonal matrix that contains the covariance matrices of $\tilde{\mathbf{z}}_t$, and $\bar{\mathbf{P}}_{t|N}$ is the smoothed second-order moment of the states in (4.14)–(4.15).

A proof of this result can be found in the Appendix of Casals, Jerez, and Sotoca [40].

Note that equations (4.20)–(4.21) apply to both, stationary and nonstationary systems, as the only terms affected by the covariance of the initial state $\bar{\mathbf{P}}_1$ are $\mathbf{x}_{1|N}$ and $\mathbf{P}_{1|N}$, and this dependence occurs through $\bar{\mathbf{P}}_1^{-1}$, which is finite, see [66] and equations (A.1)–(A.2) in the Appendix of Casals, Jerez, and Sotoca [40]. The computation of $E(\bar{\mathbf{x}}_t \tilde{\mathbf{z}}^T) \mathbf{B}^{-1} \mathbf{X}$ and $E(\bar{\mathbf{x}}_t \tilde{\mathbf{z}}^T) \mathbf{B}^{-1} \tilde{\mathbf{z}}$ in (4.20)–(4.21) depends on the specific smoothing algorithm to be used. This issue is discussed further in the next Section.

4.4 Smoothing for a general State-Space model

Results (4.20)–(4.21) can be combined with any standard smoother to derive an algorithm which can be applied to nonstationary systems. In this section we will do this using an efficient and stable method due to De Jong [67].

The smoothing algorithm proposed by De Jong [67] first runs a KF($\mathbf{0}, \mathbf{0}$) through the entire sample and then performs a backward recursion given by:

$$\mathbf{x}_{t|N} = \mathbf{x}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{r}_{t-1} \quad (4.22)$$

$$\mathbf{P}_{t|N} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{R}_{t-1} \mathbf{P}_{t|t-1} \quad (4.23)$$

$$\mathbf{r}_{t-1} = \mathbf{H}^T \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t + \bar{\Phi}_t^T \mathbf{r}_t, \quad \text{with } \mathbf{r}_N = \mathbf{0} \quad (4.24)$$

$$\mathbf{R}_{t-1} = \mathbf{H}^T \mathbf{B}_t^{-1} \mathbf{H} + \bar{\Phi}_t^T \mathbf{R}_t \bar{\Phi}_t, \quad \text{with } \mathbf{R}_N = \mathbf{0} \quad (4.25)$$

$$\bar{\Phi}_t = \Phi - \mathbf{K}_t \mathbf{H}, \quad (4.26)$$

where $\mathbf{x}_{t|t-1}$ and $\mathbf{P}_{t|t-1}$ were computed in the forward step, \mathbf{B}_t is the t -th diagonal block of \mathbf{B} , and \mathbf{K}_t is the KF gain. In this framework, the terms corresponding to (4.20)–(4.21) are given by:

$$\mathrm{E}(\bar{\mathbf{x}}_t \tilde{\mathbf{z}}^T) \mathbf{B}^{-1} \tilde{\mathbf{z}} = \mathbf{x}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{r}_{t-1} \quad (4.27)$$

$$\bar{\mathbf{P}}_{t|N} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{R}_{t-1} \mathbf{P}_{t|t-1} \quad (4.28)$$

$$\Phi^{t-1} - \mathrm{E}(\bar{\mathbf{x}}_t \tilde{\mathbf{z}}^T) \mathbf{B}^{-1} \mathbf{X} = (\mathbf{I} - \mathbf{P}_{t|t-1} \mathbf{R}_{t-1}) \bar{\bar{\Phi}}_t, \quad (4.29)$$

and combining (4.22)–(4.26) with (4.27)–(4.29) yields the following algorithm:

Forward step : Propagate a KF(**0,0**) and the following equations:

$$\bar{\bar{\Phi}}_{t+1} = (\Phi - \mathbf{K}_t \mathbf{H}) \bar{\bar{\Phi}}_t, \quad \text{with } \bar{\bar{\Phi}}_1 = \mathbf{I} \quad (4.30)$$

$$\mathbf{X}_t = \mathbf{H} \bar{\bar{\Phi}}_t \quad (4.31)$$

$$\mathbf{W}_t = \mathbf{W}_{t-1} + \mathbf{X}_t^T \mathbf{B}_t^{-1} \mathbf{X}_t, \quad \text{with } \mathbf{W}_0 = \mathbf{0} \quad (4.32)$$

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \mathbf{X}_t^T \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t, \quad \text{with } \mathbf{w}_0 = \mathbf{0}. \quad (4.33)$$

Backward step : Propagate (4.24)–(4.26) and

$$\mathbf{V}_{t|N} = (\mathbf{I} - \mathbf{P}_{t|t-1} \mathbf{R}_{t-1}) \bar{\bar{\Phi}}_t \quad (4.34)$$

$$\mathbf{x}_{t|N} = \mathbf{x}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{r}_{t-1} + \mathbf{V}_{t|N} \mathbf{x}_{1|N} \quad (4.35)$$

$$\mathbf{P}_{t|N} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{R}_{t-1} \mathbf{P}_{t|t-1} + \mathbf{V}_{t|N} \mathbf{P}_{1|N} \mathbf{V}_{t|N}^T \quad (4.36)$$

where:

$$\mathbf{x}_{1|N} = \mathbf{P}_{1|N} \left(\bar{\mathbf{P}}_1^{-1} \bar{\mathbf{x}}_1 + \mathbf{w}_N \right) \quad (4.37)$$

$$\mathbf{P}_{1|N} = \left(\bar{\mathbf{P}}_1^{-1} + \mathbf{W}_N \right)^{-1}. \quad (4.38)$$

Comparing (4.35)–(4.36) with (4.22)–(4.23), we see immediately that the distortion due to the arbitrary (**0,0**) initialization is corrected by the terms $\mathbf{V}_{t|N} \mathbf{x}_{1|N}$ and $\mathbf{V}_{t|N} \mathbf{P}_{1|N} \mathbf{V}_{t|N}^T$. This algorithm could of course have been implemented with any initial conditions. Null values were chosen only because they save computing time by simplifying the forward KF recursion.

These results can be easily applied to solve any other signal extraction problem such as, e.g., fixed-point or fixed-interval smoothing, by modifying the conditioning information set in (4.20)–(4.21), and then combining the resulting equations with a standard algorithm.

4.5 Smoothing for fixed-coefficients and single-error models

To simplify notation, we have thus far assumed that the SS model is time-invariant. However, algorithm (4.30)–(4.38) does not require fixed parameters, and can therefore be generalized to time-varying systems just by adding a time subindex to the non-constant matrices. On the other hand, if the system is indeed time-invariant, further computational gains can be obtained by using a KF($\mathbf{0}, \bar{\mathbf{P}}$) instead of a KF($\mathbf{0}, \mathbf{0}$), where $\bar{\mathbf{P}}$ denotes the stationary solution of the Riccati equation in the KF. In this case, algorithm (4.30)–(4.38) simplifies to the following procedure:

Initial step : Compute $\bar{\mathbf{P}}$ such that:

$$\bar{\mathbf{P}} = \Phi \bar{\mathbf{P}} \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T - \bar{\mathbf{K}} \bar{\mathbf{B}} \bar{\mathbf{K}}^T, \quad (4.39)$$

where

$$\bar{\mathbf{B}} = \mathbf{H} \bar{\mathbf{P}} \mathbf{H}^T + \mathbf{C} \mathbf{R} \mathbf{C}^T \quad (4.40)$$

$$\bar{\mathbf{K}} = (\Phi \bar{\mathbf{P}} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}^T) \bar{\mathbf{B}}^{-1}. \quad (4.41)$$

If the system is detectable, (4.39) has a unique and positive semi-definite strong solution, see Theorem 3.1 and Lemma 3.1 in De Souza, Gevers, and Goodwin [71] and Appendix A.10.

Forward step: The KF($\mathbf{0}, \bar{\mathbf{P}}$) in the forward step simplifies to computing:

$$\mathbf{x}_{t+1|t} = \Phi \mathbf{x}_{t|t-1} + \Gamma \mathbf{u}_t + \bar{\mathbf{K}} \tilde{\mathbf{z}}_t \quad (4.42)$$

$$\tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{H} \mathbf{x}_{t|t-1} - \mathbf{D} \mathbf{u}_t, \quad (4.43)$$

as well as equations (4.30)–(4.33) with the substitutions $\mathbf{K}_t = \bar{\mathbf{K}}$ and $\mathbf{B}_t^{-1} = \bar{\mathbf{B}}^{-1}$.

Backward step: Propagate (4.24)–(4.26) with $\mathbf{K}_t = \bar{\mathbf{K}}$ and $\mathbf{B}_t^{-1} = \bar{\mathbf{B}}^{-1}$, as well as the expressions:

$$\begin{aligned} \mathbf{x}_{t|N} &= \mathbf{x}_{t|t-1} + \mathbf{P}_{t|t-1} \mathbf{r}_{t-1} + \mathbf{V}_{t|N} \left[(\bar{\mathbf{P}}_1 - \bar{\mathbf{P}})^{-1} + \mathbf{W}_N \right]^{-1} \\ &\quad \cdot \left[(\bar{\mathbf{P}}_1 - \bar{\mathbf{P}})^{-1} \bar{\mathbf{x}}_1 + \mathbf{w}_N \right] \end{aligned} \quad (4.44)$$

$$\mathbf{P}_{t|N} = \mathbf{P}_{t|t-1} - \mathbf{P}_{t|t-1} \mathbf{R}_{t-1} \mathbf{P}_{t|t-1} + \mathbf{V}_{t|N} \left[(\bar{\mathbf{P}}_1 - \bar{\mathbf{P}})^{-1} + \mathbf{W}_N \right]^{-1} \mathbf{V}_{t|N}^T, \quad (4.45)$$

instead of (4.35)–(4.36). Note that $(\bar{\mathbf{P}}_1 - \bar{\mathbf{P}})^{-1}$ is the inverse covariance for the initial state of the equivalent SEM derived in the initial step.

The computational efficiency of this recursion is due to a positive trade-off between the setup cost required to compute $\bar{\mathbf{P}}$ from (4.39)–(4.41), and the reduced computation and memory requirements in the forward and backward steps. This follows because of three facts: (a) propagating (4.42)–(4.43) requires less computation than the analogous KF($\mathbf{0}, \mathbf{0}$); (b) in the time-invariant case $E(\tilde{\mathbf{z}}_t \tilde{\mathbf{z}}_t^T) = \bar{\mathbf{B}}$ for all t ,

so this matrix needs to be inverted only once, and (c) since $\mathbf{K}_t = \bar{\mathbf{K}}$, $\mathbf{B}_t^{-1} = \bar{\mathbf{B}}^{-1}$, and $\mathbf{P}_{t|t-1} = \bar{\mathbf{P}}$, for all t , the computational load and memory requirements of the procedure are reduced.

Further gains can be obtained by writing the model in the equivalent SEM form, as shown in Chapter 2. In this case, the previous algorithm is further simplified because the solution of the Riccati equation is $\bar{\mathbf{P}} = \mathbf{0}$. In the next subsection we will discuss other advantages of this formulation.

4.6 Uncertainty of the smoothed estimates in a SEM

The following proposition defines an important property of the smoothed states of fixed-coefficients SEM:

Proposition 4.1 *In the SEM (2.19)–(2.20), if the eigenvalues of $\Phi - \mathbf{K} \mathbf{H}$ are either on or inside the unit circle, the smoothing covariance matrix $\mathbf{P}_{t|N}$ converges to zero as t increases. When all these eigenvalues are strictly inside the unit circle, this convergence is exponentially fast.*

Therefore, under the conditions of Proposition 4.1, the smoothed state estimates converge to exact values. (Note that these conditions are closely related to the invertibility of the SS model; see Definition 2.10.) The practical significance of this result is twofold. First, the smoothed states can be treated as actual data because they have null variances. Second, they will not be revised if the sample increases. This happens because in filtering and smoothing, revisions are proportional to the uncertainty affecting the estimates; therefore an estimate with null uncertainty will never be updated.

The important property embodied by Proposition 4.1 can be derived as a Corollary of Theorem 4.2 in De Souza, Gevers, and Goodwin [71], where detectability is shown to be a necessary and sufficient condition for convergence of the difference Riccati equation to the unique strong solution of the algebraic Riccati equation. For our purposes, the following proof is simpler and more direct.

Proof of Proposition 4.1. Ignoring exogenous inputs without loss of generality, equation (2.20) implies that $\mathbf{a}_t = \mathbf{z}_t - \mathbf{H}\mathbf{x}_t$. Substituting this expression in (2.19) yields

$$\mathbf{x}_{t+1} = (\Phi - \mathbf{K} \mathbf{H})\mathbf{x}_t + \mathbf{K} \mathbf{z}_t. \quad (4.46)$$

Therefore, and if the initial state \mathbf{x}_1 were known, the entire sequence of states would be exactly determined by (4.46), because the inputs on the right-hand side of this expression would also be known. However, usually \mathbf{x}_1 is unknown, and is treated as a random variable. In this case, the covariance matrix of the smoothed estimates (by definition conditional on the entire sample) would be

$$\mathbf{P}_{t|N} = (\Phi - \mathbf{K} \mathbf{H})^{t-1} \mathbf{P}_{1|N} [(\Phi - \mathbf{K} \mathbf{H})^{t-1}]^T, \quad (4.47)$$

where the eigenvalues of $\Phi - \mathbf{K} \mathbf{H}$ coincide with the reciprocals of the roots of the MA terms implied by the SS model (see Definition 2.10). If the model is strictly

invertible, then the eigenvalues of $\Phi - \mathbf{K} \mathbf{H}$ lie strictly inside the unit circle, and, therefore, $\mathbf{P}_{t|N}$ converges to zero at an exponential rate as t increases. On the other hand, if some eigenvalues of $\Phi - \mathbf{K} \mathbf{H}$ lie exactly on the unit circle, this convergence to exact values will still occur, but is not exponentially fast. ■

The convergence of $\mathbf{P}_{t|N}$ depends on the structure of the model. In particular:

1. When the model has a pure p -th order autoregressive structure, the trace of $\mathbf{P}_{t|N}$ collapses to zero after processing p observations. This property holds even when the process has unit AR roots, and means that the states in the AR-SEM can be uniquely determined from a given amount of information about the system outputs.
2. If the model has MA roots and all of them are invertible, then $\text{tr}(\mathbf{P}_{t|N})$ converges exponentially fast to zero as t increases. The decay rate of the coefficients in the corresponding AR representation governs this convergence, so an MA root close to the unit circle slows it. Note that this type of convergence also holds for models with strictly non-invertible MA roots, as they can be written in an equivalent strictly invertible form.
3. When there are MA roots on the unit circle, $\mathbf{P}_{t|N}$ converges to zero asymptotically. In this case $\mathbf{P}_{t|N}$ converges to a value which depends on the sample size, and tends to zero as the sample increases. Therefore, $\text{tr}(\mathbf{P}_{t|N})$ converges to zero as $t \rightarrow \infty$.

The following corollaries provide further insight into the implications of Proposition 4.1.

Corollary 4.1 *Under the conditions of Proposition 4.1, the KF covariances, $\mathbf{P}_{t|t-1}$, converge to zero.*

Proof. The proof is identical to that of Proposition 4.1, replacing the subindex N in (4.47) by $t-1$.

Corollary 4.2 *Under the conditions of Proposition 4.1, the smoothed and KF state estimates converge to the same values, i.e., $\|\mathbf{x}_{t|N} - \mathbf{x}_{t|t-1}\| \rightarrow 0$ as $t \rightarrow \infty$, where $\|\cdot\|$ denotes vector norm.*

Proof. Immediate from Proposition 4.1 and Corollary 4.1.

Therefore, one-sided KF estimates also converge to exact values (Corollary 4.1) and, consequently, to the smoothed estimates of the states (Corollary 4.2). For online applications, the recursive structure of the KF would be a clear advantage, in particular if the recursion is simplified in the light of Corollary 4.1. On the other hand, using filtering instead of smoothing implies some inefficiency, as $\mathbf{P}_{t|N}$ converges to zero faster than $\mathbf{P}_{t|t-1}$.

4.7 Examples

4.7.1 Recursive Least Squares

As we saw in Example 2.1.2, the constant-coefficient linear regression model could be formulated as:

$$\beta_{t+1} = \beta_t \quad (4.48)$$

$$z_t = u_t^T \beta_t + v_t \quad (4.49)$$

Identifying terms between (4.48)–(4.49) and the MEM (2.1)–(2.2) we see that $\Phi = I$, $\Gamma = 0$, $E = 0$, $H = u_t^T$, $D = 0$, and $C = I$, with $Q = 0$, $S = 0$, $R = \sigma_v^2$ and $x_t = \beta_t$. Now, applying the KF equations (4.7)–(4.10) to model (4.48)–(4.49) yields:

$$\tilde{z}_{t|t-1} = z_t - u_t^T \hat{\beta}_t \quad (4.50)$$

$$\hat{\beta}_{t+1} = \hat{\beta}_t + k_t \tilde{z}_{t|t-1} \quad (4.51)$$

$$P_{t+1|t} = P_{t|t-1} - k_t (u_t^T P_{t|t-1} u_t + \sigma_v^2) k_t^T \quad (4.52)$$

$$k_t = (P_{t|t-1} u_t^T) (u_t^T P_{t|t-1} u_t + \sigma_v^2)^{-1}, \quad (4.53)$$

which simplify trivially to the equations of the “Recursive Least Squares” estimator:

$$\hat{\beta}_{t+1} = \hat{\beta}_t + (P_{t|t-1} u_t^T) (u_t^T P_{t|t-1} u_t + \sigma_v^2)^{-1} (z_t - u_t^T \hat{\beta}_t) \quad (4.54)$$

$$P_{t+1|t} = P_{t|t-1} - P_{t|t-1} u_t^T (u_t^T P_{t|t-1} u_t + \sigma_v^2)^{-1} u_t P_{t|t-1}. \quad (4.55)$$

Equation (4.54) computes the *ex-post* estimate $\hat{\beta}_{t+1}$ as a linear combination of the *ex-ante* estimate $\hat{\beta}_t$ and the innovation $z_t - u_t^T \hat{\beta}_t$. Starting from suitable initial conditions, this equation can be run for each observation in the sample, thus allowing one to recursively estimate the vector of regression parameters. Meanwhile, note that (4.55) furnishes the covariances for $\hat{\beta}_{t+1}$. It is interesting to see that this equation is not coupled with (4.54).

Note that equations (4.54)–(4.55) are not computable, as they depend on the error variance σ_v^2 , which is most often unknown. This issue can be solved easily by the following changes of variables:

$$P_{t+1|t} = \frac{1}{\sigma_v^2} V_{t+1} \quad , \quad P_{t|t-1} = \frac{1}{\sigma_v^2} V_t, \quad (4.56)$$

which provide the computable form of the Recursive Least Squares estimator of β .

4.7.2 Cleaning the Wolf sunspot series

This example illustrates the use of the smoothing algorithm described in previous sections to filter out the observational noise from a variable in an error-in-variables model. To this end, consider Wolf’s annual series of sunspot numbers, from 1700 to 1988, taken from Tong [210]. This dataset draws on records compiled by human

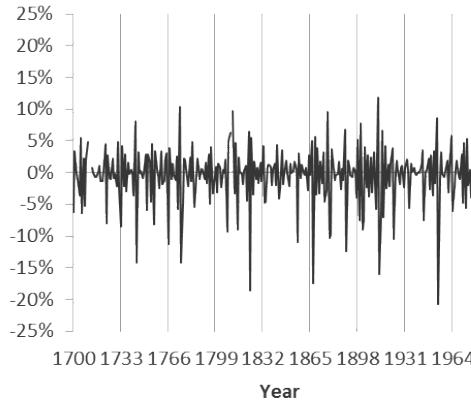


Figure 4.1 *Percent relative corrections on the original Wolf numbers: $(z_t^* / z_t - 1) \times 100$.*

observers using optical devices of varying quality, so it seems natural to assume that the data is affected by observational errors.

On the other hand, previous analyses have consistently found that this series displays a harmonic cycle with a period of about 11 years. Building on these ideas, we fitted the following AR(2) plus white noise errors model to the original data :

$$(1 - \phi_1 B - \phi_2 B^2) z_t^* = \mu + w_t, \quad \{w_t\} \sim \text{IID}(0, \sigma_w^2) \quad (4.57)$$

$$z_t = z_t^* + v_t, \quad \{v_t\} \sim \text{IID}(0, \sigma_v^2), \quad (4.58)$$

where the sequences z_t and z_t^* are, respectively, the square root of the Wolf number at year t and the underlying “decontaminated” value. Maximizing the Gaussian likelihood results in the estimates and standard errors displayed in Table 4.1. Note

Table 4.1 *Estimates and standard errors for the AR(2) plus noise model fitted to the Wolf sunspot series.*

Parameter	ϕ_1	ϕ_2	μ	σ_w^2	σ_v^2
estimate	1.444	-0.743	1.476	2.205	0.147
std. error	0.048	0.047	0.145	-	-

that the primary AR(2) structure has complex roots, implying a damped cycle with a period of 11.07 years. Using model (4.57)–(4.58) it is easy to obtain smoothed estimates for the underlying decontaminated series z_t^* . The percent relative corrections, $(z_t^* / z_t - 1) \times 100$, implied by this model are shown in Figure 4.1.

The code required to replicate this example can be found in the file Example0472.m, see Appendix D.

4.7.3 Extracting the Hodrick-Prescott trend

The potential Gross Domestic Product (GDP) is often estimated with the filter proposed by Hodrick and Prescott [126], which is equivalent to the smoothed trend resulting from the model:

$$\mu_{t+1} = \mu_t + \beta_t \quad (4.59)$$

$$\beta_{t+1} = \beta_t + \zeta_t \quad (4.60)$$

$$z_t = \mu_t + \varepsilon_t \quad (4.61)$$

$$\begin{bmatrix} \zeta_t \\ \varepsilon_t \end{bmatrix} \sim \text{IID} \left\{ \begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_\zeta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix} \right\}, \quad (4.62)$$

with a signal-to-noise variance ratio $\sigma_\zeta^2/\sigma_\varepsilon^2 = 1/1600$ for quarterly data, see Harvey and Trimbur [120]. A typical analysis in this framework consists of estimating the unknown variances, taking into account the variance ratio constraint, and then obtaining smoothed estimates of the trend component μ_t .

This procedure was applied to the quarterly and seasonally adjusted series of US GDP (GDP_t) in constant 2000 US Dollar, from 1947/Q1 to 2008/Q3. We fitted model (4.59)–(4.62) to the transformed variable $z_t = \log(GDP_t) \times 100$, and obtained the variance estimates of $\hat{\sigma}_\varepsilon^2 = 1.875$ and $\hat{\sigma}_\zeta^2 = 1.875/1600$. Figure 4.2 shows the estimates of the change in HP trend (the unobserved component β_t). This series is often used to pinpoint turning points in the economic cycle.

The code required to replicate this example can be found in the file `Example0473.m`, see Appendix D.

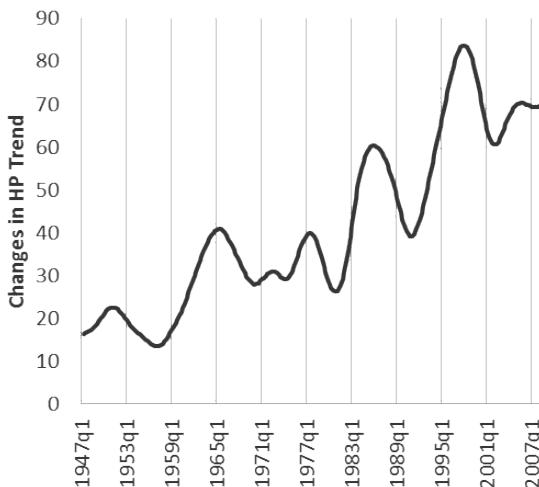


Figure 4.2 *Changes in the HP trend.*

Chapter 5

Likelihood computation for fixed-coefficients models

5.1	Maximum likelihood estimation	45
5.1.1	Problem statement	45
5.1.2	Prediction error decomposition	46
5.1.3	Initialization of the Kalman filter in the stationary case	47
5.2	The likelihood for a nonstationary model	48
5.2.1	Diffuse likelihood	49
5.2.2	Minimally conditioned likelihood	50
5.2.3	Likelihood computation for a fixed-parameter SEM	52
5.2.4	Initialization of the Kalman filter in the nonstationary case	52
5.3	The likelihood for a model with inputs	54
5.3.1	Models with deterministic inputs	54
5.3.2	Models with stochastic inputs	55
5.4	Examples	57
5.4.1	Models for the Airline Passenger series	57
5.4.2	Modeling the series of Housing Starts and Sales	60

In this chapter we discuss how to compute the Gaussian likelihood for a linear SS model. As we will see, the presence of unit roots and exogenous inputs critically affects its evaluation.

5.1 Maximum likelihood estimation

5.1.1 Problem statement

Consider a MEM without inputs:

$$\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \mathbf{E} w_t \quad (5.1)$$

$$\mathbf{z}_t = \mathbf{H}\mathbf{x}_t + \mathbf{C}\mathbf{v}_t, \quad (5.2)$$

with corresponding error covariances:

$$\mathrm{E} \left(\begin{bmatrix} \mathbf{w}_t \\ \mathbf{v}_t \end{bmatrix} \begin{bmatrix} \mathbf{w}_t & \mathbf{v}_t \end{bmatrix} \right) = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix}. \quad (5.3)$$

Our goal is to estimate a parameter vector $\boldsymbol{\theta}$ which includes all the unknown elements within the matrices $\Phi, \mathbf{E}, \mathbf{H}, \mathbf{C}, \mathbf{Q}, \mathbf{R}$ and \mathbf{S} . Assuming that the sample information is $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$, with N being the sample size, the maximum likelihood (ML) estimator of $\boldsymbol{\theta}$ is

$$\hat{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} [P(\mathbf{Z}; \boldsymbol{\theta})], \quad (5.4)$$

where $P(\mathbf{Z}; \boldsymbol{\theta})$ is the joint probability density function of the observations. If \mathbf{Z} is treated as given, this is the so-called *likelihood function*. The estimates obtained by maximizing this function have many convenient asymptotic properties, which are discussed in Appendix B.

In this framework, an ML estimation method will include two basic elements:

1. A procedure to evaluate the likelihood function for any tentative value of the parameter vector $\boldsymbol{\theta}$. In this chapter we will discuss how to do this in the case of a fixed-coefficients SS model. Chapter 6 covers the likelihood evaluation for some specific time-varying parameter models.
2. An algorithm to compute the ML estimate $\boldsymbol{\theta}$. This can be done with many methods, most of them being variations on the basic Newton-Raphson algorithm, see e.g., Dennis and Schnabel [74], requiring the first and second-order derivatives of the log-likelihood function. These can be computed either analytically, see appendices B to D in Terceiro [206], or by numerical approximation, see Dennis and Schnabel [74]. A popular method in the statistics literature, because it obviates the evaluation of derivatives, is the so-called *EM Algorithm*, which will be discussed in detail in Chapter 11.

Bearing in mind that the numerical optimization problem is substantial, but has been satisfactorily resolved in the literature, we will focus our attention on discussing how to evaluate the Gaussian likelihood for a SS model. To this end, the basic idea is to write the likelihood function in a specific form known as “prediction error decomposition.”

5.1.2 Prediction error decomposition

When the observations are not independent, one can write $P(\mathbf{Z}; \boldsymbol{\theta})$ by applying successively the definition of conditional density:

$$P(\mathbf{Z}; \boldsymbol{\theta}) = \prod_{t=1}^N P(\mathbf{z}_t | \mathbf{Z}^{t-1}; \boldsymbol{\theta}), \quad (5.5)$$

where $\mathbf{Z}^{t-1} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_{t-1})$. It is convenient to simplify the mathematical structure of (5.5) by taking logs so, instead of maximizing (5.5) with respect to $\boldsymbol{\theta}$, we will

consider the equivalent problem of minimizing the negative log-likelihood:

$$\ell(\mathbf{Z}; \boldsymbol{\theta}) = -\log P(\mathbf{Z}; \boldsymbol{\theta}) = -\sum_{t=1}^N \log P(\mathbf{z}_t | \mathbf{Z}^{t-1}; \boldsymbol{\theta}). \quad (5.6)$$

and, assuming normality for \mathbf{w}_t , \mathbf{v}_t and \mathbf{x}_1 , the general term $P(\mathbf{z}_t | \mathbf{Z}^{t-1}; \boldsymbol{\theta})$ can be computed by applying a KF to (5.1)–(5.2) for any given value of $\boldsymbol{\theta}$. In particular, the negative Gaussian log-likelihood (5.6) can be written as:

$$\ell(\mathbf{Z}; \boldsymbol{\theta}) = \frac{1}{2} \sum_{t=1}^N (m \log(2\pi) + \log |\mathbf{B}_t| + \tilde{\mathbf{z}}_t^T \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t), \quad (5.7)$$

where the sequence of innovations is defined as: $\tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbb{E}(\mathbf{z}_t | \mathbf{Z}^{t-1}; \boldsymbol{\theta})$ and $\mathbf{B}_t = \text{cov}(\tilde{\mathbf{z}}_t | \mathbf{Z}^{t-1}; \boldsymbol{\theta})$. The function (5.7) is known as the “prediction error decomposition form” of the log-likelihood, see Schweppe [192]. The reason for this terminology stems from the fact that $\tilde{\mathbf{z}}_t$ can be interpreted as a vector of one-step-ahead prediction errors which, under normality, follows the distribution, $\tilde{\mathbf{z}}_t \sim NID(\mathbf{0}, \mathbf{B}_t)$. The term *NID* will henceforth denote *Normal and Independently Distributed*.

Evaluation of the log-likelihood (5.7) requires computation of the sequences $\tilde{\mathbf{z}}_t$ and \mathbf{B}_t , obtained by propagating the KF :

$$\hat{\mathbf{x}}_{t+1|t} = \Phi \hat{\mathbf{x}}_{t|t-1} + \mathbf{K}_t \tilde{\mathbf{z}}_t \quad (5.8)$$

$$\tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{H} \hat{\mathbf{x}}_{t|t-1} \quad (5.9)$$

$$\mathbf{B}_t = \mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{C} \mathbf{R} \mathbf{C}^T \quad (5.10)$$

$$\mathbf{K}_t = (\Phi \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}^T) \mathbf{B}_t^{-1} \quad (5.11)$$

$$\mathbf{P}_{t+1|t} = \Phi \mathbf{P}_{t|t-1} \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T - \mathbf{K}_t \mathbf{B}_t \mathbf{K}_t^T. \quad (5.12)$$

Note that expressions (5.8)–(5.12) can be immediately generalized to the time-varying parameter case just by adding time subindex t to the system matrices Φ , \mathbf{E} , \mathbf{H} , \mathbf{C} , \mathbf{Q} , \mathbf{R} and \mathbf{S} . This is useful, for example, in estimating random coefficients models or periodic models, see Chapter 6. Given any value for the parameter vector $\boldsymbol{\theta}$, it is relatively easy to: (a) generate the corresponding SS matrices, (b) propagate the KF (5.8)–(5.12), and (c) use the resulting outputs to evaluate (5.7).

5.1.3 Initialization of the Kalman filter in the stationary case

The KF (5.8)–(5.12) is a recursive algorithm which requires initial conditions for the state vector and its covariance matrix. Determining these conditions is a complex problem which has been tackled in many works such as: Akaike [3], Anderson and Moore [6], De Jong[66], Ansley and Kohn [8], or De Jong and Chu-Chun-Lin [69]. We will now address this problem assuming that the state equation (5.1) is stable, see Definition 2.1. Under this assumption, adequate initial conditions for the KF are $\boldsymbol{\mu}$ and \mathbf{P}_1 such that:

$$\boldsymbol{\mu} = \mathbb{E}(\mathbf{x}_1) = \mathbf{0} \quad (5.13)$$

$$\mathbf{P}_1 = \Phi \mathbf{P}_1 \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T, \quad (5.14)$$

where the algebraic Lyapunov equation (5.14) has a unique positive semi-definite solution, see Akaike [3], Anderson and Moore [6] and Golub and Van Loan [109]. To solve it, we need to rewrite (5.14) as a system of linear equations:

$$\text{vec}(\mathbf{P}_1) = (\Phi \otimes \Phi) \text{vec}(\mathbf{P}_1) + \text{vec}(\mathbf{E} \mathbf{Q} \mathbf{E}^T), \quad (5.15)$$

where \otimes denotes the Kronecker product and $\text{vec}(\cdot)$ is the *vectorization* operator, which stacks the columns of its argument into a single vector. The solution of (5.15) is then¹:

$$\text{vec}(\mathbf{P}_1) = (\mathbf{I} - \Phi \otimes \Phi)^{-1} \text{vec}(\mathbf{E} \mathbf{Q} \mathbf{E}^T). \quad (5.16)$$

Note that the eigenvalues of $\mathbf{I} - \Phi \otimes \Phi$ are $1 - \lambda_i \lambda_j$, for all $i, j = 1, 2, \dots, n$, where λ_i is the i -th eigenvalue of Φ . Therefore, the stability assumption ensures that equation (5.15) has a unique solution. Moreover, since the matrix $\mathbf{E} \mathbf{Q} \mathbf{E}^T$ is positive semidefinite, then \mathbf{P}_1 is positive semi-definite in general. If the system is controllable², \mathbf{P}_1 would be strictly positive definite.

5.2 The likelihood for a nonstationary model

In the nonstationary case at least one eigenvalue of Φ in (5.1) has a unit modulus. Accordingly the system is not stable and, under the immemorial time hypothesis (see Definition 2.9 in Chapter 2), the uncertainty associated with the initial state is not finite. As a consequence, the initialization of the KF becomes more complex.

Many papers deal with this problem which is generically described as “diffuse initialization” see De Jong [68], Ansley and Kohn [9], Koopman and Durbin [142] and Koopman [139]. As we saw in Section 4.2, this situation could be tackled in general by initializing the KF with an arbitrarily large covariance, or by using the information filter, see Anderson and Moore [6]. Ansley and Kohn [8] point out that the former method induces numerical errors, and the latter cannot be applied to all cases. Alternatively, one could develop specialized algorithms for each specific application. In this line, the literature provides three main approaches to evaluate the likelihood for a model with nonstationary states:

1. Ansley and Kohn [8] propose a data transformation which cancels the nonstationary states. This approach has been further developed in other works, such as Kohn and Ansley [137] or Ansley and Kohn [9] and, in the univariate case, in Bell and Hillmer [19] or Gomez and Maravall [110].
2. Koopman [139] derives an exact solution to initialize a KF with diffuse initial conditions by: (a) setting $\mu = \mathbf{0}$ and $\mathbf{P}_1 = \mathbf{P}_* + k \mathbf{P}_\infty$ ($k \rightarrow \infty$), where the matrices \mathbf{P}_* and \mathbf{P}_∞ are known, and (b) transforming the KF variables so that they do not depend on k .

¹There are computationally efficient procedures that do not require solving a system of dimension n^2 such as the one in (5.15). The basic idea consists of using a real Schur decomposition of Φ that computes \mathbf{P}_1 by solving a system of dimension 1×1 or 2×2 , see , Christov and Konstantinov [181] and Appendix A.

²See Chapter 2, Definition 2.2

3. De Jong, [66, 68], suggest initializing the KF with $\mu = \mathbf{0}$ and $P_1 = \mathbf{0}$, and then adjusting the log-likelihood to compensate for these arbitrary initial conditions.

These methods have specific problems and limitations. In particular, Ansley and Kohn [8] recognize that method 1 needs a complex and nonstandard filtering, and requires the data transformation to be independent of the parameter values. This requirement is not fulfilled, for example, when one wants to estimate structural time series models such as those described by Harvey [118]. On the other hand, methods 2 and 3 have two problems: (a) these methods do not allow for stochastic inputs, and (b) its likelihood depends in some cases on the scale of the state vector. In Subsection 5.2.1 we provide further details about this issue.

Subsection 5.2.1 describes the diffuse likelihood method 3 in detail. Subsection 5.2.2 presents the *minimally conditioned likelihood*, a procedure which draws many ideas from the diffuse likelihood but avoids its scaling problem. Last, Subsection 5.3.1 discusses the effect of exogenous inputs on the calculation of the likelihood.

5.2.1 Diffuse likelihood

The diffuse likelihood can be written as:

$$\ell_\infty(\mathbf{Z}; \boldsymbol{\theta}) = \ell(\mathbf{Z}; \boldsymbol{\theta}) - \log |P_1|, \quad (5.17)$$

where $\ell_\infty(\mathbf{Z}; \boldsymbol{\theta})$ denotes the diffuse log-likelihood of model (5.1)–(5.2) and $\ell(\mathbf{Z}; \boldsymbol{\theta})$ is the corresponding standard log-likelihood, see De Jong [68]. On this basis, De Jong [68] proposes an evaluation algorithm based on the so-called diffuse KF, while Koopman [139] suggests using two specialized filters for the diffuse and non-diffuse components, respectively.

The first term on the right-hand side of (5.17), $\ell(\mathbf{Z}; \boldsymbol{\theta})$, can be computed following De Jong [66], who assumes that the initial state is such that $\mathbf{x}_1 \sim N(\boldsymbol{\mu}, P_1)$. Under these conditions:

$$\begin{aligned} \ell(\mathbf{Z}; \boldsymbol{\theta}) &= \log |P_1| + \boldsymbol{\mu}^T P_1^{-1} \boldsymbol{\mu} + \sum_{t=1}^N \log |\mathbf{B}_t| + \sum_{t=1}^N \tilde{\mathbf{z}}_t^T \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t + \log |P_1^{-1} + \mathbf{W}_N| \\ &\quad - (\mathbf{P}_1^{-1} \boldsymbol{\mu} + \mathbf{w}_N)^T (\mathbf{P}_1^{-1} + \mathbf{W}_N)^{-1} (\mathbf{P}_1^{-1} \boldsymbol{\mu} + \mathbf{w}_N), \end{aligned} \quad (5.18)$$

where $\tilde{\mathbf{z}}_t$ and \mathbf{B}_t are, respectively, the innovations and their covariance matrices resulting from a KF($\mathbf{0}, \mathbf{0}$) applied to (5.1)–(5.2). On the other hand, the values of \mathbf{w}_N and \mathbf{W}_N in (5.18) can be computed as follows:

$$\mathbf{w}_t = \mathbf{w}_{t-1} + \bar{\Phi}_{t-1}^T \mathbf{H}^T \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t \quad (5.19)$$

$$\mathbf{W}_t = \mathbf{W}_{t-1} + \bar{\Phi}_{t-1}^T \mathbf{H}^T \mathbf{B}_t^{-1} \mathbf{H} \bar{\Phi}_{t-1} \quad (5.20)$$

$$\bar{\Phi}_t = (\boldsymbol{\Phi} - \mathbf{K}_t \mathbf{H}) \bar{\Phi}_{t-1}, \quad (5.21)$$

where \mathbf{w}_t and \mathbf{W}_t should be initialized with null values, $\bar{\Phi}_0 = \mathbf{I}$, and \mathbf{K}_t is the gain of a KF($\mathbf{0}, \mathbf{0}$) applied to (5.1)–(5.2). Note that:

1. The terms $\log|\mathbf{P}_1|$, $\boldsymbol{\mu}^T \mathbf{P}_1^{-1} \boldsymbol{\mu}$, $\log|\mathbf{P}_1^{-1} + \mathbf{W}_N|$, $\mathbf{P}_1^{-1} \boldsymbol{\mu} + \mathbf{w}_N$, and $(\mathbf{P}_1^{-1} + \mathbf{W}_N)^{-1}$ in (5.18) depend on the mean and covariance of the initial state, $\boldsymbol{\mu}$ and \mathbf{P}_1 . When the system is stationary, all of them can be computed.
2. When there are some unit roots, the first term in (5.18) diverges to infinity, whence it is ignored for computing the diffuse likelihood. The other terms are not problematic because they depend on \mathbf{P}_1^{-1} which is finite³.
3. The diffuse approach has a clear shortcoming, as the likelihood value depends on the scale of the state vector in some models. The following example illustrates this situation.

Example 5.2.1 Scale-dependency of the diffuse likelihood

Consider the observationally equivalent models:

$$\begin{aligned} x_{t+1} &= x_t + w_t \\ z_t &= \alpha x_t + v_t, \end{aligned} \tag{5.22}$$

and

$$\begin{aligned} x_{t+1}^* &= x_t^* + \alpha w_t \\ z_t &= x_t^* + v_t, \end{aligned} \tag{5.23}$$

where α is an arbitrary constant and the system needs some identification constraints such as $\text{var}(w_t) = 1$ and $\text{cov}(w_t, v_t) = 0$. According to (5.17), the diffuse likelihood for models (5.22) and (5.23) are, respectively:

$$\ell_{\infty}^1(\mathbf{Z}; \alpha) = \ell(\mathbf{Z}; \boldsymbol{\theta}) - \frac{1}{2} \log |\text{cov}(\mathbf{x}_1)| \tag{5.24}$$

$$\ell_{\infty}^2(\mathbf{Z}; \alpha) = \ell(\mathbf{Z}; \boldsymbol{\theta}) - \frac{1}{2} \log |\text{cov}(\alpha \mathbf{x}_1)|, \tag{5.25}$$

and these values do not coincide because $\ell_{\infty}^1(\mathbf{Z}; \alpha) - \ell_{\infty}^2(\mathbf{Z}; \alpha) = \log(\alpha)$. Therefore, the values of the diffuse likelihood corresponding to observationally equivalent representations, such as (5.22) and (5.23), can be different. Note also that the first-order derivatives of the likelihood are also scale-dependent. The procedure described in the following subsection solves these problems.

5.2.2 Minimally conditioned likelihood

An alternative to the diffuse likelihood would be a Gaussian likelihood conditional on the minimum subsample required to eliminate the effect of the diffuse states. This strategy is closely related to the diffuse likelihood, but does not suffer

³If all the modes in MEM (5.1)–(5.2) were unstable, \mathbf{P}_1^{-1} would be null. On the other hand, if the model has some stable modes, the matrix \mathbf{P}_1^{-1} converges to a nonzero value, see De Jong and Chu-Chun-Lin [69].

from its shortcomings. Following Casals, Sotoca and Jerez [47], the diffuse and minimally conditioned likelihoods are related by the following expression:

$$\ell^*(\mathbf{Z}; \boldsymbol{\theta}) = \ell_\infty(\mathbf{Z}; \boldsymbol{\theta}) - \log |\mathbf{O}_1^T \mathbf{O}_1|, \quad (5.26)$$

where $\ell^*(\mathbf{Z}; \boldsymbol{\theta})$ denotes the minimally conditioned log-likelihood and $\ell_\infty(\mathbf{Z}; \boldsymbol{\theta})$ is the diffuse function. The difference between both functions is therefore the term $-\log |\mathbf{O}_1^T \mathbf{O}_1|$ which plays a very important role, as it frees the minimally conditioned likelihood from any scale dependency. To compute this correction term, one needs to:

1. Write observation equation (5.2) in matrix form:

$$\mathbf{Z} = \mathbf{O}\mathbf{x}_1 + \mathbf{Z}^*, \quad (5.27)$$

where \mathbf{Z}^* is the part of the sample which does not depend on \mathbf{x}_1 , and \mathbf{O} is the extended observability matrix, defined as:

$$\mathbf{O} = \begin{bmatrix} \mathbf{H} \\ \mathbf{H}\Phi \\ \vdots \\ \mathbf{H}\Phi^{n-1} \end{bmatrix}. \quad (5.28)$$

2. Separate the diffuse and non-diffuse components of the initial state by means of the similar transformation:

$$\mathbf{T}\mathbf{x}_1 = \begin{bmatrix} \mathbf{x}_1^N \\ \mathbf{x}_1^S \end{bmatrix}, \quad (5.29)$$

where \mathbf{T} is the matrix characterizing the transformation such that $\mathbf{T}\Phi\mathbf{T}^{-1} = \text{diag}(\Phi^N, \Phi^S)$, \mathbf{x}_1^N is a vector including the diffuse states, such that $\text{cov}(\mathbf{x}_1^N) \rightarrow \infty$, and \mathbf{x}_1^S is a vector of stationary components. Denoting $\mathbf{T}^{-1} = [\mathbf{M} \ \mathbf{N}]$, we can write equation (5.29) as:

$$\mathbf{x}_1 = \mathbf{M}\mathbf{x}_1^N + \mathbf{N}\mathbf{x}_1^S \quad (5.30)$$

3. Apply the decomposition (5.30) to (5.27), which yields:

$$\mathbf{Z} = \mathbf{O}_N \mathbf{x}_1^N + \mathbf{Z}^S, \quad (5.31)$$

where $\mathbf{O}_N = \mathbf{O}\mathbf{M}$ is the extended observability matrix corresponding to the diffuse initial states and \mathbf{Z}^S is the part of the sample which does not depend on the diffuse states.

4. Compute the correction term $-\log |\mathbf{O}_1^T \mathbf{O}_1|$, where \mathbf{O}_1 includes the first columns of \mathbf{O}_N so that $\text{rank}(\mathbf{O}_N) = k$, where k is the number of diffuse states in the system. This specific choice implies that the likelihood computation is conditioned on the first k observations that can reduce the covariance matrix of \mathbf{x}_1^N (conditional on the first k observations) to a finite value.

Casals, Sotoca and Jerez [47] proved that the minimally conditioned likelihood: (a) can be efficiently computed by two equivalent methods which, in comparison with other approaches, (b) only requires standard filtering, so it is able to accommodate missing data and cointegration constraints. Last, (c) the choice made for the correction term $-\log |\mathbf{O}_1^T \mathbf{O}_1|$ assures that the values provided by the minimally conditioned likelihood coincide with those that would be obtained by differencing the data and then computing the likelihood for the stationary model; a desirable consistency property.

5.2.3 Likelihood computation for a fixed-parameter SEM

Likelihood evaluation can be substantially simplified when the model is a constant-parameter SEM. First, in an SEM $\mathbf{C} = \mathbf{I}$ and there is only one error covariance, \mathbf{Q} , so in this case the log-likelihood function (5.18) simplifies to:

$$\ell(\mathbf{Z}; \boldsymbol{\theta}) = \log |\mathbf{P}_1| + \boldsymbol{\mu}^T \mathbf{P}_1^{-1} \boldsymbol{\mu} + N \log |\mathbf{Q}| + \sum_{t=1}^N \tilde{\mathbf{z}}_t^T \mathbf{Q}^{-1} \tilde{\mathbf{z}}_t + \log |\mathbf{P}_1^{-1} + \mathbf{W}_N| - (\mathbf{P}_1^{-1} \boldsymbol{\mu} + \mathbf{w}_N)^T (\mathbf{P}_1^{-1} + \mathbf{W}_N)^{-1} (\mathbf{P}_1^{-1} \boldsymbol{\mu} + \mathbf{w}_N). \quad (5.32)$$

The KF($\mathbf{0}, \mathbf{0}$) thus simplifies drastically, because $\mathbf{P} = \mathbf{0}$ is a solution of the Riccati equation (5.12), which implies that $\mathbf{P}_{t+1|t} = \mathbf{0}$, $\mathbf{B}_t = \mathbf{Q}$, and $\mathbf{K}_t = \mathbf{E}$. As these KF variables have exact and constant values which do not need to be updated, the resulting likelihood computation algorithm is faster and more stable than those previously described, see Casals, Sotoca and Jerez [45].

5.2.4 Initialization of the Kalman filter in the nonstationary case

To evaluate (5.18) we need to determine \mathbf{P}_1^{-1} . To this end, we will separate the stable and non-stable modes by applying a similar transformation to model (5.1)–(5.2):

$$\begin{bmatrix} \mathbf{x}_{t+1}^N \\ \mathbf{x}_{t+1}^S \end{bmatrix} = \mathbf{T} \mathbf{x}_t, \quad (5.33)$$

such that the state equation is decomposed in two blocks:

$$\begin{bmatrix} \mathbf{x}_{t+1}^N \\ \mathbf{x}_{t+1}^S \end{bmatrix} = \begin{bmatrix} \Phi^N & \mathbf{0} \\ \mathbf{0} & \Phi^S \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^N \\ \mathbf{x}_t^S \end{bmatrix} + \begin{bmatrix} \mathbf{E}^N \\ \mathbf{E}^S \end{bmatrix} \mathbf{w}_t, \quad (5.34)$$

where the vectors \mathbf{x}_{t+1}^N and \mathbf{x}_{t+1}^S correspond, respectively, to the nonstationary and stationary modes of the system. Note that this transformation is a particular case of the block-diagonalization procedure presented in Subsection 3.1.3 and Appendix A. De Jong and Chu-Chun-Lin [69] originally suggested using the Jordan decomposition to obtain (5.34). In our opinion, the procedures discussed in Subsection 3.1.3 are more efficient and stable.

Assume now that the model has been initialized with arbitrary values. The covariance of the state vector can then be propagated according to the following equations:

$$\mathbf{P}_{t+1}^N = \Phi^N \mathbf{P}_t^N (\Phi^N)^T + \mathbf{E}^N \mathbf{Q} (\mathbf{E}^N)^T \quad (5.35)$$

$$\mathbf{P}_{t+1}^S = \Phi^S \mathbf{P}_t^S (\Phi^S)^T + \mathbf{E}^S \mathbf{Q} (\mathbf{E}^S)^T \quad (5.36)$$

$$\mathbf{P}_{t+1}^{NS} = \Phi^N \mathbf{P}_t^{NS} (\Phi^N)^T + \mathbf{E}^N \mathbf{Q} (\mathbf{E}^N)^T, \quad (5.37)$$

where \mathbf{P}_t^N , \mathbf{P}_t^S , and \mathbf{P}_t^{NS} are, respectively, the covariances of the nonstationary and stationary states, as well as the cross-covariance between both types of states. As the eigenvalues of the nonstationary subsystem are equal to one, the variances of all its modes diverge to infinity, so that $\lim_{t \rightarrow \infty} \mathbf{P}_t^N = \infty$ and, accordingly, $\lim_{t \rightarrow \infty} (\mathbf{P}_t^N)^{-1} = 0$. On the other hand, \mathbf{P}_t^S and \mathbf{P}_t^{NS} converge to the following equilibrium values:

$$\mathbf{P}^S = \Phi^S \mathbf{P}^S (\Phi^S)^T + \mathbf{E}^S \mathbf{Q} (\mathbf{E}^S)^T \quad (5.38)$$

$$\mathbf{P}^{NS} = \Phi^N \mathbf{P}^{NS} (\Phi^N)^T + \mathbf{E}^N \mathbf{Q} (\mathbf{E}^N)^T. \quad (5.39)$$

In their otherwise excellent paper, De Jong and Chu-Chun-Lin [69] assumed that the covariance between the stationary and nonstationary states converges to zero. To see that this is not true, apply the vectorization operator to (5.39):

$$\text{vec}(\mathbf{P}^{NS}) = (\mathbf{I} - \Phi^S \otimes \Phi^N)^{-1} \text{vec}[\mathbf{E}^N \mathbf{Q} (\mathbf{E}^S)^T]. \quad (5.40)$$

Taking into account that the eigenvalues of $\mathbf{I} - \Phi^S \otimes \Phi^N$ are $1 - \lambda_i^N \lambda_j^S$, where λ_i^N and λ_j^S are the i -th and j -th eigenvalues of Φ^N and Φ^S , respectively, and bearing in mind that there are no reciprocal eigenvalues in these matrices, the value \mathbf{P}^{NS} is the unique solution of (5.40), see Section A.7. Therefore, the following convergence result holds:

$$\lim_{t \rightarrow \infty} \text{cov} \left(\begin{bmatrix} \mathbf{x}_t^N \\ \mathbf{x}_t^S \end{bmatrix} \right) = \begin{bmatrix} \infty & (\mathbf{P}^{NS})^T \\ \mathbf{P}^{NS} & \mathbf{P}^S \end{bmatrix}. \quad (5.41)$$

Applying the matrix block-inversion lemma to equation (5.41), we obtain:

$$\lim_{t \rightarrow \infty} \left(\text{cov} \left(\begin{bmatrix} \mathbf{x}_t^N \\ \mathbf{x}_t^S \end{bmatrix} \right) \right)^{-1} = \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & (\mathbf{P}^S)^{-1} \end{bmatrix}, \quad (5.42)$$

and using $\mathbf{T}^{-1} = [\mathbf{M} \ \mathbf{N}]$ and $[\mathbf{M} \ \mathbf{N}]^{-1} = \begin{bmatrix} \mathbf{U} \\ \mathbf{V} \end{bmatrix}$ we have:

$$\lim_{t \rightarrow \infty} [\text{cov}(\mathbf{x}_t)]^{-1} = \lim_{t \rightarrow \infty} \text{cov} \left(\mathbf{T}^{-1} \begin{bmatrix} \mathbf{x}_t^N \\ \mathbf{x}_t^S \end{bmatrix} (\mathbf{T}^{-1})^T \right)^{-1} = \mathbf{V}^T (\mathbf{P}^S)^{-1} \mathbf{V}, \quad (5.43)$$

so that, under the immemorial time hypothesis, expression (5.43) provides an adequate initial condition for $[\text{cov}(\mathbf{x}_t)]^{-1}$.

On the other hand, the state vector should be initialized to null values. This choice is adequate because: (a) in a system without inputs the vector of stationary states (\mathbf{x}_t^S) converges to zero, and (b) the variances of the nonstationary states (\mathbf{x}_t^N) diverge to infinity, so that the particular choice made for their initial values is irrelevant. Thus, and since any initial condition is adequate for these states, the simplest choice would be $\boldsymbol{\mu} = \mathbf{0}$.

5.3 The likelihood for a model with inputs

Consider now a MEM with exogenous inputs:

$$\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{E} \mathbf{w}_t \quad (5.44)$$

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_t + \mathbf{D} \mathbf{u}_t + \mathbf{C} \mathbf{v}_t, \quad (5.45)$$

where the unknown elements in Γ and \mathbf{D} are added to the vector $\boldsymbol{\theta}$ to be estimated, and the sample information is now given by $\mathbf{Z} = (\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$ and $\mathbf{U} = (\mathbf{u}_1, \mathbf{u}_2, \dots, \mathbf{u}_N)$. Under these conditions, the log-likelihood function in (5.6) becomes:

$$\ell(\boldsymbol{\theta}) = -\log P(\mathbf{Z}, \mathbf{U}; \boldsymbol{\theta}) = -\sum_{t=1}^N \log P(\mathbf{z}_t | \mathbf{Z}^{t-1}, \mathbf{U}; \boldsymbol{\theta}). \quad (5.46)$$

Evaluating (5.46) requires propagating a KF analogous to (5.8)–(5.12), but replacing (5.8) and (5.9) by the following expressions:

$$\hat{\mathbf{x}}_{t+1|t} = \Phi \hat{\mathbf{x}}_{t|t-1} + \Gamma \mathbf{u}_t + \mathbf{K}_t \tilde{\mathbf{z}}_t \quad (5.47)$$

$$\tilde{\mathbf{z}}_t = \mathbf{z}_t - \mathbf{H} \hat{\mathbf{x}}_{t|t-1} - \mathbf{D} \mathbf{u}_t. \quad (5.48)$$

The likelihood function can be written in a form analogous to (5.18) using the following result:

$$\ell(\mathbf{Z} | \mathbf{U}) = \ell(\mathbf{x}_1 | \mathbf{U}) + \ell(\mathbf{Z} | \mathbf{U}, \mathbf{x}_1) - \ell(\mathbf{x}_1 | \mathbf{Z}, \mathbf{U}) \quad (5.49)$$

and, under normality, evaluation of (5.49) depends crucially on the values of $\boldsymbol{\mu} = \mathbf{E}(\mathbf{x}_1 | \mathbf{U})$ and $\mathbf{P}_1 = \text{cov}(\mathbf{x}_1 | \mathbf{U})$. Following De Jong [66], minus twice the log-likelihood (discarding constants) is then:

$$\begin{aligned} \ell(\mathbf{Z}, \mathbf{U}; \boldsymbol{\theta}) &= \log |\mathbf{P}_1| + \boldsymbol{\mu}^T \mathbf{P}_1^{-1} \boldsymbol{\mu} + \sum_{t=1}^N \log |\mathbf{B}_t| + \sum_{t=1}^N \tilde{\mathbf{z}}_t^T \mathbf{B}_t^{-1} \tilde{\mathbf{z}}_t + \log |\mathbf{P}_1^{-1} + \mathbf{W}_N| \\ &\quad - (\mathbf{P}_1^{-1} \boldsymbol{\mu} + \mathbf{w}_N)^T (\mathbf{P}_1^{-1} + \mathbf{W}_N)^{-1} (\mathbf{P}_1^{-1} \boldsymbol{\mu} + \mathbf{w}_N), \end{aligned} \quad (5.50)$$

where $\tilde{\mathbf{z}}_t$ and \mathbf{B}_t are the innovations and their covariances resulting from a KF($\mathbf{0}, \mathbf{0}$) given by (5.47)–(5.48) and (5.10)–(5.12). Again, the values of \mathbf{w}_N and \mathbf{W}_N in (5.50) can be computed via (5.19)–(5.21).

5.3.1 Models with deterministic inputs

The literature typically assumes that the inputs are: (a) exactly known within the sample period, (b) exogenous to the system, and (c) deterministic, which affects the expected value of state vector, but not its covariance matrix. Under these assumptions, the initial conditions for the KF depend on the stationarity of the system.

Stationary systems. If the inputs are constant and known, a stable and adequate initialization for the state vector in the stationary case would be given by:

$$\boldsymbol{\mu} = (\mathbf{I} - \boldsymbol{\Phi})^{-1} \boldsymbol{\Gamma} \mathbf{u}_1 \quad (5.51)$$

$$\mathbf{P}_1 = \boldsymbol{\Phi} \mathbf{P}_1 \boldsymbol{\Phi}^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T, \quad (5.52)$$

where we assume that $\mathbf{u}_1 = \mathbf{u}_0 = \mathbf{u}_{-1} \dots$

Nonstationary systems. When the model has inputs we cannot apply directly the concepts of stationarity and stability defined in Chapter 2. This is because a stable system could have nonstationary inputs which would then realize nonstationary states. Bearing in mind this disclaimer, we will decompose the state equation (5.44) into the corresponding “stationary” and “nonstationary” subsystems as in (5.34):

$$\begin{bmatrix} \mathbf{x}_{t+1}^N \\ \mathbf{x}_{t+1}^S \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Phi}^N & \mathbf{0} \\ \mathbf{0} & \boldsymbol{\Phi}^S \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^N \\ \mathbf{x}_t^S \end{bmatrix} + \begin{bmatrix} \boldsymbol{\Gamma}^N \\ \boldsymbol{\Gamma}^S \end{bmatrix} \mathbf{u}_t + \begin{bmatrix} \mathbf{E}^N \\ \mathbf{E}^S \end{bmatrix} \mathbf{w}_t. \quad (5.53)$$

To avoid discontinuities, we will also assume that the pair $(\boldsymbol{\Phi}^N, \mathbf{E}^N)$ is controllable. This is enough to avoid a situation where some modes of the nonstationary subsystem have a persistent excitation from \mathbf{u}_t , thus causing them to diverge to infinity. Under these assumptions, and keeping in mind that the initial state vector has been transformed according to:

$$\mathbf{T} \mathbf{x}_1 = \begin{bmatrix} \mathbf{x}_1^N \\ \mathbf{x}_1^S \end{bmatrix}, \quad (5.54)$$

we can multiply the expected value of the initial state by $\mathbf{T}^{-1} = [\mathbf{M} \ \mathbf{N}]$, obtaining:

$$\boldsymbol{\mu} = \mathbf{M} \mathbf{E}(\mathbf{x}_1^N) + \mathbf{N} \mathbf{E}(\mathbf{x}_1^S). \quad (5.55)$$

Under (5.53) and the immemorial time hypothesis, the value of $\mathbf{E}(\mathbf{x}_1^N)$ is again irrelevant, and so we can take it to be null. On the other hand, the initial condition for the stationary states is $\mathbf{E}(\mathbf{x}_1^S) = (\mathbf{I} - \boldsymbol{\Phi}^S)^{-1} \boldsymbol{\Gamma}^S \mathbf{u}_1$. Accordingly, expression (5.55) is equivalent to:

$$\boldsymbol{\mu} = \mathbf{N} \mathbf{E}(\mathbf{x}_1^S) = \mathbf{N} (\mathbf{I} - \boldsymbol{\Phi}^S)^{-1} \boldsymbol{\Gamma}^S \mathbf{u}_1, \quad (5.56)$$

which determines the desired initial condition.

5.3.2 Models with stochastic inputs

In many cases the exogenous variables should be regarded as stochastic. The literature (mostly in engineering) does not pay enough attention to this issue, probably due to the fact that in most physical systems inputs can be safely assumed to be either deterministic or controllable. However, it is very relevant in other situations such as, e.g., when modeling an economic time series.

The key for a proper initialization of the KF with stochastic inputs consists of decomposing the state equation (5.44) into its deterministic and stochastic subsystems, as shown in Subsection 3.1.1. After doing so, the state equations (3.5) and (3.7) can be jointly written as:

$$\begin{bmatrix} \mathbf{x}_{t+1}^d \\ \mathbf{x}_{t+1}^s \end{bmatrix} = \begin{bmatrix} \Phi & \mathbf{0} \\ \mathbf{0} & \Phi \end{bmatrix} \begin{bmatrix} \mathbf{x}_t^d \\ \mathbf{x}_t^s \end{bmatrix} + \begin{bmatrix} \Gamma \\ \mathbf{0} \end{bmatrix} \mathbf{u}_t + \begin{bmatrix} \mathbf{0} \\ \mathbf{E} \end{bmatrix} \mathbf{w}_t, \quad (5.57)$$

and the problem reduces then to initializing both subsystems. Initialization for the stochastic system was discussed in Sections 5.1 and 5.2. As for the deterministic subsystem, its initial conditions depend on whether the dynamic structure of the inputs is known.

The dynamic structure of the inputs is known. In this case, a proper initialization would require determining $\mu = E(\mathbf{x}_1 | \mathbf{U})$ and $P_1 = \text{cov}(\mathbf{x}_1 | \mathbf{U})$. These exact expressions were derived by Casals and Sotoca [43] under the immemorial time hypothesis⁴.

The mean and covariance of the initial state vector conditional on the inputs have different expressions depending on the particular stochastic structure of the inputs. For example, if the inputs follow a white noise process, it is easy to see that $\mu = E(\mathbf{x}_1 | \mathbf{U}) = \mathbf{0}$ and $P_1 = \text{cov}(\mathbf{x}_1 | \mathbf{U}) = \Phi P_1 \Phi^T + \mathbf{E} Q \mathbf{E}^T + \Gamma \Sigma_a \Gamma^T$, where Σ_a is the covariance matrix of the white noise.

The dynamic structure of the inputs is unknown. The previous framework can be extended to the case when the model for the inputs is unknown. The resulting initial conditions are approximate but, on the other hand, their computational cost is smaller than that of the exact method. To this end, we use again the decomposition (5.57) assuming that the states of both subsystems are independent, that is, $\text{cov}(\mathbf{x}_t^s, \mathbf{x}_t^d) = \mathbf{0}$ for all t , see Van Overschee and De Moor [215]⁵. The observation equations corresponding to the states in (5.57) are:

$$\mathbf{z}_t^d = \mathbf{H} \mathbf{x}_t^d + \mathbf{D} \mathbf{u}_t \quad (5.58)$$

$$\mathbf{z}_t^s = \mathbf{H} \mathbf{x}_t^s + \mathbf{v}_t. \quad (5.59)$$

From (5.57), (5.58) and (5.59) it is immediate that:

$$\mu = \mu^d + \mu^s \quad (5.60)$$

$$P_1 = P_1^d + P_1^s, \quad (5.61)$$

where $\mu = E(\mathbf{x}_1 | \mathbf{U})$, $\mu^d = E(\mathbf{x}_1^d | \mathbf{U})$, and $\mu^s = E(\mathbf{x}_1^s | \mathbf{U})$. Then, the problem reduces to estimating μ^d , μ^s , P_1^d , and P_1^s .

⁴In the nonstationary case this procedure may yield indeterminate results such as, e.g., $(0)(\infty)$. However, this is irrelevant for likelihood computation because (5.18) does not depend on these conditional moments, but on P_1^{-1} , $P_1^{-1}\mu$, and $\mu^T P_1^{-1}\mu$ instead, and these terms always converge to finite and easy to compute values.

⁵This hypothesis is not too restrictive, as the independence between \mathbf{x}_1 and \mathbf{w}_t implies that $\text{cov}(\mathbf{x}_1^s, \mathbf{x}_1^d) = \mathbf{0}$ is a sufficient condition for $\text{cov}(\mathbf{x}_t^s, \mathbf{x}_t^d) = \mathbf{0}$ for all t .

For the stochastic system, the immemorial time argument assures that an adequate initialization is $\mu^s = \mathbf{0}$, because $E(\mathbf{x}_1^s | \mathbf{U}) = E(\mathbf{x}_1^s)$. The solution for \mathbf{P}_1^s is the same as described for models without inputs (see Subsections 5.1.3 and 5.2.4). In the deterministic subsystem (5.58), we assume that the initial state is a fixed and unknown value with $\mathbf{P}_1^d = \mathbf{0}$. ML estimates of μ^d can be computed as $\mu^d = \mathbf{W}_N^{-1} \mathbf{w}_N$, where \mathbf{W}_N and \mathbf{w}_N are by-products of the propagation of (5.19)–(5.21). In this case, the likelihood (5.50) can be concentrated with respect to μ^d and $\mathbf{P}_1^d = \mathbf{0}$.

Therefore, the approximation consists in: (a) computing μ^s using the immemorial time argument, (b) computing μ^d by ML, and (c) using (5.60) and (5.61) to calculate the approximate initial conditions for the whole system. The resulting initial conditions are again different, depending on the stationary or nonstationary nature of the model.

1. If the model is stationary, both the deterministic and stochastic subsystems are stationary. Hence $\mathbf{M} = \mathbf{0}$, see Expression (5.55), $\Phi^S = \Phi$, and the approximation to the initial conditions is given by:

$$\mu = \mathbf{W}_N^{-1} \mathbf{w}_N \quad (5.62)$$

$$\mathbf{P}_1 = \Phi \mathbf{P}_1 \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T. \quad (5.63)$$

Note that expressions (5.62) and (5.63) do not depend on the parameters of the model for the inputs.

2. When the model is nonstationary, the conditional mean of the initial state is $\mu = \mathbf{W}_N^{-1} \mathbf{w}_N$, and the inverse of the corresponding covariance is $\mathbf{P}_1^{-1} = \mathbf{0}$. This result coincides with the initial conditions obtained by De Jong and Chu-Chun-Lin [69] for the same case. If $\mathbf{P}_1^{-1} = \mathbf{0}$, the value of μ is irrelevant.
3. The initial conditions for a partially nonstationary model are $\mu = \mathbf{W}_N^{-1} \mathbf{w}_N$ and $\mathbf{P}_1^{-1} = \mathbf{V}^T (\mathbf{P}^S)^{-1} \mathbf{V}$, where \mathbf{P}^S is calculated according to the Lyapunov equation given by (5.38), and \mathbf{V} is defined in (5.43).

The key idea in deriving suitable initial conditions then, consists in computing the decomposition (5.57)–(5.59) followed by an application of the immemorial time hypothesis. On the other hand, the correct initialization of a model with stochastic inputs consists of obtaining ML estimates of the initial state for the system.

5.4 Examples

5.4.1 Models for the Airline Passenger series

Examples 5.4.1 and 5.4.2 are based on the famous monthly time series of international airline passengers, which is shown in Figure 5.1.

Example 5.4.1 Estimating the Airline model in stationary and nonstationary form

This example emphasizes the estimation of seasonal ARIMA models and illustrates the fact that SS methods allow one to obtain identical estimates for a model

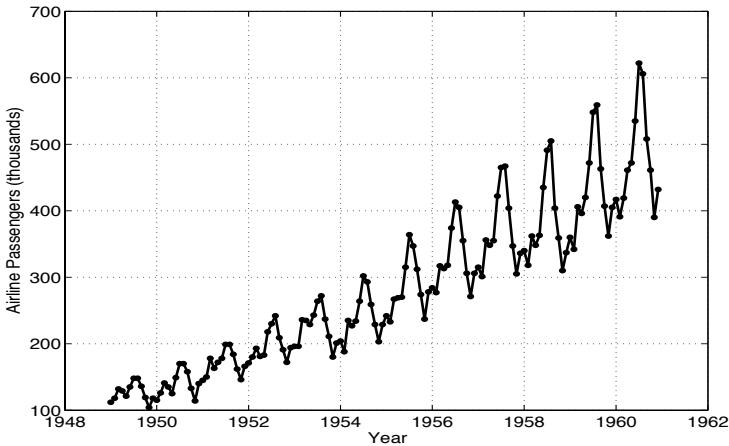


Figure 5.1 *International Airline Passengers (thousands) from 1949 to 1960*. Source: see Box, Jenkins and Reinsel [30].

in stationary and nonstationary forms. While both representations are strictly equivalent, the latter has clear advantages for some specific applications, which will be discussed later.

As is well known, the (log) airline series is adequately represented by an $IMA(1,1) \times (1,1)_{12}$ model:

$$(1 - B)(1 - B^{12}) \log y_t = (1 - \theta B)(1 - \Theta B^{12}) a_t \quad (5.64)$$

where y_t denotes the number of airline passengers in month t . On the other hand, if we denote $z_t = (1 - B)(1 - B^{12}) \log y_t$, the nonstationary model (5.64) could be written in the equivalent stationary $MA(1) \times MA_{12}$ form:

$$z_t = (1 - \theta B)(1 - \Theta B^{12}) a_t \quad (5.65)$$

where the parameters and errors in both representations should be the same.

Table 5.1 summarizes the main estimation results obtained for both representations. The table reports a single set of values because the figures obtained for both forms are identical up to the fourth decimal place. This high level of numerical consistency is only assured when the nonstationary model is estimated by optimizing the minimally conditioned likelihood described in Subsection 5.2.2. Maximizing the likelihood evaluated according to the diffuse filters of De Jong [68] or Koopman [139] (see Subsection 5.2.1) often produces estimates which differ from those of the stationary model. The code required to replicate this example can be found in the file Example0541.m, see Appendix D.

Example 5.4.2 A transfer function model with calendar effects and missing values

Table 5.1 *Estimation results for the airline models (5.64) and (5.65).*

Parameter	Estimate	Standard error
$\hat{\theta}$	0.4018	0.0766
$\hat{\Theta}$	0.5569	0.0738
$\hat{\sigma}_a$	0.0367	—
Log-likelihood	-244.6965	—

The previous example showed that the same basic model can be estimated in both stationary and nonstationary forms, providing results which are close to identical if the likelihood computation procedure is suitably chosen. On the other hand, it is natural to ask: why should we take on the additional cost in terms of complexity and computational burden to estimate a model in nonstationary form? The answer to this question is that this representation is more convenient for many purposes. Without loss of generality, we will discuss this idea using the equivalent models (5.64) and (5.65).

Assume for example that the goal of the modeling exercise is to forecast the nonstationary variable y_t . In this case, estimating the nonstationary form (5.64) is more convenient, as it immediately provides point and interval forecasts for $\log y_t$, and hence y_t . In comparison, the stationary form (5.65) provides forecasts for $z_t = (1 - B)(1 - B^{12}) \log y_t$, requiring a more complex treatment to recover the original metrics.

Another situation where the nonstationary form is more practical is when one wants to perform the trend-cycle-seasonal decomposition of y_t , see Chapter 8. In this case, the nonstationary form (5.64) can be immediately decomposed into the additive components of $\log y_t$ whose anti-logs provide the corresponding multiplicative decomposition. Once again, the same results could be obtained on the basis of the stationary formulation (5.65) by means of a more complex procedure.

A third case where the nonstationary representation has clear advantages is when the sample contains missing values. They could arise, either by a discontinuity in the sampling, or as an effective treatment for additive impulse-type outliers see Gomez, Maravall and Peña [111]. In this case one should never difference the data because differencing propagates the missing values, thus destroying valuable information. Therefore, working with the nonstationary model is a must in this case.

To illustrate these ideas we will fit the following transfer function to the airline series:

$$\begin{aligned} \log y_t &= \omega_0^L L_t + \omega_0^W W_t + \omega_0^E E_t + N_t \\ (1 - B)(1 - B^{12})N_t &= (1 - \theta B)(1 - \Theta B^{12}) a_t, \end{aligned} \quad (5.66)$$

where L_t is the number of labor days (Monday to Friday), W_t , is the number of weekend days (Saturday and Sunday) in a month, E_t , is a dummy variable whose value is one if Easter occurs in month t and zero otherwise, and N_t denotes an autocorrelated error term which follows an $IMA(1, 1) \times (1, 1)_{12}$ process. Therefore, the explanatory

variables in this model are devised to capture calendar effects, see Cleveland and Devlin [54].

Estimating model (5.66) by ML provides the following results:

$$\begin{aligned}\log y_t &= .039L_t + .049W_t + .028E_t + \hat{N}_t \\ (1 - B)(1 - B^{12})\hat{N}_t &= (1 - .222B)(1 - .533B^{12})\hat{a}_t \quad , \quad \hat{\sigma}_a = .033\end{aligned}\quad (5.67)$$

and Figure 5.2 displays the residuals \hat{a}_t .

Building on these results we detected three additive outliers, in observations number 29, 54 and 62. The standard treatment for this situation consists of adding three impulse-type intervention variables to the specification; see Box and Tiao [31]. According to the previous discussion, we re-estimated the model (5.66) treating these three observations as missing values. This provided the following estimation results:

$$\begin{aligned}\log y_t^* &= .034L_t + .044W_t + .023E_t + \hat{N}_t \\ (1 - B)(1 - B^{12})\hat{N}_t &= (1 - .082B)(1 - .484B^{12})\hat{a}_t^* \quad , \quad \hat{\sigma}_{a^*} = .029,\end{aligned}\quad (5.68)$$

where the “asterisk” (*) symbol denotes the variable with missing values. Note that the residuals of this model, displayed in Figure 5.3, have fewer outliers than those in Figure 5.2, so this atypical intervention procedure has been effective.

The code required to replicate this example can be found in the file `Example0542.m`, see Appendix D.

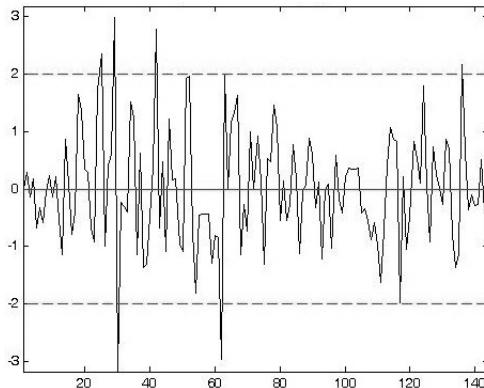


Figure 5.2 *Standardized residuals of the transfer function (5.67).*

5.4.2 Modeling the series of Housing Starts and Sales

Examples 5.4.3 and 5.4.4 are based on the monthly U.S. single-family housing starts (ST_t) and sales (SS_t) series, see Figure 5.4. This dataset has been analyzed in many works, such as those of Hillmer and Tiao [124], Ahn and Reinsel [1] or

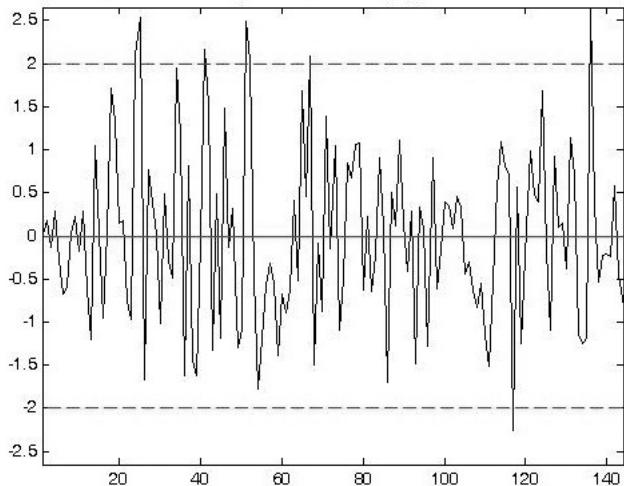


Figure 5.3 *Standardized residuals of the transfer function (5.68).*

Tiao [208]. Some stylized facts found in previous analyses are: (a) both series are adequately represented by an airline model (b) their seasonal structure is probably deterministic, and (c) they are cointegrated.

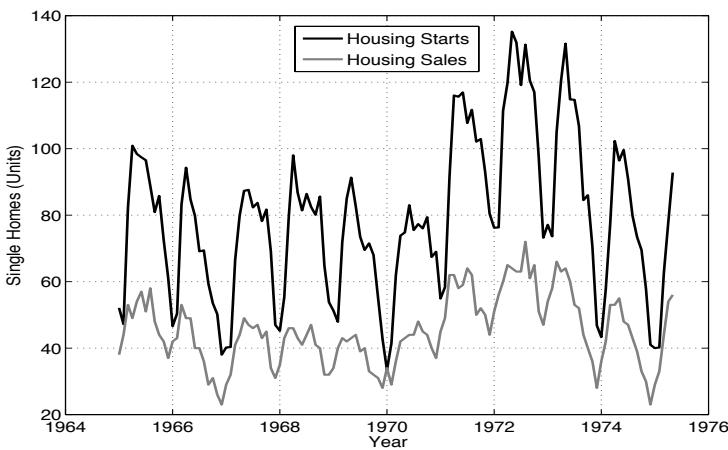


Figure 5.4 *Monthly U.S. single-family Housing Starts and Housing Sales, in thousands of units, from January 1965 through May 1975.*

Example 5.4.3 Univariate modeling

According to these stylized facts and our own univariate analysis, we fitted the following initial models:

$$\begin{aligned} \log ST_t &= .016L_t + \hat{N}_t \\ (1 - B)(1 - B^{12})\hat{N}_t &= (1 - .259B)(1 - 1.000B^{12})\hat{a}_t^{ST} \\ \hat{\sigma}_{ST} &= .081 \quad , \quad Q(39) = 37.50, \end{aligned} \tag{5.69}$$

and

$$\begin{aligned} (1 - B)(1 - B^{12}) \log SA_t &= (1 - 1.000B^{12})\hat{a}_t^{SA} \\ \hat{\sigma}_{SA} &= .078 \quad , \quad Q(39) = 45.62, \end{aligned} \tag{5.70}$$

where $Q(39)$ denotes the Ljung and Box [151] portmanteau statistic for the null that the model errors are white noise, computed with the first 39 sample auto-correlations.

Note that the housing starts series depends on the number of work days (Monday to Friday) in each month, while the housing sales data does not display any significant calendar effect.

Another relevant feature of models (5.69) and (5.70) is that the seasonal MA parameter is unity in both cases, so the corresponding polynomial has twelve roots on the unit circle. This is a frequent result in two cases: first, when a seasonal difference is applied to a time series which does not need it; in this case the misspecified seasonal difference tends to be compensated by an equivalent seasonal MA polynomial; second, a unity seasonal MA parameter may also imply that the seasonal structure is deterministic. To illustrate this idea, consider the following monthly process:

$$\gamma_t = \sum_{i=1}^{12} \beta_i D_t^i + a_t, \tag{5.71}$$

where γ_t is a purely seasonal variable, β_i is the monthly fluctuation in month i , and D_t^i is a seasonal dummy variable, which value is 1 when the period t corresponds to month i and zero otherwise. Obviously, model (5.71) implies that the seasonal fluctuation corresponding to a given month is always the same. Applying a seasonal difference to both sides of (5.71) we obtain:

$$(1 - B^{12})\gamma_t = (1 - B^{12})a_t, \tag{5.72}$$

so a seasonal MA root on the unit circle is also an indication of a deterministic seasonal structure in the time series. To illustrate these ideas, Figure 5.5 displays the seasonal components implied by models (5.69) and (5.70), extracted by the methods explained in Chapter 8. Note that the seasonal fluctuation is exactly the same in different years, meaning that seasonality for this series is deterministic. If the seasonal MA term in models (5.69) and (5.70) were invertible, seasonality would be stochastic and, accordingly, the seasonal fluctuation would change over the sample.

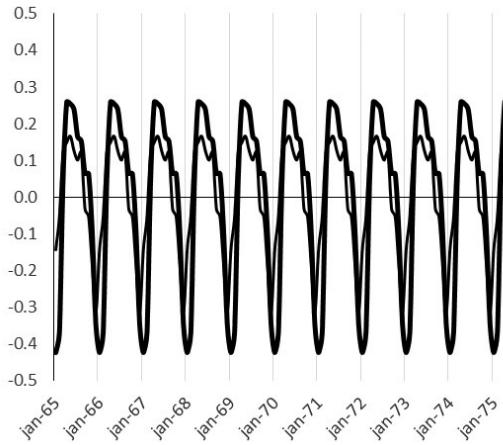


Figure 5.5 Additive seasonal component of (log) Housing Starts (thick line) and Housing Sales (thin line), from January 1965 through May 1975.

The code required to replicate this example can be found in the file Example0543.m, see Appendix D.

Example 5.4.4 A nonstandard VARMAX model with a cointegration constraint and calendar effects

As we said before, previous analyses often found the above series to be cointegrated, meaning that there is a long-term equilibrium between the housing starts and sales. The results in the previous example help us to visualize this equilibrium. Figure 5.6 shows the profile of both series (in logs) after adjusting for seasonality and calendar effects. Note that they follow each other closely which is a clear (if somewhat informal) evidence of cointegration.

On the basis of the previous analysis, and after a standard multivariate specification analysis, see Jenkins and Alavi [131], we fitted the following model to these series:

$$\begin{bmatrix} 1 & -1.028 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \log ST_t \\ \log SA_t \end{bmatrix} = \begin{bmatrix} .016 \\ (.006) \\ 0 \end{bmatrix} L_t + \begin{bmatrix} \hat{N}_t^{ST} \\ \hat{N}_t^{SA} \end{bmatrix} \quad (5.73)$$

$$\begin{aligned} & \begin{bmatrix} 1 - .213B & 0 \\ 0 & 1 - B \end{bmatrix} \begin{bmatrix} 1 - B^{12} & 0 \\ 0 & 1 - B^{12} \end{bmatrix} \begin{bmatrix} \hat{N}_t^{ST} \\ \hat{N}_t^{SA} \end{bmatrix} = \\ & \begin{bmatrix} 1 & 0 \\ 0.237B & 1 \end{bmatrix} \begin{bmatrix} 1 - B^{12} & 0 \\ 0 & 1 - .880B^{12} \end{bmatrix} \begin{bmatrix} \hat{a}_t^{ST} \\ \hat{a}_t^{SA} \end{bmatrix} \\ \hat{\Sigma}_a = & \begin{bmatrix} .007 & -.004 \\ -.004 & .006 \end{bmatrix}, \quad Q(10) = \begin{bmatrix} 7.20 & 7.92 \\ 8.24 & 10.93 \end{bmatrix}, \quad (5.74) \end{aligned}$$

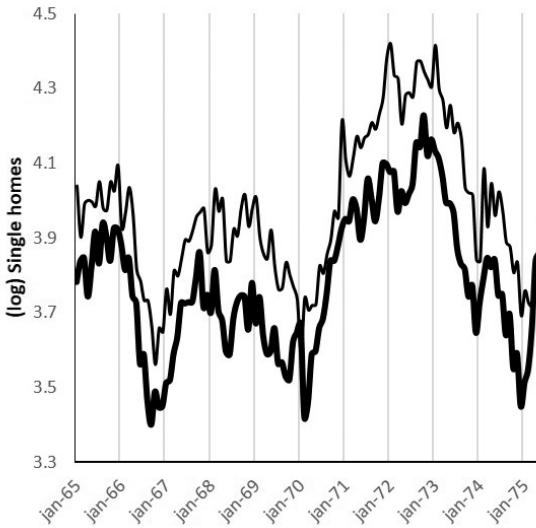


Figure 5.6 Seasonality and calendar-adjusted (log) Housing Starts and Housing Sales series, from January 1965 through May 1975.

where $Q(10)$ denotes the matrix of portmanteau statistics for the null that the model errors are white noise, computed with the first 10 sample auto- and cross-correlations.

Note that equations (5.73)–(5.74) define a simultaneous equation model with VARMAX errors⁶. This model has some interesting implications:

1. The autoregressive parameter in (5.74) provides a cointegration test. A value close enough to unity would imply that there was no cointegration, because the series would require the same orders of differencing as those in the previously fitted univariate models, thus rejecting cointegration. On the other hand, the estimate obtained (.213, with a .085 standard error) is small enough to provide positive evidence in favor of cointegration.
2. The first coefficient in Equation (5.73) is the cointegration parameter. Its estimate (-1.028 with a 0.010 standard error) means that there exists a long term equilibrium between both series. Deviations from this equilibrium are transient and given by the cointegrating relationship $\log ST_t - 1.028 \log SA_t$.
3. The estimated cointegrating parameter is close to unity. If its true value were -1 it would mean that, in the long term, the ratio between Housing Starts and Housing Sales is approximately stable.

The code required to replicate this example can be found in the file Example0544.m, see Appendix D.

⁶See Subsection 3.2.2 for the details about the SS formulation of these combined models.

Chapter 6

The likelihood of models with varying parameters

6.1	Regression with time-varying parameters	66
6.1.1	SS formulation	66
6.1.2	Maximum likelihood estimation	67
6.2	Periodic models	68
6.2.1	All the seasons have the same dynamic structure	69
6.2.2	The s models do not have the same dynamic structure	70
6.2.3	Stationarity and invertibility	71
6.2.4	Maximum likelihood estimation	72
6.3	The likelihood of models with GARCH errors	74
6.4	Examples	76
6.4.1	A time-varying CAPM regression	76
6.4.2	A periodic model for West German Consumption	78
6.4.3	A model with vector-GARCH errors for two exchange rate series	79

In a statistical model, any value to be estimated must be either a constant, or a function of constant parameters. Thus, the reference to “varying parameters” in the title of this chapter is somewhat misleading. However, it is standard terminology, referring to models which allow for the variation over time of some feature which is considered fixed by most standard specifications.

In particular, we will discuss how to compute the Gaussian likelihood for three specifications of this kind. First, a time-varying parameter regression model, where the coefficients are allowed to change according to a given law of motion. Second, a periodic VARMAX model (PVARMAX), in which the parameters may change over different seasons. Third and last, a general dynamic model with conditionally heteroscedastic (GARCH) errors, where the traditional assumption of homoscedasticity is relaxed by parameterizing the conditional variance of the disturbances as a function of its own past.

6.1 Regression with time-varying parameters

The literature shows an increasing interest in estimating regressions with time-varying parameters. For example, Cooley and Prescott [56] and Rosenberg [188] study how different specification errors, such as structural changes, heteroscedasticity or incorrect functional forms, may result in a model with changing parameters. We can see a survey about varying coefficient regression models and new developments in Park *et al.* [177]. Economic theory suggests that many relationships may change over time. Along this line, Lucas [154] showed that econometric models may be inadequate to assess economic policy in the long term because they inherently assume parametric stability.

According to this interest, the application of time-varying regressions is becoming widespread in areas such as financial economics, see Fabozzi and Francis [89], Alexander and Benson [4] or Bera, Higgins, and Lee [22], macroeconomics, see Margaritis [162], Evans, Kim, and Oh [88], Stock and Watson [199], microeconomics, see Hackl [114], Chang and Hsing [50], Park and Zhao [178], as well as in other fields such as e.g. politics, epidemiology, or medical science, where the constant parameters assumption is not necessary or realistic because of the increasing availability of disaggregated data.

Varying coefficient models (not regression models) are natural extensions of standard parametric models with good interpretability and are becoming popular in data analysis, thus experiencing fast development on the methodological, theoretical, and applied side; see, for example, Fan and Zhang [93], Cai, Fan, and Yao [36], Fan *et al.* [91], Fan and Zhang [92], Kauermann and Tutz [135], Hastie and Tibshirani [122], or Nicholls and Quinn [167].

6.1.1 SS formulation

A general regression model with changing parameters is¹:

$$\mathbf{z}_t = \boldsymbol{\chi}_t \boldsymbol{\beta}_t \quad (6.1)$$

$$\boldsymbol{\beta}_t = \mathbf{H} \mathbf{x}_t + \mathbf{D} \mathbf{u}_t + \mathbf{C} \mathbf{v}_t \quad (6.2)$$

$$\mathbf{x}_{t+1} = \boldsymbol{\Phi} \mathbf{x}_t + \boldsymbol{\Gamma} \mathbf{u}_t + \mathbf{E} \mathbf{w}_t, \quad (6.3)$$

where \mathbf{z}_t is an observable endogenous variable, $\boldsymbol{\chi}_t$ is a row ($1 \times k$) of explanatory variables with $\chi_{1t} = 1$, for all t , $\boldsymbol{\beta}_t$ is a k -dimensional vector of unknown parameters, and \mathbf{u}_t is a r -dimensional vector of variables affecting the coefficients, with $u_{1t} = 1$, for all t .

To write the model (6.1)–(6.3) in SS form, we need to substitute (6.2) in (6.1):

$$\mathbf{z}_t = (\boldsymbol{\chi}_t \mathbf{H}) \mathbf{x}_t + (\boldsymbol{\chi}_t \mathbf{D}) \mathbf{u}_t + (\boldsymbol{\chi}_t \mathbf{C}) \mathbf{v}_t, \quad (6.4)$$

or, equivalently:

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{D}_t \mathbf{u}_t + \mathbf{C}_t \mathbf{v}_t, \quad (6.5)$$

¹The model (6.1)–(6.3) generalizes the formulation in Swamy and Tavlas [205].

with $\mathbf{H}_t = \chi_t \mathbf{H}$, $\mathbf{D}_t = \chi_t \mathbf{D}$ and $\mathbf{C}_t = \chi_t \mathbf{C}$. Therefore, equations (6.5) and (6.3) define an SS model with a time-varying observation equation, which can be trivially generalized to the case of multiple outputs.

There are three specifications for the parameter change which have special interest: random walk, random coefficients, and mean-reverting parameters.

The *random walk model* assumes that the coefficients follow a random walk process:

$$z_t = \chi_t \beta_t + \varepsilon_t \quad (6.6)$$

$$\beta_{t+1} = \beta_t + a_t, \quad (6.7)$$

where ε_t is a white noise sequence that is uncorrelated with a_t . The formulation of this model in SS form involves defining $\Phi = \mathbf{I}$, $\Gamma = \mathbf{0}$, $\mathbf{E} = \mathbf{I}$, $\mathbf{H}_t = \chi_t$, $\mathbf{D}_t = \mathbf{0}$, $\mathbf{C}_t = \mathbf{I}$, $\mathbf{Q} = \Sigma_a$, $\mathbf{S} = \mathbf{0}$, and $\mathbf{R} = \sigma_\varepsilon^2$.

In the *random coefficients model* the transition matrix is null and the parameters are assumed to change randomly around a constant mean value $\bar{\beta}$:

$$z_t = \chi_t \beta_t + \varepsilon_t \quad (6.8)$$

$$\beta_t = \bar{\beta} + a_t, \quad (6.9)$$

where $\Phi = \mathbf{0}$, $\Gamma = \mathbf{0}$, $\mathbf{E} = \mathbf{I}$, $\mathbf{H}_t = \chi_t$, $\mathbf{D}_t = \chi_t \bar{\beta}$, $\mathbf{C}_t = \mathbf{I}$, $\mathbf{Q} = \Sigma_a$, $\mathbf{S} = \mathbf{0}$ and $\mathbf{R} = \sigma_\varepsilon^2$.

Last, the parameters in the *mean-reverting model* follow a stationary AR(1) process given by:

$$z_t = \chi_t \beta_t + \varepsilon_t \quad (6.10)$$

$$(\beta_{t+1} - \bar{\beta}) = F(\beta_t - \bar{\beta}) + a_t, \quad (6.11)$$

where $\Phi = F$, $\Gamma = \mathbf{0}$, $\mathbf{E} = \mathbf{I}$, $\mathbf{H}_t = \chi_t$, $\mathbf{D}_t = \chi_t \bar{\beta}$, $\mathbf{C}_t = \mathbf{I}$, $\mathbf{Q} = \Sigma_a$, $\mathbf{S} = \mathbf{0}$, and $\mathbf{R} = \sigma_\varepsilon^2$.

6.1.2 Maximum likelihood estimation

Estimation of the models defined in Subsection 6.1.1 is generally addressed by the procedures described in Chapter 5 for the MEM formulation. In particular, the log-likelihood function to be evaluated is the one given in expression (5.46), its addends being the result of the following KF:

$$\hat{x}_{t+1|t} = \Phi \hat{x}_{t|t-1} + \Gamma u_t + K_t \tilde{z}_t \quad (6.12)$$

$$\tilde{z}_t = z_t - H_t \hat{x}_{t|t-1} - D_t u_t \quad (6.13)$$

$$B_t = H_t P_{t|t-1} H_t^T + C_t R C_t^T \quad (6.14)$$

$$K_t = (\Phi P_{t|t-1} H_t^T + E S C_t^T) B_t^{-1} \quad (6.15)$$

$$P_{t+1|t} = \Phi P_{t|t-1} \Phi^T + E Q E^T - K_t B_t K_t^T. \quad (6.16)$$

The model given by (6.3) and (6.5) can also be estimated using the log-likelihood (5.50) where the terms: (a) \tilde{z}_t and B_t result from the $KF(\mathbf{0}, \mathbf{0})$ given by (6.12)–(6.16), and (b) w_t and W_t in (5.50) are computed by the recursions:

$$w_t = w_{t-1} + \bar{\Phi}_{t-1} H_t^T B_t^{-1} \tilde{z}_t \quad (6.17)$$

$$W_t = W_{t-1} + \bar{\Phi}_{t-1} H_t^T B_t^{-1} H_t \bar{\Phi}_{t-1} \quad (6.18)$$

$$\bar{\Phi}_t = (\Phi_t - K_t H_t) \bar{\Phi}_{t-1}, \quad (6.19)$$

where w_t and W_t are initialized to zero, $\bar{\Phi}_0 = I$, and K_t is the $KF(\mathbf{0}, \mathbf{0})$ gain.

The proper choice of initial conditions μ and P_1 in (5.50) depends on whether the model (6.3) and (6.5) is stationary or not, and on the nature (deterministic or stochastic) of its inputs, see Sections 5.3.1 and 5.3.2.

6.2 Periodic models

A periodic PVARMAX model extends the standard VARMAX framework by allowing the model parameters to change over different seasons. The PVARMAX model assumes that the observations in each of the seasons can be described using a different model. Such a property may be useful to describe some economic time series, since in several situations one may expect economic agents to behave differently in different seasons; see Ghysels and Osborn [107]. Most applications exclude the MA part of models. Periodic AR models have been used successfully in macroeconomic time series; see, for example, Novales and Flores [168], Franses [100] or Osborn and Smith [171].

As is known, it is said that the random vector z_t , for all $t = 1, 2, \dots, N$, follows a process $PVARMAX_s(p_i, q_i, g_i)$, for all $i = 1, 2, \dots, s$ where s is the number of sample periods required to complete a seasonal cycle, if:

$$F_t(B) z_t = G_t(B) u_t + \Theta_t(B) a_t, \quad (6.20)$$

where the parameters' matrices are:

$$F_t(B) = I + F_{t,1}B + F_{t,2}B^2 + \dots + F_{t,p_i}B^{p_i} \quad (6.21)$$

$$G_t(B) = I + G_{t,1}B + G_{t,2}B^2 + \dots + G_{t,g_i}B^{g_i} \quad (6.22)$$

$$\Theta_t(B) = I + \Theta_{t,1}B + \Theta_{t,2}B^2 + \dots + \Theta_{t,q_i}B^{q_i} \quad (6.23)$$

$$a_t \sim \text{IID}(\mathbf{0}, \Sigma_t). \quad (6.24)$$

and the coefficient matrices change according to a deterministic variation law with period s , such that:

$$F_t(B) = F_{t+s}(B); \quad G_t(B) = G_{t+s}(B); \quad \Theta_t(B) = \Theta_{t+s}(B), \quad (6.25)$$

and

$$\Sigma_t = \Sigma_{t+s}; \quad \text{for all } t = 1, 2, \dots, N \quad (6.26)$$

so that the PVARMAX model can be seen to be a VARMAX model whose coefficients change periodically over time; see Lutkepohl [158].

Equations (6.20)–(6.26) can be written in SEM form as:

$$\mathbf{x}_{t+1} = \Phi_t \mathbf{x}_t + \Gamma_t \mathbf{u}_t + \mathbf{E}_t \mathbf{a}_t \quad (6.27)$$

$$\mathbf{z}_t = \mathbf{H}_t \mathbf{x}_t + \mathbf{D}_t \mathbf{u}_t + \mathbf{a}_t, \quad (6.28)$$

where the covariance matrix of the error vector \mathbf{a}_t is \mathbf{Q}_t , and the dimensions of Φ_t , Γ_t , \mathbf{E}_t , \mathbf{H}_t , and \mathbf{D}_t are, respectively, $(n_{t+1} \times n_t)$, $(n_{t+1} \times r)$, $(n_{t+1} \times m)$, $(m \times n_t)$, and $(m \times r)$. Finally, the parameter variation laws are: $\Phi_t = \Phi_{t+s}$, $\Gamma_t = \Gamma_{t+s}$, $\mathbf{E}_t = \mathbf{E}_{t+s}$, $\mathbf{H}_t = \mathbf{H}_{t+s}$, $\mathbf{D}_t = \mathbf{D}_{t+s}$, and $\mathbf{Q}_t = \mathbf{Q}_{t+s}$, for all $t = 1, 2, \dots, N$.

Note that the dimension of state vector \mathbf{x}_{t+1} may change over different seasons and, therefore, the transition matrix Φ may not be square in some cases; see e.g. Example 6.2.2 below.

To determine the relationship between the model (6.20)–(6.26) and matrices of (6.27)–(6.28), it is convenient to distinguish between the case when the dynamic structure is always the same and when it may be different in some seasons.

6.2.1 All the seasons have the same dynamic structure

When all the seasons have the same dynamic structure, it holds that $k_i = k$ for all $i = 1, 2, \dots, s$ where $k_i = \max\{p_i, q_i, g_i\}$ and the matrices of (6.27)–(6.28) are:

$$\Phi_t = \begin{bmatrix} -\mathbf{F}_{t+1,1} & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ -\mathbf{F}_{t+2,2} & \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \dots & \vdots \\ -\mathbf{F}_{t+k-2,k-1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} \\ -\mathbf{F}_{t+k-1,k} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}, \quad \Gamma_t = \begin{bmatrix} \mathbf{G}_{t+1,1} - \mathbf{F}_{t+1,1}\mathbf{G}_{t,0} \\ \mathbf{G}_{t+2,2} - \mathbf{F}_{t+2,2}\mathbf{G}_{t,0} \\ \vdots \\ \mathbf{G}_{t+k-2,k-1} - \mathbf{F}_{t+k-2,k-1}\mathbf{G}_{t,0} \\ \mathbf{G}_{t+k-1,k} - \mathbf{F}_{t+k-1,k}\mathbf{G}_{t,0} \end{bmatrix}, \quad (6.29)$$

$$\mathbf{E}_t = \begin{bmatrix} \Theta_{t+1,1} - \mathbf{F}_{t+1,1} \\ \Theta_{t+2,2} - \mathbf{F}_{t+2,2} \\ \vdots \\ \Theta_{t+k-2,k-1} - \mathbf{F}_{t+k-2,k-1} \\ \Theta_{t+k-1,k} - \mathbf{F}_{t+k-1,k} \end{bmatrix}, \quad \mathbf{H}_t = [\mathbf{I} \quad \mathbf{0} \quad \dots \quad \mathbf{0}], \quad \mathbf{D}_t = \mathbf{G}_{t,0}, \quad (6.30)$$

where $\mathbf{F}_{t+j,j+1} = \mathbf{F}_{t+j+s,j+1}$, $\mathbf{G}_{t+j,j+1} = \mathbf{G}_{t+j+s,j+1}$, $\Theta_{t+j,j+1} = \Theta_{t+j+s,j+1}$ for all $j = 0, 1, \dots, k-1$.

Example 6.2.1

If $s = 3$ and the model for each season is a stationary AR(1) process, the standard representation of the periodic model $PAR_3(1)$ is:

$$\begin{aligned} z_{t+1} &= f_{1,1} z_t + a_{t+1} \\ z_{t+2} &= f_{2,1} z_{t+1} + a_{t+2} \\ z_{t+3} &= f_{3,1} z_{t+2} + a_{t+3}, \quad \text{for } t = 0, 3, 6, 9, \dots, \end{aligned} \quad (6.31)$$

where the SS formulation of each of the terms above is respectively given by:

$$z_{t+1} = x_{t+1} + a_{t+1} \quad (6.32)$$

$$x_{t+2} = f_{2,1}x_{t+1} + f_{2,1}a_{t+1}, \quad (6.33)$$

$$z_{t+2} = x_{t+2} + a_{t+2} \quad (6.34)$$

$$x_{t+3} = f_{3,1}x_{t+2} + f_{3,1}a_{t+2}, \quad (6.35)$$

$$z_{t+3} = x_{t+3} + a_{t+3} \quad (6.36)$$

$$x_{t+4} = f_{1,1}x_{t+3} + f_{1,1}a_{t+3}. \quad (6.37)$$

6.2.2 The s models do not have the same dynamic structure

When the s models do not have the same structure the standard conversion of the periodic VARMAX model to its corresponding SS representation may yield a non-minimal state vector. Casals, Sotoca and Jerez [44] proposed an SS representation of (6.20)–(6.26) which assures minimality. In this SS representation the number of rows of Φ_t is $n_{t+1} = m \times \sum_{j=1}^s (k_j^+ + k_j^-)$, where k_j^+ is the integer part of k_j/s , and

$$k_j^- = \begin{cases} 1, & \text{if } k_j - k_j^+ s \geq j - t \text{ when } j > t, \\ 1, & \text{if } k_j - k_j^+ s \geq j + s - t \text{ when } j \leq t, \\ 0, & \text{otherwise.} \end{cases}$$

The conversion rule to achieve a minimum dimension representation is to remove redundant states, that is, those that are not affected by any observable (\mathbf{u}_t) or unobservable (\mathbf{a}_t) input. The way to build the matrices Φ_t , Γ_t , and E_t is as follows. Let $k = \max\{k_i\}, i = 1, 2, \dots, s$. Then, if $k_{i+j} \geq j$ for all $j = 1, 2, \dots, k$:

- the Γ_t matrix is constructed by adding blocks of rows $[G_{i+j,j} - F_{i+j,j}G_{i,0}]$;
- the E_t matrix is constructed by adding blocks of rows $[\Theta_{i+j,j} - F_{i+j,j}]$;
- the Φ_t matrix is constructed by adding blocks of rows $[-F_{i+j,j} \mathbf{0} \dots \mathbf{0}]$.

If $k_{i+j} > j$ for all $j = 1, 2, \dots, k$, columns $[\mathbf{0} \mathbf{0} \dots \mathbf{I}]^T$ are added to the matrix Φ_t .

Example 6.2.2

Consider a model where $s = 3$, $k_1 = 6$, $k_2 = 1$ and $k_3 = 1$. According the conventional conversion rule, the non minimum dimension SS formulation is:

$$\Phi_1 = \begin{bmatrix} -f_{2,1} & 1 & 0 & 0 & 0 & 0 \\ -f_{3,2} & 0 & 1 & 0 & 0 & 0 \\ -f_{1,3} & 0 & 0 & 1 & 0 & 0 \\ -f_{2,4} & 0 & 0 & 0 & 1 & 0 \\ -f_{3,5} & 0 & 0 & 0 & 0 & 1 \\ -f_{1,6} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} -f_{3,1} & 1 & 0 & 0 & 0 & 0 \\ -f_{1,2} & 0 & 1 & 0 & 0 & 0 \\ -f_{2,3} & 0 & 0 & 1 & 0 & 0 \\ -f_{3,4} & 0 & 0 & 0 & 1 & 0 \\ -f_{1,5} & 0 & 0 & 0 & 0 & 1 \\ -f_{2,6} & 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad (6.38)$$

$$\Phi_3 = \begin{bmatrix} -f_{1,1} & 1 & 0 & 0 & 0 & 0 \\ -f_{2,2} & 0 & 1 & 0 & 0 & 0 \\ -f_{3,3} & 0 & 0 & 1 & 0 & 0 \\ -f_{1,4} & 0 & 0 & 0 & 1 & 0 \\ -f_{2,5} & 0 & 0 & 0 & 0 & 1 \\ -f_{3,6} & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (6.39)$$

However, the application of the rule above removes redundant states providing the following minimum dimension matrices:

$$\Phi_1 = \begin{bmatrix} -f_{2,1} & 0 \\ -f_{1,3} & 1 \\ -f_{1,6} & 0 \end{bmatrix}, \quad \Phi_2 = \begin{bmatrix} -f_{3,1} & 0 & 0 \\ -f_{1,2} & 1 & 0 \\ -f_{1,5} & 0 & 1 \end{bmatrix}, \quad (6.40)$$

$$\Phi_3 = \begin{bmatrix} -f_{1,1} & 1 & 0 \\ -f_{1,4} & 0 & 1 \end{bmatrix}, \quad (6.41)$$

where the dimension of the state vector changes from being 6 (equal to the maximum dynamic, $k_i = 6$), to be varying with the season, but never exceeding 3.

6.2.3 Stationarity and invertibility

The stationarity and invertibility conditions of a periodic process often are characterized by its equivalent representation VARMA with constant parameters; see, for example, Franses and Paap [101] or Boswijk and Franses [27]. Proceeding in the same way, to characterize the stationarity and invertibility conditions of a PVARMAX process, it is necessary to write the state equation (6.27) in the equivalent form:

$$\begin{aligned} \mathbf{x}_{t+s} = & (\Phi_{t+s-1} \times \Phi_{t+s-2} \times \cdots \times \Phi_t) \mathbf{x}_t + \\ & \sum_{i=1}^s \prod_{j=1}^{s-i} (\Phi_{t+s-j} \boldsymbol{\Gamma}_{t+i-1} \mathbf{u}_{t+i-1} + \Phi_{t+s-1} \mathbf{E}_{t+i-j} \mathbf{a}_{t+i-1}), \end{aligned} \quad (6.42)$$

whose matrices do not change every s periods due to the law of variation $\Phi_t = \Phi_{t+s}$, $\boldsymbol{\Gamma}_t = \boldsymbol{\Gamma}_{t+s}$, and $\mathbf{E}_t = \mathbf{E}_{t+s}$, $t = 1, 2, \dots, N$. Once described, the model in (6.28) and (6.42), is said to be *cycle-stationary* if all eigenvalues of $\Phi^* = \prod_{j=1}^s \Phi_{s-j+1}$ are within the unit circle.

Similarly, to characterize the invertibility conditions of a PVARMAX model one should use its equivalent VARX(∞) representation. Thus the model can be written as:

$$\mathbf{x}_{t+1} = \bar{\Phi}^* \mathbf{x}_{t+1-s} + \bar{\boldsymbol{\Gamma}}^* \begin{bmatrix} \mathbf{u}_t \\ \mathbf{u}_{t-1} \\ \vdots \\ \mathbf{u}_{t+1-s} \end{bmatrix} + \bar{\mathbf{E}}^* \begin{bmatrix} \mathbf{z}_t \\ \mathbf{z}_{t-1} \\ \vdots \\ \mathbf{z}_{t+1-s} \end{bmatrix} \quad (6.43)$$

$$z_{t+1} = \mathbf{H}_{t+1} \bar{\Phi}^* x_{t+1-s} + \mathbf{H}_{t+1} \bar{\Gamma}^* \begin{bmatrix} u_t \\ u_{t-1} \\ \vdots \\ u_{t+1-s} \end{bmatrix} + \mathbf{H}_{t+1} \bar{\mathbf{E}}^* \begin{bmatrix} z_t \\ z_{t-1} \\ \vdots \\ z_{t+1-s} \end{bmatrix} + \mathbf{D}_t u_t + a_{t+1}, \quad (6.44)$$

where

$$\bar{\Phi}^* = (\Phi_t - \mathbf{E}_t \mathbf{H}_t) \times (\Phi_{t-1} - \mathbf{E}_{t-1} \mathbf{H}_{t-1}) \times \cdots \times (\Phi_{t-1+s} - \mathbf{E}_{t-1+s} \mathbf{H}_{t-1+s}) \quad (6.45)$$

$$\bar{\mathbf{E}}^* = [\mathbf{E}_t, (\Phi_t - \mathbf{E}_t \mathbf{H}_t) \mathbf{E}_{t-1}, (\Phi_t - \mathbf{E}_t \mathbf{H}_t)(\Phi_{t-1} - \mathbf{E}_{t-1} \mathbf{H}_{t-1} \mathbf{E}_{t-2}), \dots, \prod_{i=0}^{s-2} (\Phi_{t-i} - \mathbf{E}_{t-i} \mathbf{H}_{t-i}) \mathbf{E}_{t+1-s}] \quad (6.46)$$

$$\bar{\Gamma}^* = [\bar{\Gamma}_t, (\Phi_t - \mathbf{E}_t \mathbf{H}_t) \bar{\Gamma}_{t-1}, (\Phi_t - \mathbf{E}_t \mathbf{H}_t)(\Phi_{t-1} - \mathbf{E}_{t-1} \mathbf{H}_{t-1}) \bar{\Gamma}_{t-2}, \dots, \prod_{i=0}^{s-2} (\Phi_{t-i} - \mathbf{E}_{t-i} \mathbf{H}_{t-i}) \bar{\Gamma}_{t+1-s}], \quad (6.47)$$

with $\bar{\Gamma}_t = \Gamma_t - \mathbf{E}_t \mathbf{D}_t$. Since (6.44) can be interpreted as a VARX(∞) process, substituting recursively x_{t+1-s} according to (6.43), it is straightforward to see that, for the past values of z_t to have a decreasing effect on the current value (invertibility) it is necessary that the eigenvalues in $\bar{\Phi}^*$ lie within the unit circle. Bentarzi and Hallin [21] obtained the same necessary and sufficient condition for invertibility.

6.2.4 Maximum likelihood estimation

The Gaussian likelihood of model (6.27)–(6.28) is given by (5.32). Its evaluation can be accomplished in three phases by computing: (a) \tilde{z}_t and B_t , (b) \mathbf{W}_N and w_N , and (c) using the initial conditions μ and P_1 . As usual, the innovations and their covariances are calculated by propagating a KF($\mathbf{0}, \mathbf{0}$):

$$\hat{x}_{t+1|t} = \Phi_t \hat{x}_{t|t-1} + \Gamma_t u_t + K_t \tilde{z}_t \quad (6.48)$$

$$\tilde{z}_t = z_t - \mathbf{H}_t \hat{x}_{t|t-1} - D_t u_t \quad (6.49)$$

$$B_t = Q_t \quad (6.50)$$

$$K_t = E_t \quad (6.51)$$

$$P_{t+1|t} = \mathbf{0}, \quad (6.52)$$

where we have taken into account that $P = \mathbf{0}$ is the solution of the Riccati equation of the KF for the SEM (6.27)–(6.28), even when some matrices are time-varying. Moreover, the vector w_N and the matrix \mathbf{W}_N are computed by once again propagating (6.17)–(6.19) with the same initial conditions.

The choice of initial conditions μ and P_1 should take into account whether the model (6.27)–(6.28) is stationary or not, and if it has deterministic and/or stochastic inputs (see Sections 5.2 and 5.3), according to the following procedure.

1. *Stationary PVARMAX models.* If the model is stationary and has no exogenous variables, the proper initialization of the KF is $\mu = \mathbf{0}$, with P_1 chosen to be the solution of the Lyapunov equation corresponding to (6.42). The likelihood function in this case would be:

$$\ell(Z; \theta) = \log|P_1| + \sum_{t=1}^N \log|Q_t| + \sum_{t=1}^N \tilde{z}_t^T Q_t^{-1} \tilde{z}_t + \log|P_1^{-1} + W_N| - w_N^T (P_1^{-1} + W_N)^{-1} w_N, \quad (6.53)$$

where w_N and W_N are calculated from $w_t = w_{t-1} + \bar{\Phi}_{t-1} H_t^T Q_t^{-1} \tilde{z}_t$, and $W_t = W_{t-1} + \bar{\Phi}_{t-1} H_t^T Q_t^{-1} H_t \bar{\Phi}_{t-1}$, being $\Phi_t = (\Phi_t - E_t H_t) \bar{\Phi}_{t-1}$, with $\bar{\Phi}_1 = I$. Now, denoting by M the Cholesky factor of P_1 , we have that $P_1 = MM^T$, $|P_1| = |MM^T| = |M|^2$, and $\log|P_1| + \log|P_1^{-1} + W_N| = \log|I + M^T W_N M|$. Therefore, (6.53) can be written as:

$$\ell(Z; \theta) = \log|I + M^T W_N M| + \sum_{t=1}^N \log|Q_t| + \sum_{t=1}^N \tilde{z}_t^T Q_t^{-1} \tilde{z}_t - w_N^T M (I + M^T W_N M)^{-1} M^T w_N. \quad (6.54)$$

If L is the Cholesky factor of $I + M^T W_N M$ such that $(I + M^T W_N M)^{-1} = (LL^T)^{-1}$, it is easy to obtain recursively a λ vector that satisfies the triangular system of equations $L\lambda = M^T w_N$, and consequently, (6.54) can be written in simplified form as:

$$\ell(Z; \theta) = \log|L|^2 + \sum_{t=1}^N \log|Q_t| + \sum_{t=1}^N \tilde{z}_t^T Q_t^{-1} \tilde{z}_t - \lambda^T \lambda, \quad (6.55)$$

where if N is the total number of observations in the sample and N_i the number of observations in the i -th season, so that $\sum_{i=1}^s N_i = N$. Note that the term $\sum_{t=1}^N \log|Q_t|$ in (6.55) can therefore be written as $\sum_{i=1}^s N_i \log|Q_i|$.

2. *Nonstationary PVARMAX models.* If the model is nonstationary and has no exogenous variables, the Kalman filter must be initialized with $\mu = \mathbf{0}$ and $P_1^{-1} = \mathbf{0}$, which is associated with infinite initial uncertainty. In this case, we can evaluate the minimally conditioned likelihood defined in Subsection 5.2.2.
3. *Partially Nonstationary PVARMAX models.* If the model is partially nonstationary and has no exogenous variables, the eigenvalues of $\prod_{j=1}^s \Phi_{s-j+1}$ will be either on or inside the unit circle. In this situation, the appropriate initial conditions are $\mu = \mathbf{0}$, with P_1^{-1} set to a finite non-null matrix; see De Jong and Chu-Chun-Lin [69].

The presence of exogenous variables in the model and its deterministic or stochastic nature will affect the initial conditions for the KF; see Sections 5.2 and 5.3.

6.3 The likelihood of models with GARCH errors

The idea that a model may have conditional heteroscedastic errors, originally proposed by Engle [86], motivated the development of a large family of ARCH/GARCH-type models. In such models, the state vector may provide information about the future variances and covariances as a function of the past history.

A model with GARCH errors has two main elements: the *model for the mean*, which is a standard econometric model for the endogenous variables, and the *model for the variance*, which is a dynamic equation describing the law governing changes in variances and covariances².

We will consider that the model for the mean is (or can be written as) the SEM:

$$\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{E} \varepsilon_t \quad (6.56)$$

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_t + \mathbf{D} \mathbf{u}_t + \varepsilon_t, \quad (6.57)$$

where the distribution of the error ε_t , conditional on the information available up to time $t - 1$, is $NID(\mathbf{0}, \Sigma_t)$ ³, where the error covariance matrix satisfies:

$$\begin{aligned} \text{vech}(\Sigma_t) &= \text{vech}(\bar{\Sigma}) + \sum_{i=1}^{p_\varepsilon} \mathbf{F}_i^\varepsilon \text{vech}(\varepsilon_{t-i} \varepsilon_{t-i}^T) + \sum_{i=0}^{q_\varepsilon} \mathbf{G}_i^\varepsilon \mathbf{u}_t^\varepsilon \\ &\quad + \sum_{i=1}^{s_\varepsilon} \mathbf{\Xi}_i^\varepsilon \text{vech}(\Sigma_{t-i}), \end{aligned} \quad (6.58)$$

and $\text{vech}()$ denotes the half-vectorization operator, which vectorizes the upper triangle and main diagonal of its (matrix) argument.

The vector \mathbf{u}_t^ε allows for exogenous variables explaining changes in the volatility of the process. These variables are often dummies associated with an event, or with some calendar effect, e.g., a holiday. Therefore the conditional variance matrix Σ_t has a complex dependence on past information.

Assuming that the initial state follows a normal distribution, it is easy to prove that the process \mathbf{z}_t is conditionally normally distributed, being its first and second-

²Models with GARCH errors have been very successful, especially in the financial area, because high frequency time series in finance (returns, transaction volumes, etc.) often display periods of high volatility followed by more quiescent phases. This fact, which is known as *volatility clustering*, suggests that there is an autocorrelation structure for the volatility which can be captured by GARCH-type specifications. The theoretical and empirical literature in this field is too numerous to provide even a quick overview. For a survey, see Bollerslev, Engle and Nelson [25].

³In many cases and, in particular, when working with high-frequency financial data, the assumption of normality is not realistic. Despite this, one can still obtain consistent estimates for the model parameters by optimizing the Gaussian log-likelihood; see also Appendix B. These are sometimes referred to as pseudo-likelihood estimates.

order moments the output of the following KF:

$$\tilde{z}_t = z_t - H_t \hat{x}_{t|t-1} - D_t u_t \quad (6.59)$$

$$\hat{x}_{t+1|t} = \Phi_t \hat{x}_{t|t-1} + \Gamma_t u_t + K_t \tilde{z}_t \quad (6.60)$$

$$K_t = (\Phi P_{t|t-1} H^T + E \Sigma_t) B_t^{-1} \quad (6.61)$$

$$P_{t+1|t} = \Phi P_{t|t-1} \Phi^T + E \Sigma_t E^T - K_t B_t K_t^T \quad (6.62)$$

$$B_t = H P_{t|t-1} H^T + \Sigma_t. \quad (6.63)$$

Therefore, the propagation of (6.59)–(6.63) provides the variables required to evaluate the (minus) log-likelihood of the innovations \tilde{z}_t (5.46), taking into account the nature of Σ_t .

As $E[\text{vech}(\varepsilon_t \varepsilon_t^T) | \Omega_{t-1}] = \text{vech}(\Sigma_t)$, we can define a vector, $\xi_t = \text{vech}(\varepsilon_t \varepsilon_t^T) - \text{vech}(\Sigma_t)$, such that $E(\xi_t | \Omega_{t-1}) = \mathbf{0}$. Now, adding ξ_t to both sides of (6.58), we obtain:

$$\begin{aligned} \text{vech}(\varepsilon_t \varepsilon_t^T) &= \text{vech}(\bar{\Sigma}) + \sum_{i=1}^{p_\varepsilon} F_i^\varepsilon \text{vech}(\varepsilon_{t-i} \varepsilon_{t-i}^T) + \sum_{i=0}^{q_\varepsilon} G_i^\varepsilon u_t^\varepsilon \\ &\quad + \sum_{i=1}^{s_\varepsilon} \Xi_i^\varepsilon \text{vech}(\varepsilon_i \varepsilon_t^T) + \xi_t - \sum_{i=1}^{s_\varepsilon} \Xi_i^\varepsilon \xi_{t-i}, \end{aligned} \quad (6.64)$$

or the equivalent VARMAX representation:

$$\begin{aligned} \text{vech}(\varepsilon_t \varepsilon_t^T) &= \text{vech}(\bar{\Sigma}) + \sum_{i=1}^{\bar{p}_\varepsilon} \bar{F}_i^\varepsilon \text{vech}(\varepsilon_{t-i} \varepsilon_{t-i}^T) \\ &\quad + \sum_{i=0}^{q_\varepsilon} G_i^\varepsilon u_t^\varepsilon + \xi_t - \sum_{i=1}^{s_\varepsilon} \Xi_i^\varepsilon \xi_{t-i}, \end{aligned} \quad (6.65)$$

where $\bar{p}_\varepsilon = \max(p_\varepsilon, q_\varepsilon)$, and $\bar{F}_i^\varepsilon = F_i^\varepsilon + \Xi_i^\varepsilon$, with $F_i^\varepsilon = \mathbf{0}$ and $\Xi_i^\varepsilon = \mathbf{0}$ for all $i > q_\varepsilon$. The VARMAX formulation (6.65) has an SS representation given by:

$$x_{t+1}^\varepsilon = \Phi^\varepsilon x_t^\varepsilon + \Gamma^\varepsilon u_t^\varepsilon + E^\varepsilon \xi_t \quad (6.66)$$

$$\text{vech}(\varepsilon_t \varepsilon_t^T) = H^\varepsilon x_t^\varepsilon + D^\varepsilon u_t^\varepsilon + \xi_t, \quad (6.67)$$

and, as the vector $\text{vech}(\varepsilon_t \varepsilon_t^T)$ in equation (6.67) is not observable, it is replaced by:

$$\text{vech}(\tilde{z}_t \tilde{z}_t^T) = H^\varepsilon x_t^\varepsilon + D^\varepsilon u_t^\varepsilon + \xi_t, \quad (6.68)$$

where $\tilde{z}_t = z_t - \hat{z}_{t|t-1}$, and its covariance matrix is given by (6.63).

Note that, in a representation such as (6.56)–(6.57), $\bar{P} = \mathbf{0}$ is the solution of the algebraic Riccati equation:

$$\begin{aligned} \bar{P} &= \Phi \bar{P} \Phi^T + E \Sigma_t E^T \\ &\quad - (\Phi \bar{P} H^T + E \Sigma_t)(H \bar{P} H^T + \Sigma_t)^{-1} (\Phi \bar{P} H^T + E \Sigma_t)^T, \end{aligned} \quad (6.69)$$

associated with equation (6.62); see Appendix A.

Since $\text{vech}(\Sigma_t) = \mathbb{E}[\text{vech}(\varepsilon_t \varepsilon_t^T) | \Omega_{t-1}]$, a KF applied to the model given by (6.66) and (6.68) provides the minimum variance linear estimator of $\text{vech}(\varepsilon_t \varepsilon_t^T)$ conditional on the information available up to time $t-1$, regardless of the error distribution ξ_t ; see Anderson and Moore [6]. This property allows us to define $\text{vech}(\Sigma_t) = \mathbf{H}^\varepsilon \hat{x}_{t|t-1}^\varepsilon + \mathbf{D}^\varepsilon u_t^\varepsilon$, where:

$$\hat{x}_{t+1|t}^\varepsilon = (\Phi^\varepsilon - \mathbf{K}_t^\varepsilon \mathbf{H}^\varepsilon) \hat{x}_{t|t-1}^\varepsilon + \Gamma^\varepsilon u_t^\varepsilon + \mathbf{K}_t^\varepsilon \text{vech}(\tilde{z}_t \tilde{z}_t^T) \quad (6.70)$$

$$\mathbf{K}_t^\varepsilon = [\Phi^\varepsilon \mathbf{P}_{t|t-1}^\varepsilon (\mathbf{H}^\varepsilon)^T + \mathbf{E}^\varepsilon \Sigma^\varepsilon] (\mathbf{B}_t^\varepsilon)^{-1} \quad (6.71)$$

$$\mathbf{P}_{t+1|t}^\varepsilon = \Phi^\varepsilon \mathbf{P}_{t|t-1}^\varepsilon (\Phi^\varepsilon)^T + \mathbf{E}^\varepsilon \mathbf{B}_t^\varepsilon (\mathbf{E}^\varepsilon)^T - \mathbf{K}_t^\varepsilon \mathbf{B}_t^\varepsilon (\mathbf{K}_t^\varepsilon)^T \quad (6.72)$$

$$\mathbf{B}_t^\varepsilon = \mathbf{H}^\varepsilon \mathbf{P}_{t|t-1}^\varepsilon (\mathbf{H}^\varepsilon)^T + \Sigma^\varepsilon. \quad (6.73)$$

Assuming that x_1^ε is a null vector, the filter (6.70)–(6.73) collapses to the expression:

$$\hat{x}_{t+1|t}^\varepsilon = (\Phi^\varepsilon - \mathbf{E}^\varepsilon \mathbf{H}^\varepsilon) \hat{x}_{t|t-1}^\varepsilon + \Gamma^\varepsilon u_t^\varepsilon + \mathbf{E}^\varepsilon \text{vech}(\tilde{z}_t \tilde{z}_t^T). \quad (6.74)$$

Note that this assumption implies that $\mathbf{P}_1^\varepsilon = \mathbf{0}$ is the solution to the algebraic Riccati equation associated with (6.71), and therefore $\mathbf{P}_{t|t-1}^\varepsilon = \mathbf{0}$ and $\mathbf{K}_t^\varepsilon = \mathbf{E}^\varepsilon$ for all t .

If the eigenvalues of Φ^ε are inside the unit circle, model (6.56)–(6.57) implies that the unconditional variance is finite. On the other hand, if the model has some unit roots it would reduce to an IGARCH; see Engle and Bollerslev [84].

The conditional variance model (6.58) does not assure that Σ_t will be positive semi-definite. In the univariate case is easy to constrain the parameters to avoid negative variances. For example, in a model where the conditional variance is:

$$\sigma_t^2 = \bar{\sigma}^2 + f^\varepsilon \varepsilon_{t-1}^2 + \xi^\varepsilon \sigma_{t-1}^2, \quad (6.75)$$

the conditions assuring non-negativity of the variance are $\bar{\sigma}^2 > 0$, $f^\varepsilon \geq 0$, and $\xi^\varepsilon \geq 0$. In the multivariate case however, the constraints required to ensure that Σ_t is positive-semidefinite are not trivial. Because of this, Engle and Kroner [85] proposed an alternative formulation known as BEKK, whose functional form assures this condition without additional constraints. The BEKK model can also be interpreted as a constrained form of (6.58).

6.4 Examples

6.4.1 A time-varying CAPM regression

A well known econometric model in finance is the Capital-Asset Pricing Model (CAPM), given by a linear regression between the return of the i -th asset or portfolio, r_t^i and that of a market index r_t^m :

$$r_t^i = \alpha^i + \beta^i r_t^m + \varepsilon_t^i, \quad (6.76)$$

where (a) the slope parameter (β^i) measures the change in the asset profitability due to changes in the market return, (b) the constant term (α^i) is an extra return not related

to fluctuations in the market, and (c) the error term ε_t^i captures specific changes in r_t^i not explained by the market performance.

The β^i parameter is useful in many ways. For example, it is often used to determine an optimal hedging strategy for the market risk associated with the i -th asset, using synthetic assets related to the market return r_t^m . On the other hand, its value is useful in characterizing the performance of the asset in comparison with that of the market. For example, the returns of an asset with $\beta^i > 1$ typically amplify the market fluctuations, so it would make a good investment in “bull” markets. On the other hand, an asset with $0 < \beta^i < 1$ mitigates the market fluctuations, so it would be a sensible defensive investment in “bear” markets.

Despite its usefulness, there is a broad consensus in the literature about the instability of the β^i coefficient, so it is relevant to develop methods to test and measure its changes over time. We tested this hypothesis with a sample of closing prices of General Electric (GE) shares and the Standard & Poor’s 500 (SP500) Market index, observed during 2,362 consecutive market days from November 1, 1993 to April 3, 2003. Building on this dataset, we first computed the daily return of the GE shares and the Market index, and then subtracted the simultaneous return of a US Treasury Bill divided by 360 to obtain the corresponding excess return series, $erGE_t$ and $erSP500_t^m$. Last, we formulated a time-varying coefficient version of the CAPM model (6.76), assuming that the “alpha” and “beta” parameters follow a random-walk process as in (6.6)–(6.7):

$$erGE_t = \alpha_t^{GE} + \beta_t^{GE} erSP500_t^m + \varepsilon_t^{GE} \quad (6.77)$$

$$\alpha_{t+1}^{GE} = \alpha_t^{GE} + a_t^\alpha \quad (6.78)$$

$$\beta_{t+1}^{GE} = \beta_t^{GE} + a_t^\beta. \quad (6.79)$$

The model was fitted via ML estimation, resulting in a log-likelihood value of 3929.41. The estimates and standard deviations are displayed in Table 6.1.

Table 6.1 *Estimation results for model (6.77)–(6.79).*

Parameter	Estimate	Standard Error
standard deviation of a_t^α	0.000	0.001
standard deviation of a_t^β	0.015	0.006
standard deviation of ε_t^{GE}	1.267	0.019

The parameter sequences implied by this model can be easily estimated by fixed-interval smoothing; see Chapter 4. Note that the estimate for the variance of the “alpha” parameter is very close to zero, so α_t^{GE} is approximately constant, with a fixed-interval smoothed estimate of $\hat{\alpha}_t^{GE} = .043$ and a standard error of .026. The “beta” parameter, on the other hand, has nonzero variance, and so it is indeed time-varying. Figure 6.1 displays the corresponding sequence of smoothed estimates, $\hat{\beta}_t^{GE}$, as well as approximate 95% confidence bounds.

The code required to replicate this example can be found in the file `Example0641.m`; see Appendix D.

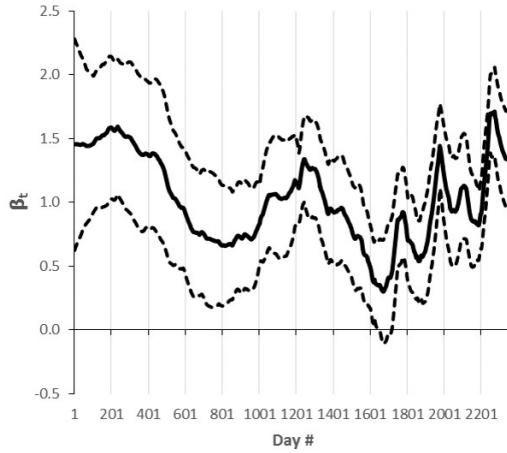


Figure 6.1 Fixed-interval smoothed estimates for β_t (solid line) in the time-varying parameter CAPM regression with $\pm 2\sigma$ limits (dashed line).

6.4.2 A periodic model for West German Consumption

In this example we will model the quarterly unadjusted West German Consumption data for the years 1960–1987, see Table E.4 in Appendix E of Lutkepohl [156]. The profile of this series, see Figure 6.2, displays a strong seasonal pattern and Lutkepohl [156], after testing for parametric variation, concludes that there could be a periodic structure, perhaps affected by structural shifts during the sample period used. We will ignore such potential problems because our goal here is to provide an illustrative example for the periodic models given in Section 6.2.

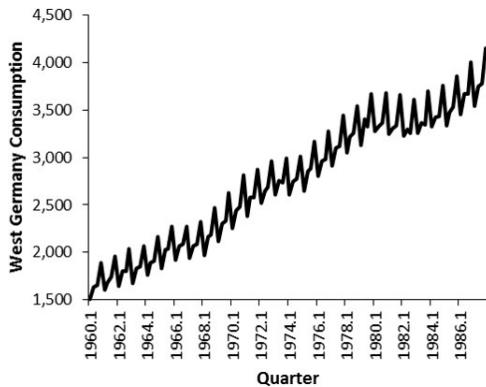


Figure 6.2 Quarterly unadjusted West German Consumption. The data is in billions of Deutsch Marks.

Table 6.2 *Estimation results for the model (6.80). The values on convergence of the (log) likelihood function and the information criteria are, respectively, $\ell^*(\hat{\theta}) = -313.766$; $AIC = -5.715$, $SBC = -5.515$*

t	$\hat{\phi}_1$	$\hat{\phi}_2$	d_t	$\hat{\theta}_1$	$\hat{\sigma}_{a,t}^2$
t = 1, 5, 9,56 (.16)	.48 (.17)	1	—	.011 (.002)
t = 2, 6, 10, ...	—	—	1	.58 (.17)	.019 (.003)
t = 3, 7, 11, ...	—	—	0	—	.010 (.001)
t = 4, 8, 12, ...	—	—	1	.27 (.26)	.011 (.001)

A standard ARIMA specification analysis for each season provided the following general structure:

$$(1 - \phi_{1,t}B - \phi_{2,t}B^2)(1 - B)^{d_t} \log C_t = a_t, \quad a_t \sim \text{IID}(0, \sigma_{a,t}^2), \quad (6.80)$$

where C_t denotes the West German Consumption at quarter t . The estimation results for model (6.80) are summarized in Table 6.2.

On the other hand, C_t can be represented by the following non-periodic ARIMA model:

$$(1 - B)(1 - B^4) \log C_t = (1 - .259B)(1 - .548B^4) \hat{a}_t, \quad \sigma_a^2 = .014 \quad (6.81)$$

$$\ell^*(\hat{\theta}) = -307.017, \quad AIC = -5.687, \quad SBC = -5.608,$$

where the value of likelihood at the optimum is -307.017 , while the corresponding value for the periodic model is -313.766 .

The code required to replicate this example can be found in the file `Example0642.m`; see Appendix D.

6.4.3 A model with vector-GARCH errors for two exchange rate series

This example focuses on estimating the short-run correlation between the spot bid exchange rates of Deutsch Mark (dm) and Japanese Yen (jy) against the US Dollar, observed in the London Market during the 695 week period extending from January 1985 to April 1998. The data has been logged, differenced, and scaled by a factor of 100, to obtain the corresponding log percent yields. Excess returns were then computed by subtracting the sample mean.

Univariate analyses suggest that these (log) excess returns: (a) are stationary, (b) may have a short autoregressive structure in the mean, and (c) some conditional heteroscedasticity in the variance. We will model these features using VAR(1) model for the mean and a diagonal GARCH(1,1) model for the variance. The pseudo-ML

estimation results for this model are:

$$\begin{bmatrix} 1 - .244B \\ (.032) \\ 0 \\ 1 - .300B \\ (.034) \end{bmatrix} \begin{bmatrix} dm_t \\ jy_t \end{bmatrix} = \begin{bmatrix} \hat{\epsilon}_t^{dm} \\ \hat{\epsilon}_t^{jy} \end{bmatrix} \quad (6.82)$$

$$\begin{bmatrix} (\hat{\epsilon}_t^{dm})^2 \\ \hat{\epsilon}_t^{dm} \cdot \hat{\epsilon}_t^{jy} \\ (\hat{\epsilon}_t^{jy})^2 \end{bmatrix} = \begin{bmatrix} 1.810 \\ (.282) \\ 1.145 \\ (.155) \\ 1.666 \\ (.171) \end{bmatrix} + \begin{bmatrix} 1 - .973B & 0 & 0 \\ (.012) & 1 - .942B & 0 \\ 0 & (.015) & 1 - .861B \\ 0 & 0 & (.035) \end{bmatrix}^{-1} \cdot \begin{bmatrix} 1 - .916B & 0 & 0 \\ (.021) & 1 - .873B & 0 \\ 0 & 0 & 1 - .739B \\ (.024) & (.047) \end{bmatrix} \begin{bmatrix} \hat{\xi}_t^{dm} \\ \hat{\xi}_t^{dm,jy} \\ \hat{\xi}_t^{jy} \end{bmatrix}, \quad (6.83)$$

where dm_t and jy_t denote, respectively, the excess returns of the Deutsch Mark and the Japanese Yen. Note that Equation (6.83) is a vector GARCH model in VARMAX form; see equation (6.65).

Finally, Figure 6.3 shows the resulting conditional correlations. Note that they fluctuate around its unconditional mean value of 0.64, with highs and lows of 0.9 and 0.3, respectively. A high conditional correlation value means that the co-movement of both series is strong in this period. On the contrary, low correlations indicate periods when both currencies fluctuate independently.

The code required to replicate this example can be found in the file Example0643.m; see Appendix D.

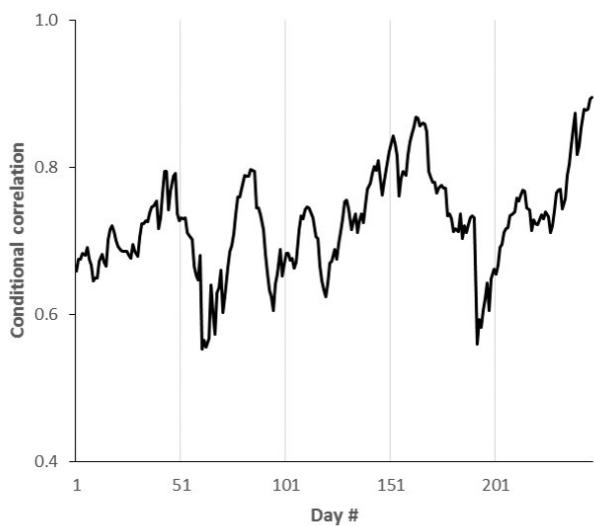


Figure 6.3 *Conditional correlation between Deutsch Mark and Japanese Yen returns.*

This page intentionally left blank

Chapter 7

Subspace methods

7.1	Theoretical foundations	83
7.1.1	Subspace structure and notation	84
7.1.2	Assumptions, projections and model reduction	86
7.1.3	Estimating the system matrices	87
7.2	System order estimation	89
7.2.1	Preliminary data analysis methods	89
7.2.2	Model comparison methods	91
7.3	Constrained estimation	92
7.3.1	State sequence structure	92
7.3.2	Subspace-based likelihood	93
7.4	Multiplicative seasonal models	94
7.5	Examples	95
7.5.1	Univariate models	95
7.5.2	A multivariate model for the interest rates	99

This chapter presents the basic theory of a family of system identification procedures known generically as “subspace methods.” Their aim is to estimate the matrices of an SS model given a series of input and output measurements.

7.1 Theoretical foundations

Subspace methods are a set of algorithms based on systems theory, linear algebra, and statistics, which estimate the parameters in an SS model, i.e., the matrices in (2.1)–(2.2) or (2.19)–(2.20), given a series of input (u_t) and output (z_t) measures. They are based on fast and efficient linear algebra tools, such as the QR decompositions or the Singular Value (SVD), see Appendix A.1 and A.4, respectively. As a consequence, subspace methods are not iterative and, therefore, are very appropriate when dealing with large time series datasets.

The basic idea of these methods consists of estimating the sequence of system states from \mathbf{u}_t and \mathbf{z}_t , without further information on the model dynamics. On the basis of these estimates, the system matrices can be estimated consistently by linear Least Squares (hereafter, LS), so that the model residuals can be obtained in order to ultimately estimate the error covariance matrices.

7.1.1 Subspace structure and notation

Assume that the system inputs, $\mathbf{u}_t \in \mathbb{R}^k$, and outputs, $\mathbf{z}_t \in \mathbb{R}^m$, are related by an SS model in SEM form (2.19)–(2.20)¹. By substituting (2.20) into (2.19) in \mathbf{a}_t and solving the resulting recursion, we obtain:

$$\begin{aligned}\mathbf{x}_t = & (\Phi - \mathbf{K}\mathbf{H})^{t-1}\mathbf{x}_1 + \sum_{j=1}^{t-1} (\Phi - \mathbf{K}\mathbf{H})^{t-1-j} \mathbf{K} \mathbf{z}_j \\ & + \sum_{j=1}^{t-1} (\Phi - \mathbf{K}\mathbf{H})^{t-1-j} (\Gamma - \mathbf{K}\mathbf{D}) \mathbf{u}_j,\end{aligned}\quad (7.1)$$

so that the states at time t depend on the initial state vector, \mathbf{x}_1 , as well as the past values of the input and output.

On the other hand, substituting the recursive solution of (2.19) into the observation equation (2.20), yields:

$$\mathbf{z}_t = \mathbf{H}\Phi^{t-1}\mathbf{x}_1 + \mathbf{H} \sum_{j=1}^{t-1} \Phi^{t-1-j} \Gamma \mathbf{u}_j + \mathbf{D}\mathbf{u}_t + \mathbf{H} \sum_{j=1}^{t-1} \Phi^{t-1-j} \mathbf{K} \mathbf{a}_j + \mathbf{a}_t,\quad (7.2)$$

which implies that the endogenous variable, \mathbf{z}_t , depends on the initial state, as well as current and past values of the innovations, \mathbf{a}_t , and inputs, \mathbf{u}_t .

Equations (7.1) and (7.2) can be written in matrix form as:

$$\mathbf{X}_f = \mathbf{L}_x^i \mathbf{X}_p + \mathbf{L}_z \mathbf{Z}_p + \mathbf{L}_u \mathbf{U}_p,\quad (7.3)$$

and

$$\mathbf{Z}_f = \mathbf{O}_i \mathbf{X}_f + \mathbf{T}_{d,i} \mathbf{U}_f + \mathbf{T}_{s,i} \mathbf{A}_f,\quad (7.4)$$

respectively, where i is a user-defined index which represents the number of past block rows, p . For simplicity we will assume henceforth that the dimension of past and future blocks is equal to i , so that $p = i$ and $f = 2i$. This dimensioning implies that the \mathbf{L} matrices in (7.3) are defined as:

$$\begin{aligned}\mathbf{L}_x &:= (\Phi - \mathbf{K}\mathbf{H})_{n \times n}, \\ \mathbf{L}_z &:= [(\Phi - \mathbf{K}\mathbf{H})^{i-1} \mathbf{K} \quad \dots \quad (\Phi - \mathbf{K}\mathbf{H}) \mathbf{K} \quad \mathbf{K}]_{n \times im},\end{aligned}$$

¹The basic ideas of subspace methods can also be implemented using an MEM form. For example, the N4SID algorithm (see Van Overschee and De Moor [215]) builds on model (2.1)–(2.2), particularly with \mathbf{E} and \mathbf{C} equal to the identity.

and

$$\begin{aligned} \mathbf{L}_u := & [(\Phi - \mathbf{K}\mathbf{H})^{i-1}(\Gamma - \mathbf{K}\mathbf{D}) \\ & \dots \quad (\Phi - \mathbf{K}\mathbf{H})(\Gamma - \mathbf{K}\mathbf{D}) \quad (\Gamma - \mathbf{K}\mathbf{D})]_{n \times ik}. \end{aligned}$$

Thus the matrices \mathbf{L}_x , \mathbf{L}_z , and \mathbf{L}_u have a number of rows equal to the system order, n , while the number of columns is n , im , and ik , respectively.

The system order is typically unknown in practice, but the literature provides several methods to estimate it consistently; see e.g. Garcia-Hieraux, Jerez, and Casals [106] or Bauer [16]. We will describe some of those procedures later. In this section, and for simplicity, we will assume that n is known or has been previously determined.

In what follows, we will define additional matrices in equations (7.3) and (7.4), following the scheme presented by Favoreel, De Moor and Van Overschee [94]:

1. The output and input block Hankel matrices, whose dimensions are determined by the integers p and f , are such that:

$$\mathbf{Z}_p := \begin{bmatrix} z_1 & z_2 & \dots & z_{T-p-f+1} \\ z_2 & z_3 & \dots & z_{T-p-f+2} \\ \vdots & \vdots & & \vdots \\ z_p & z_{p+1} & \dots & z_{T-f} \end{bmatrix}; \quad \mathbf{Z}_f := \begin{bmatrix} z_{p+1} & z_{p+2} & \dots & z_{T-f+1} \\ z_{p+2} & z_{p+3} & \dots & z_{T-f+2} \\ \vdots & \vdots & & \vdots \\ z_{p+f} & z_{p+f+1} & \dots & z_T \end{bmatrix}. \quad (7.5)$$

On the other hand, \mathbf{A}_p , \mathbf{A}_f , \mathbf{U}_p , and \mathbf{U}_f are defined, respectively, as \mathbf{Z}_p and \mathbf{Z}_f , but with their own corresponding observations instead of z_t . Therefore, \mathbf{Z}_p , \mathbf{Z}_f , \mathbf{A}_p and \mathbf{A}_f are $im \times T - 2i + 1$ matrices, while \mathbf{U}_p and \mathbf{U}_f are $ik \times T - 2i + 1$.

2. State sequences are crucial in the derivation of subspace methods. The past and future state sequences are defined as $\mathbf{X}_p := [x_1 \ x_2 \ \dots \ x_{T-p-f+1}]$ and $\mathbf{X}_f := [x_{p+1} \ x_{p+2} \ \dots \ x_{T-f+1}]$. Since $p = i$ and $f = 2i$, both matrices are $n \times T - 2i + 1$.
3. The Extended Observability matrix:

$$\mathbf{O}_i := [\mathbf{H}^T \quad (\mathbf{H}\Phi)^T \quad (\mathbf{H}\Phi^2)^T \quad \dots \quad (\mathbf{H}\Phi^{i-1})^T]_{im \times n}^T. \quad (7.6)$$

4. The deterministic lower block triangular Toeplitz matrix, \mathbf{T}_d , defined as:

$$\mathbf{T}_{d,i} := \begin{bmatrix} D_m & 0 & 0 & \dots & 0 \\ H\Gamma & D_m & 0 & \dots & 0 \\ H\Phi\Gamma & H\Gamma & D_m & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ H\Phi^{i-2}\Gamma & H\Phi^{i-3}\Gamma & H\Phi^{i-4}\Gamma & \dots & D_m \end{bmatrix}_{im}. \quad (7.7)$$

5. The stochastic lower block triangular Toeplitz matrix, \mathbf{T}_s , defined as:

$$\mathbf{T}_{s,i} := \begin{bmatrix} I_m & 0 & 0 & \dots & 0 \\ HK & I_m & 0 & \dots & 0 \\ H\Phi K & HK & I_m & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ H\Phi^{i-2}K & H\Phi^{i-3}K & H\Phi^{i-4}K & \dots & I_m \end{bmatrix}_{im}. \quad (7.8)$$

7.1.2 Assumptions, projections and model reduction

To set up the foundation of the subspace methods, we need the following assumptions:

- A1. The system is strictly minimum-phase. That is, the eigenvalues of \mathbf{L}_x are inside the unit circle.
- A2. The system is minimal, which implies that the pair (Φ, \mathbf{H}) is observable and $(\Phi, [\Gamma, \mathbf{K}])$ is controllable; see Subsection 2.1.3.
- A3. *Open-loop* data assumption: the innovations, \mathbf{a}_t , and the input, \mathbf{u}_t , are uncorrelated.
- A4. The innovations, \mathbf{a}_t , follow a stationary, zero-mean, white noise process with $\mathbb{E}(\mathbf{a}_t \mathbf{a}_t^T) = \mathbf{Q}$.
- A5. The input \mathbf{u}_t is quasi-stationary in the sense of Ljung [152].

From A1 and for sufficiently large i , $\mathbf{L}_x^i \approx 0$. Consequently, substituting equation (7.3) in (7.4) gives:

$$\mathbf{Z}_f = \mathbf{O}_i (\mathbf{L}_z \mathbf{Z}_p + \mathbf{L}_u \mathbf{U}_p) + \mathbf{T}_{d,i} \mathbf{U}_f + \mathbf{T}_{s,i} \mathbf{A}_f, \quad (7.9)$$

which can be written in regression form as:

$$\mathbf{Z}_f = \Theta_i \mathbf{V}_p + \mathbf{T}_{d,i} \mathbf{U}_f + \mathbf{T}_{s,i} \mathbf{A}_f, \quad (7.10)$$

where $\Theta_i := \mathbf{O}_i \mathbf{L}_i$, $\mathbf{L}_i := [\mathbf{L}_z \quad \mathbf{L}_u]$, and $\mathbf{V}_p := [\mathbf{Z}_p^T \quad \mathbf{U}_p^T]^T$.

Equation (7.10) is the basis of subspace methods. It is a regression model which can be consistently estimated by LS because the columns in $\mathbf{T}_{s,i} \mathbf{A}_f$ are uncorrelated with the regressors, given by columns in \mathbf{V}_p and \mathbf{U}_f .²

Note also that the matrix Θ_i is the product of the observability and controllability matrices of the process, and has a reduced rank, n , equal to the system order, which is less than the matrix dimensions $im \times (im + ik)$. Because of this, estimation of (7.10) requires reduced-rank LS; see Appendix A.9. The estimation problem can be written as:

$$\min_{\{\Theta_i, \mathbf{T}_{d,i}\}} \left\| \mathbf{W}_1 (\mathbf{Z}_f - \Theta_i \mathbf{V}_p - \mathbf{T}_{d,i} \mathbf{U}_f) \right\|_F^2, \quad \text{such that } \text{rank}(\Theta_i) = n, \quad (7.11)$$

where $\|\cdot\|_F$ is the Frobenius norm. This problem is solved by first eliminating the term that includes \mathbf{U}_f . To this end, we post-multiply (7.10) by $\Pi_{\mathbf{U}_f}^\perp = \mathbf{I} - \mathbf{U}_f^T (\mathbf{U}_f \mathbf{U}_f^T)^{-1} \mathbf{U}_f$, which is the projection on the orthogonal complement of \mathbf{U}_f . Then we can run a LS regression on the result, which yields:

$$\hat{\Theta}_i = \mathbf{Z}_f \Pi_{\mathbf{U}_f}^\perp \mathbf{V}_p^T (\mathbf{V}_p \Pi_{\mathbf{U}_f}^\perp \mathbf{V}_p^T)^{-1}, \quad (7.12)$$

where we used the idempotency of $\Pi_{\mathbf{U}_f}^\perp$. An SVD decomposition, see Appendix A.4, is now required on $\hat{\Theta}_i$ to deal with the rank reduction. Van Overschee and De Moor

²This is assured by assumptions A3 and A4.

[216] proved that several subspace methods are equivalent to performing:

$$\mathbf{W}_1 \hat{\Theta}_i \mathbf{W}_2 \stackrel{svd}{=} \mathbf{U}_n \mathbf{S}_n \mathbf{V}_n^T + \hat{\mathbf{E}}_n, \quad (7.13)$$

where $\hat{\mathbf{E}}_n$ denotes the approximation error arising from the use of a system order n , and $\mathbf{W}_1, \mathbf{W}_2$ are two nonsingular weighting matrices, whose composition varies according to the specific algorithm being employed; see Table 7.1. Furthermore, \mathbf{W}_2 should not reduce the rank for $\hat{\Theta}_i \mathbf{W}_2$, and must be uncorrelated with \mathbf{A}_f .

Table 7.1 *Definition of the weight matrices \mathbf{W}_1 and \mathbf{W}_2 in equation (7.13). Note that when the CVA approach is used, then the singular values obtained from the SVD coincide with the canonical correlation coefficients between \mathbf{Z}_f and \mathbf{V}_p .*

Weighting	CVA [147]	MOESP [217]	N4SID [215]
\mathbf{W}_1	$(\mathbf{Z}_f \Pi_{\mathbf{U}_f}^\perp \mathbf{Z}_f^T)^{-\frac{1}{2}}$	\mathbf{I}	\mathbf{I}
\mathbf{W}_2	$(\mathbf{V}_p \Pi_{\mathbf{U}_f}^\perp \mathbf{V}_p^T)^{\frac{1}{2}}$	$(\mathbf{V}_p \Pi_{\mathbf{U}_f}^\perp \mathbf{V}_p^T)^{\frac{1}{2}}$	$(\mathbf{V}_p \mathbf{V}_p^T)^{\frac{1}{2}}$

7.1.3 Estimating the system matrices

Once the matrices Θ_i and $\mathbf{T}_{d,i}$ have been estimated by solving problem (7.11), the matrices in (2.19)–(2.20) can be computed via two different approaches. The first one uses the extended observability matrix, while the second uses the state sequence.

Let us first focus on the methods based on the extended observability matrix. From (7.13), we can estimate the extended observability matrix, defined in (7.6), as $\hat{\mathbf{O}}_i = \mathbf{W}_1^{-1} \mathbf{U}_n \mathbf{S}_n^{1/2}$, where we have used the definition $\Theta_i := \mathbf{O}_i [\mathbf{L}_z \quad \mathbf{L}_u]$ presented in equation (7.10), as well as the result in (7.13). Note that if a factorization differing from (7.13) is used, the resulting estimates will be observationally equivalent.

Following Kung [145], Arun and Kung [12], and Verhaegen [217], we can obtain the matrix Φ in (2.19)–(2.20) as:

$$\hat{\Phi} = \underline{\hat{\mathbf{O}}_i^\dagger} \overline{\hat{\mathbf{O}}_i}, \quad (7.14)$$

where $\hat{\mathbf{O}}_i$ and $\overline{\hat{\mathbf{O}}_i}$ denote the result of removing the last and first m rows, respectively, from $\hat{\mathbf{O}}_i$, and where A^\dagger denotes the Moore–Penrose pseudo-inverse of matrix A ; see Appendix A.4, Property 2. The estimate of \mathbf{H} is given by the first m rows of $\hat{\mathbf{O}}_i$.

On the other hand, we can compute the matrices Γ and D using the matrix $\hat{\mathbf{T}}_{d,i}$ resulting from the solution of problem (7.11). In fact, from equation (7.10), we have:

$$\mathbf{O}_i^\perp \mathbf{Z}_f \mathbf{U}_f^\dagger = \mathbf{O}_i^\perp \mathbf{T}_{d,i}, \quad (7.15)$$

where $\mathbf{O}_i^\perp = \mathbf{I} - \mathbf{O}_i \mathbf{O}_i^\dagger$ is a reduced rank matrix that satisfies $\mathbf{O}_i^\perp \mathbf{O}_i = \mathbf{0}$, and assumption A3 assures that $\mathbf{T}_{s,i} \mathbf{A}_f \mathbf{U}_f^\dagger = \mathbf{0}$.

It is straightforward to prove that $\mathbf{Z}_f \mathbf{U}_f^\dagger$ coincides with $\hat{\mathbf{T}}_{d,i}$. Then, taking into account the structure of $\mathbf{T}_{d,i}$ in (7.7), the procedure for obtaining $\hat{\mathbf{D}}$ and $\hat{\mathbf{\Gamma}}$ consists in solving a linear system where Φ , \mathbf{H} , and \mathbf{O}_i^\perp are replaced by their corresponding estimates³. The problem can also be written as:

$$\arg \min_{\{\Gamma, D\}} \left\| \hat{\mathbf{O}}_i^\perp (\hat{\mathbf{T}}_{d,i} - \mathbf{T}_{d,i}(\hat{\Phi}, \Gamma, \hat{\mathbf{H}}, \mathbf{D})) \right\|_F^2. \quad (7.16)$$

Finally, the matrix \mathbf{K} can be determined from an estimate of the extended controllability matrix, \mathbf{L}_i , introduced in (7.10). From the SVD (7.13) we get:

$$\hat{\mathbf{L}}_i = \mathbf{S}_n^{1/2} \mathbf{V}_n^T \mathbf{W}_2^{-1}. \quad (7.17)$$

Similar to the estimation of \mathbf{H} , we can get $\hat{\mathbf{K}}$ from $[\hat{\mathbf{L}}_z \quad \hat{\mathbf{L}}_u] = \hat{\mathbf{L}}_i$ and its corresponding structure defined in (7.3). Therefore, all the system matrices can be estimated using the observability matrix approach.

We will now focus on the methods based on the estimation of the state sequence such as, e.g., those presented in Larimore [147] or Van Overschee and De Moor [215]. These algorithms estimate all model parameters by solving a simple set of LS problems.

From the estimate of the controllability matrix in (7.17), we obtain the state sequence as:

$$\hat{\mathbf{X}}_f \simeq \hat{\mathbf{L}}_i \mathbf{V}_p, \quad (7.18)$$

and similarly, $\hat{\mathbf{X}}_{f^+} \simeq \hat{\mathbf{L}}_i \mathbf{V}_{p^+}$, where \mathbf{V}_{p^+} is like \mathbf{V}_p , defined in (7.10), but adding a ‘+1’ to all the subscripts⁴. Once the state sequence is approximated, the system matrix estimates ($\hat{\Phi}$, $\hat{\Gamma}$, $\hat{\mathbf{H}}$, $\hat{\mathbf{D}}$, $\hat{\mathbf{K}}$) are obtained by LS as:

$$[\hat{\mathbf{H}} \quad \hat{\mathbf{D}}] = \mathbf{Z}_{f_1} \begin{bmatrix} \hat{\mathbf{X}}_f^T & \mathbf{U}_{f_1}^T \end{bmatrix} \left[\begin{bmatrix} \hat{\mathbf{X}}_f \\ \mathbf{U}_{f_1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{X}}_f^T & \mathbf{U}_{f_1}^T \end{bmatrix} \right]^{-1}, \quad (7.19)$$

where \mathbf{Z}_{f_1} and \mathbf{U}_{f_1} denote the first row of \mathbf{Z}_f and \mathbf{U}_f , respectively, in (7.5). From these estimates we then compute the residuals $\hat{\mathbf{A}}_{f_1} = \mathbf{Z}_{f_1} - \hat{\mathbf{H}} \hat{\mathbf{X}}_f - \hat{\mathbf{D}} \mathbf{U}_{f_1}$, finally arriving at the expression:

$$[\hat{\Phi} \quad \hat{\Gamma} \quad \hat{\mathbf{K}}] = \hat{\mathbf{X}}_{f^+} \begin{bmatrix} \hat{\mathbf{X}}_f^T & \mathbf{U}_{f_1}^T & \hat{\mathbf{A}}_{f_1}^T \end{bmatrix} \left[\begin{bmatrix} \hat{\mathbf{X}}_f \\ \mathbf{U}_{f_1} \\ \hat{\mathbf{A}}_{f_1} \end{bmatrix} \begin{bmatrix} \hat{\mathbf{X}}_f^T & \mathbf{U}_{f_1}^T & \hat{\mathbf{A}}_{f_1}^T \end{bmatrix} \right]^{-1}. \quad (7.20)$$

Equations (7.19)–(7.20) therefore demonstrate how the parameter matrices in the SEM can also be estimated from the state sequence approach.

³We follow the convention of topping a matrix with a hat, \hat{A} , to denote an estimate of that matrix, A .

⁴The \simeq sign recalls that the approximation $\mathbf{L}_x^i \approx 0$ for large i is used in the two expressions.

7.2 System order estimation

In Section 7.1 we assumed that the system order, the minimum number of dynamic components required to represent the data dynamics, was known. However, n is unknown in almost all empirical analysis, and specifying its value is critical in applied data modeling. Accordingly, there is an extensive literature about this subject, which can be divided into two broad categories: (i) preliminary data analysis methods, and (ii) procedures based on final estimates.

The first group includes very different proposals. Some of them check whether the model dimension is too high using the condition number of the information matrix; see Stoica and Soderstrom [203]. Other authors, such as Cooper and Wood [57], Tsay [212] or Tiao and Tsay [209], estimate the system order using canonical correlations. In a recent paper, Toscano and Reisen [211] revise and compare these methods. Aoki [11] applies the same technique to state space models. Closely related to this approach, Peternell [180] and Bauer [16] present some subspace-based criteria to estimate the system order, while Sorelius [197] test the rank of a particular sample covariance matrix; see also Woodside [228] with the same aim. In a different context, Bujosa, Garcia–Ferrer and Young [33] apply spectral techniques to obtain valuable information about resonance peaks and phase shifts that aids in identifying the system order required to represent the dynamics.

On the other hand, the methods based on final estimates include the procedures that reduce the system order by Wald-type tests, and those that typically rely on information criteria. This last methodology consists of comparing several alternative models by balancing their “goodness-of-fit” with a penalty depending on the sample size and a measure of model complexity. The pioneering criterion in this vein is due to Akaike [2], whose seminal work motivated other proposals such as those of Schwarz [191] or Hannan and Quinn [117].

Finally, it is worth noting that determining the system order in an SS model is intrinsically more parsimonious than, for instance, the popular VAR approach, which represents a fair alternative to the subspace methods in terms of balancing model simplicity with computational cost. To clarify this remark, consider a m -vector of time series and the decision to increase the order of a $\text{VAR}(p)$ model to a $\text{VAR}(p+1)$. In a LS VAR framework, this would imply increasing the dimension of the state vector by adding m new states, while in a SEM representation one could use the above methods to test all the possible choices in the sequence $1, 2, \dots, m$ of additional states, without imposing any parameter restrictions.

For the sake of simplicity, in the remaining Section 7.2 we will assume the SEM (2.19)–(2.20) without deterministic inputs u_t . In this case, the model dynamics is determined exclusively from the relation between z_t with its own past.

7.2.1 Preliminary data analysis methods

In this section we present three procedures specifically adapted or designed for subspace methods.

1. First, one may determine the system order by testing the null hypothesis that $\sigma_{n+1} = \sigma_{n+2} = \dots = \sigma_{im} = 0$, where σ_{n+1} denotes the $(n+1)$ -th *canonical correlation* between Z_p and Z_f ; see Appendix A.5 for additional details on canonical correlations. These two matrices were introduced in the previous section and defined in (7.5). To do so, we suggest the statistic

$$D(n) = -(T - 2i + 1) \sum_{k=n+1}^{im} \log \left(1 - \frac{\hat{\sigma}_k^2}{\hat{d}_k} \right), \quad (7.21)$$

with $\hat{d}_k = 1 + 2 \sum_{l=1}^i \hat{\rho}_{pk}(l) \hat{\rho}_{fk}(l)$, and where $\hat{\rho}_{fk}(l)$ and $\hat{\rho}_{pk}(l)$ are the sample l -order autocorrelations of the canonical variables. The term \hat{d}_k is related to the asymptotic variance of the estimated canonical correlation coefficients. Under the null, $D(n)$ follows a χ^2 distribution with $2(im - n)$ degrees of freedom; see Tsay [212] for further details.

2. Second, Bauer [16] proposes the *SVC* (Singular Values Criterion) which is based on the analysis of the largest neglected singular value in the SVD step; see (7.13). This criterion is defined as:

$$SVC(n) = \hat{\sigma}_{n+1}^2 + C(T)d(n), \quad (7.22)$$

where $\hat{\sigma}_{n+1}$ is the $(n+1)$ -th singular value of $W_1 \hat{\Theta}_i W_2$, and $d(n) = 2nm$ denotes the number of parameters in the original SEM model. Note that we also denote the singular values by σ_j , for all $j = 1, 2, \dots, im$, since, when CVA methods are used, the weighting matrices W_1 and W_2 are such that the singular values of $W_1 \hat{\Theta}_i W_2$ coincide with the canonical correlations between Z_p and Z_f ; see Table 7.1. In this case, the method closely approximates that of Tsay [212]. The estimated order is obtained as the argument that minimizes (7.22). Bauer [16] shows in a simulation exercise that *SVC* with $C(T) = \log T / T$ outperforms several information criteria when the sample is relatively large, but recognizes that: (i) the choice of W_1 has a large impact on the performance of *SVC* in finite samples, and (ii) the penalty function is somewhat arbitrary.

3. Third, to overcome the latter drawbacks, Garcia-Hiernaux, Jerez, and Casals [106] propose two refinements. They first substitute $W_1 = (Z_f Z_f^T)^{-\frac{1}{2}}$ by the matrix $\tilde{W}_1 = (\tilde{Z}_f \tilde{Z}_f')^{-\frac{1}{2}}$, where \tilde{Z}_f contains the prediction errors of the output future block, e.g., the residuals of previously regressing Z_f onto Z_p . W_1 is chosen for its asymptotic properties, not its finite sample behavior. \tilde{W}_1 is suggested for two reasons: (i) a lessening in the effects of heteroscedasticity (a consequence of the fact that uncertainty grows when the prediction horizon increases) as well as a reduction in the correlation among prediction errors, and (ii) the estimation results obtained with \tilde{W}_1 and W_1 are asymptotically equivalent⁵, so that the asymptotic properties of *SVC* hold.

⁵The parameters estimated with both matrices have the same asymptotic distributions; see e.g. Bauer and Ljung [17].

The second procedure consists of using a refined penalty function, denoted by $H(T, i)$, which depends not only on the sample size, but also on the number of block rows in the output block Hankel matrix. From a geometric point of view, σ_j can be interpreted as the cosine of the angles between subspaces. In this way, σ_1 corresponds to the minimum angle between all pairs of vectors that belong to each subspace. Obviously, increasing the dimension of the subspace reduces the angle. The assessment of the canonical correlations should necessarily take this fact into account; a basic argument in favor of the definition of this penalty function so as to improve its power in finite samples. Intuitively, a larger subspace dimension implies, *caeteris paribus*, higher singular values, since adding new explanatory variables increases the coefficient of determination in a regression model. With this in mind, the “ n identification criterion” (*NIDC*), is defined as:

$$NIDC(n) = \tilde{\sigma}_{n+1}^2 + H(T, i)d(n), \quad (7.23)$$

where $\tilde{\sigma}_{n+1}^2$ is calculated with the weighting matrix $\tilde{W}_1 = (\tilde{Z}_f \tilde{Z}'_f)^{-\frac{1}{2}}$, and $d(n) = 2nm$ denotes the number of parameters, for $n = 0, 1, \dots, i-1$. Garcia–Hiernaux, Jerez, and Casals [106] show that the penalty function optimized through Monte Carlo simulations, $H(T, i) = e^{-2}T^{-9/10}i^{3/2}$, ensures strong consistency of the system order estimated by minimizing *NIDC*(n).

7.2.2 Model comparison methods

Other useful tools to determine the system order are the so-called information criteria, such as *AIC*, see Akaike [2], *SBC* (also known as *BIC*), see Schwarz [191], or *HQ*, see Hannan and Quinn [117], among others. These criteria are encompassed by the expression:

$$IC(n) = \frac{-2 \log l(n)}{T} + C(T)d(n), \quad (7.24)$$

where, $l(n)$ and $d(n)$ denote, respectively, the maximized log-likelihood value, and the number of parameters corresponding to a n -order system. $C(T)$ is again a penalty function that depends on the criterion; for instance, in *AIC*, $C(T) = 2/T$, in *SBC*, $C(T) = \log T/T$ and in *HQ*, $C(T) = 2 \log \log T/T$. These procedures are well known and widely employed; for recent references see for example Stoica, Yan, and Jian [202] or Bengtsson and Cavanaugh [20]. Their computational cost could be relevant for high-dimensional outputs, where the extra cost of one gradient step might be large (although it is considered acceptable in many cases).

Using the procedures introduced in this section, Garcia–Hiernaux, Jerez, and Casals [106] define an automatic model-building criterion based on the mode of the system order chosen by the previous *a priori* and *a posteriori* methods. A simulation analysis confirms that the mode-based criterion is more robust than any specific procedure alone. Its performance from a forecasting perspective is studied in Garcia–Hiernaux, Jerez, and Casals[103].

7.3 Constrained estimation

As the pros and cons of LS and ML are complementary, many works focus on devising methods that provide the best of both worlds. In particular, Spliid [198], Ko-reisha and Pukkila [143, 144], Flores and Serrano [97] and Dufour and Pelletier [81] extend LS and GLS to VARMAX modeling. Also Lutkepohl and Poskitt [159] propose an LS-based method to specify and estimate a canonical VARMA structure and Francq and Zakoian [99] propose weak GARCH representations which can be estimated by two-stage LS. All these methods share two substantial limitations. First, when there are exclusion constraints they require iterating, which partially offsets their computational advantage. Second, they are typically devised for specific parameterizations, such as VARMAX, and their implementation in other frameworks is not trivial in general.

In this section we present a fast and stable algorithm to estimate time series models written in their equivalent SS form. This procedure is useful, either to obtain adequate initial conditions for iterative ML, or to provide final estimates when ML is considered inadequate or too expensive. On the other hand, its SS foundation provides a wide scope, as it can be applied to any linear fixed-coefficients model with an equivalent SS form; see Chapter 9 and references therein for further details.

The algorithm is based on the subspace methods introduced in Section 7.1. Any exclusion or fixed-value constraint is treated in the same way. The procedure, termed SUBML from now on, provides a Gaussian ML solution to the subspace regressions. Iteration is required as there are constraints relating the parameters in the standard and SS representation of the model. Monte Carlo experiments show that SUBML is numerically more stable, computationally faster, and provides estimates close to those of ML in finite samples; see Garcia-Hiernaux, Jerez, and Casals [104].

The basic problem consists in estimating β , which is the vector of parameters in the original model (2.19)–(2.20). The system matrices Φ, Γ, K, H, D and Q will be, in general, nonlinear functions of β to be estimated by iterating on β . To this end, we will first discuss how to compute the state sequences defined in the previous section.

7.3.1 State sequence structure

Our starting point is the regression model (7.4). This expression implies that the expected value of future outputs conditional on U (which contains the past and future information of the observed inputs) and on Z_p (which contains the past values of the output), is:

$$E(Z_f|U, Z_p) = O_i E(X_f|U, Z_p) + T_{d,i} E(U_f|U, Z_p) + T_{s,i} E(A_f|U, Z_p). \quad (7.25)$$

Note that U_f is included in U . On the other hand, assumption A3 guarantees that the innovations are uncorrelated with the inputs and, by construction, $E(z_{t-j} a_t^T) = 0$, for all $j > 0$. Thus, (7.25) simplifies to:

$$E(Z_f|U, Z_p) = O_i E(X_f|U, Z_p) + T_{d,i} U_f. \quad (7.26)$$

By rearranging the terms in this equation, we obtain:

$$\mathbb{E}(\mathbf{Z}_f|\mathbf{U}, \mathbf{Z}_p) = \mathbf{O}_i^\dagger [\mathbb{E}(\mathbf{Z}_f|\mathbf{U}, \mathbf{Z}_p) - \mathbf{T}_{d,i}\mathbf{U}_f]. \quad (7.27)$$

As \mathbf{Z}_p and \mathbf{U} are observable, from (7.27) we can estimate the states using orthogonal projections:

$$\hat{\mathbf{X}}_f = \mathbf{O}_i^\dagger (\mathbf{Z}_f \mathbf{\Pi}_{U, Z_p} - \mathbf{T}_{d,i}\mathbf{U}_f), \quad (7.28)$$

where $\mathbf{\Pi}_{U, Z_p}$ is the orthogonal projection of the stacked matrix $[\mathbf{Z}_p^T \quad \mathbf{U}^T]^T$, and $\hat{\mathbf{X}}_f$ is an estimate of the state sequence conditional on the past of the output (\mathbf{Z}_p) as well as the entire input (\mathbf{U}). Parameters are included in matrices \mathbf{O}_i^\dagger and $\mathbf{T}_{d,i}$, which are nonlinear functions of β .

7.3.2 Subspace-based likelihood

Equation (7.4) implies that the error term can be written as:

$$\mathbf{T}_{s,i}\mathbf{A}_f = \mathbf{Z}_f - \mathbf{O}_i\mathbf{X}_f - \mathbf{T}_{i,d}\mathbf{U}_f, \quad (7.29)$$

where the states, \mathbf{X}_f , will be replaced with the expression given by equation (7.28).

The error term $\mathbf{T}_{s,i}\mathbf{A}_f$ has a particular structure, with one-step-ahead errors in the first row, two-step-ahead errors in the second row, and so on. On this basis, the Gaussian log-likelihood can then be written as,

$$\log l(\beta) = -\frac{im}{2} \log(2\pi) - \frac{i}{2} \log(|\Sigma|) - \frac{1}{2} \text{tr}(\mathbf{A}_f^T \mathbf{T}_{s,i}^T \Sigma^{-1} \mathbf{T}_{s,i} \mathbf{A}_f), \quad (7.30)$$

where $\text{tr}(\cdot)$ is the trace operator and Σ is the covariance matrix of the prediction errors. The noise term is obviously autocorrelated, and thus Σ has the following structure:

$$\Sigma = \mathbf{O}_i \mathbf{P}_i^* \mathbf{O}_i^T + \mathbf{T}_{s,i} (\mathbf{I}_i \otimes \mathbf{B}_i^*) \mathbf{T}_{s,i}^T, \quad (7.31)$$

where \mathbf{P}_i^* is the covariance matrix of the estimation error of the Kalman states and \mathbf{B}_i^* is the covariance of the Kalman innovations. Note that the first addend on the right-hand side of (7.31) refers to the error covariance of the states, while the second addend corresponds to the future output error variance, conditional on the estimated states. On account of the structure of $\mathbf{T}_{s,i}\mathbf{A}_f$, its covariance matrix is:

$$\Sigma := \begin{cases} \Sigma_{jk} \equiv \text{PE}_j \text{ covariance matrix,} & \text{when } j = k \\ \Sigma_{jk} \equiv \text{cov}(\text{PE}_j, \text{PE}_k), & \text{when } j \neq k \end{cases} \quad j, k = 1, 2, \dots, i, \quad (7.32)$$

where PE_j denotes the j -step-ahead prediction errors. Computing expression (7.31) requires propagating the Kalman filter covariance equations i times:

$$\hat{\mathbf{K}}_i^* = (\hat{\Phi} \hat{\mathbf{P}}_i^* \hat{\mathbf{H}}^T + \hat{\mathbf{E}} \hat{\mathbf{S}} \hat{\mathbf{C}}^T) \hat{\mathbf{B}}_i^{*-1}, \quad (7.33)$$

$$\hat{\mathbf{P}}_{i+1}^* = \hat{\Phi} \hat{\mathbf{P}}_i^* \hat{\Phi}^T + \hat{\mathbf{E}} \hat{\mathbf{Q}} \hat{\mathbf{E}}^T - \hat{\mathbf{K}}_i^* \hat{\mathbf{B}}_i^* \hat{\mathbf{K}}_i^{*T},$$

$$\hat{\mathbf{B}}_i^* = \hat{\mathbf{H}} \hat{\mathbf{P}}_i^* \hat{\mathbf{H}}^T + \hat{\mathbf{C}} \hat{\mathbf{R}} \hat{\mathbf{C}}^T,$$

in order to obtain an estimate of \mathbf{P}_i^* and \mathbf{B}_i^* . Note that $\hat{\mathbf{K}}_i^*$ is the Kalman filter gain introduced in Chapter 4.

This subspace-based maximum likelihood (SUBML) procedure requires the SS model to be in the SEM form (2.19)–(2.20). However, it is easy to extend it to the MEM (2.1)–(2.2) by just transforming the MEM structure in each iteration to the equivalent SEM form (using the method described in Subsection 2.2.3). Note that the likelihood value corresponding to two equivalent models is the same, and thus in practice it is indifferent whether the log-likelihood value is computed on the basis of the original MEM form or its equivalent SEM.

7.4 Multiplicative seasonal models

In models for seasonal data, the state vector typically has to undergo a substantial increase in size in order to capture the seasonal dynamic structure. This multiplies the computational cost of most estimation algorithms applied to these models. Additionally, seasonality also increases the dimension of the Block Hankel matrices (7.5) where data is organized, which is a second source of higher computational load.

We can cope with these drawbacks by decomposing a seasonal dynamic model (without deterministic inputs) into the following SEM subsystems:

$$\begin{aligned}\mathbf{x}_{s,t+s} &= \Phi_s \mathbf{x}_{s,t} + \mathbf{E}_s \mathbf{r}_t \\ \mathbf{z}_t &= \mathbf{H}_s \mathbf{x}_{s,t} + \mathbf{r}_t,\end{aligned}\tag{7.34}$$

and,

$$\begin{aligned}\mathbf{x}_{t+1} &= \Phi \mathbf{x}_t + \mathbf{E} \varepsilon_t \\ \mathbf{r}_t &= \mathbf{H} \mathbf{x}_t + \varepsilon_t.\end{aligned}\tag{7.35}$$

which can be combined into a single “cascade” formulation:

$$\begin{aligned}\begin{bmatrix} \mathbf{x}_{s,t+s} \\ \mathbf{x}_{t+1} \end{bmatrix} &= \begin{bmatrix} \Phi_s & \mathbf{E}_s \mathbf{H} \\ \mathbf{0} & \Phi \end{bmatrix} \begin{bmatrix} \mathbf{x}_{s,t} \\ \mathbf{x}_t \end{bmatrix} + \begin{bmatrix} \mathbf{E}_s \\ \mathbf{E} \end{bmatrix} \varepsilon_t \\ \mathbf{z}_t &= [\mathbf{H}_s \quad \mathbf{H}] \begin{bmatrix} \mathbf{x}_{s,t} \\ \mathbf{x}_t \end{bmatrix} + \varepsilon_t,\end{aligned}\tag{7.36}$$

where the state vector propagates simultaneously in two frequencies: 1 and s . From subsystem (7.35), it can be shown that $E(\mathbf{r}_{t+s} \mathbf{r}_t^T) = \mathbf{H} \Phi^{s-1} E(\mathbf{x}_{t+1} \mathbf{r}_t^T)$. Then, when s is much larger than the order of the subsystem (7.35), we have $E(\mathbf{r}_{t+s} \mathbf{r}_t^T) \simeq 0$, for all $i = s, 2s, 3s, \dots$, a result which holds with equality when the regular structure corresponds to a moving average of order q , and $q < s$. This property allows one to consider the input \mathbf{r}_t in the subsystem (7.34) as a white noise process in the frequency s and, therefore, uncorrelated for all t with the state vector $\mathbf{x}_{s,t}$.

To estimate the subsystem (7.34) using the subspace methods presented in Section 7.1, we define the s -frequency Block Hankel matrices:

$$\mathbf{Z}_p^s := \begin{bmatrix} z_1 & z_2 & \dots & z_{T-s(p+f)} \\ z_{s+1} & z_{s+2} & \dots & z_{T-s(p+f-1)} \\ \vdots & \vdots & & \vdots \\ z_{ps+1} & z_{ps+2} & \dots & z_{T-sf} \end{bmatrix}, \quad (7.37)$$

$$\mathbf{Z}_f^s := \begin{bmatrix} z_{s(p+1)+1} & z_{s(p+1)+2} & \dots & z_{T-s(f-1)} \\ z_{s(p+2)+1} & z_{s(p+2)+2} & \dots & z_{T-s(f-2)} \\ \vdots & \vdots & & \vdots \\ z_{s(p+f)+1} & z_{s(p+f)+2} & \dots & z_T \end{bmatrix}, \quad (7.38)$$

analogous to the *regular* Block Hankel matrices introduced in (7.5).

Therefore, when working with seasonal data, the estimation procedure would be:

1. Estimate the seasonal subsystem (7.34) using the above seasonal Block Hankel matrices, (7.37)–(7.38).
2. Compute the residuals, \hat{r}_t , from the previous subsystem.
3. Estimate the matrices in the regular subsystem (7.35), by means of \hat{r}_t organized in the Block Hankel matrices (7.5).

Note that the same strategy can be applied to estimate the system matrices using the subspace methods presented in Section 7.1, or the SUBML algorithm described in Section 7.3.

7.5 Examples

In this section we present some examples using the algorithms previously described. The first block illustrates the performance of these methods with univariate models. The second block is devoted to the identification of a multivariate model.

7.5.1 Univariate models

We first detail the steps to identify a univariate seasonal model using the well-known airline passengers data. Later, we will show the results obtained from applying the same methodology to nine time series commonly found in the literature.

The airline passengers time series was used by Box, Jenkins and Reinsel [30] to illustrate how to model seasonal data. These and other authors specify and estimate a double IMA structure. Figure 7.1 (top) depicts this data.

We start from the Box–Cox [28] transformed (logarithms of the) original series, and apply one regular and one seasonal difference as is usually done in the literature, $z_t = \nabla \nabla_{12} \log(y_t)$, where $\nabla_d = (1 - B^d)$. The identification (Box–Jenkins sense) of the number of regular and seasonal units roots is beyond the scope of this book. However, the reader can refer to Garcia–Hiernaux [102] and Garcia–Hiernaux, Jerez,

and Casals [105] to see how to cope with this problem from a subspace point of view. Table 7.2 displays the results of applying E^4 function `nid`, see appendices C and D, to z_t in order to identify the system order via the procedures presented in Section 7.2.

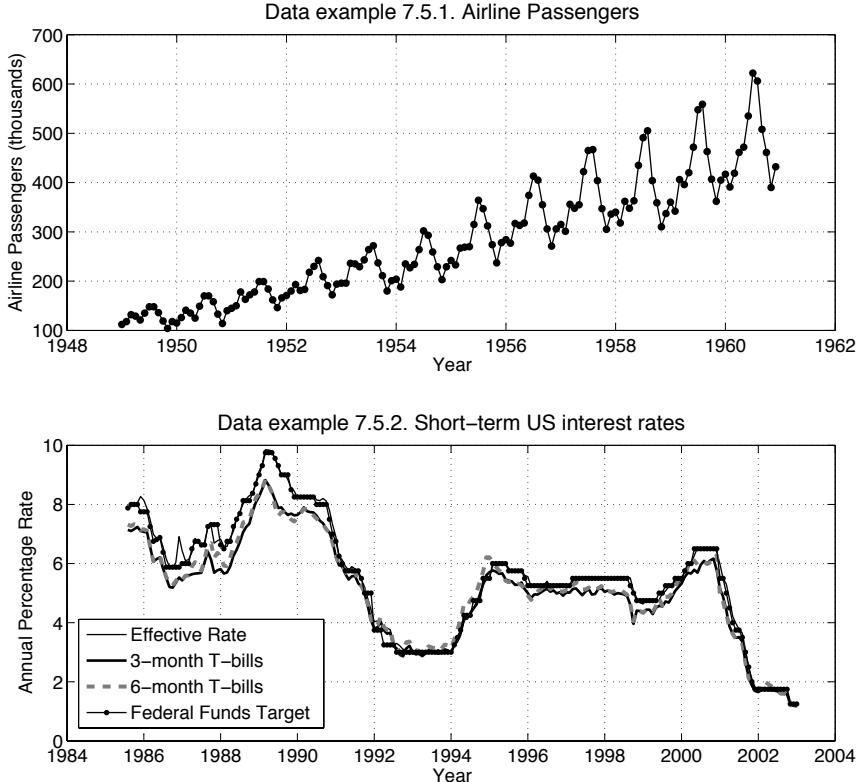


Figure 7.1 *Data in examples 7.5.1 and 7.5.2. Top: Airline passengers time series from Box, Jenkins and Reinsel [30]. Bottom: US short-term interest rates.*

All the procedures find a regular system order equal to one, i.e., $\hat{n}_r = 1$. To estimate the system order of the seasonal subsystem we also use `mid`, but with seasonal block Hankel matrices where the monthly frequency is set to $s = 12$. The output is shown in Table 7.3. All the procedures again find the same system order for the seasonal subsystem, i.e., $\hat{n}_s = 1$.

With these system orders, we then apply the subspace methods in cascade, see (7.34)–(7.35), obtaining the following balanced model:

$$(1 + .05 B)(1 - .08 B^{12})z_t = (1 - .35 B)(1 - .58 B^{12})\hat{a}_{1t}, \quad \hat{\sigma}_{a_1} = 3.68\%, \quad (7.39)$$

where standard errors for the parameters are given in parentheses below the corresponding estimates, and $\hat{\sigma}_{a_1}$ denotes the standard deviation of the residuals.

Table 7.2 *Regular system order estimation for z_t .*

Results for different methods and values of n						
n	AIC	SBC	HQ	SVC	NIDC	P-value χ^2
0	-3.319	-3.297	-3.310	0.196	0.244	-
1	-3.459	-3.393	-3.432	0.152	0.128	0.000
2	-3.433	-3.323	-3.388	0.208	0.150	0.066
3	-3.405	-3.252	-3.343	0.237	0.144	0.205

Chosen regular system order						
\hat{n}_r	1	1	1	1	1	1

Table 7.3 *Seasonal system order estimation for z_t with seasonal structure ($s = 12$) Block Hankel matrices.*

Results for different methods and values of n						
n	AIC	SBC	HQ	SVC	NIDC	P-value χ^2
0	-3.319	-3.300	-3.310	0.280	0.389	-
1	-3.566	-3.500	-3.540	0.108	0.079	0.013
2	-3.543	-3.433	-3.498	0.168	0.107	0.341
3	-3.513	-3.359	-3.450	0.235	0.143	0.341

Chosen seasonal system order						
\hat{n}_s	1	1	1	1	1	1

If the user wishes to prune the model of non-significant parameters and obtain, with a low computational cost, a more precise estimation (or good initial conditions for an ML estimation), it may be useful to employ the SUBML algorithm. Doing so provides the following results:

$$z_t = (1 - .39B)(1 - .52B^{12})\hat{a}_{2t}, \quad \hat{\sigma}_{a_2} = 3.68\%. \quad (7.40)$$

Regardless of when a decision is made to stop, (7.39), (7.40), or even after a full-blown ML estimation, it is recommended that a diagnostic exercise be carried out. To illustrate, we apply `nid` once again to the residuals from model (7.40), for both the $s = 1$ and $s = 12$ frequencies. Tables 7.4 and 7.5 reveal that there is no additional regular or seasonal dynamical structure, because the system order estimated by all the criteria (except NIDC in Table 7.4) is zero.

The residuals \hat{a}_{2t} do not display any evidence of misspecification, either in the regular or seasonal components of the model. The model identified by the subspace algorithm corresponds to the structure typically found in the literature. ML estimation using the procedures detailed in Chapter 5 yields:

$$z_t = (1 - .40B)_{(.08)}(1 - .56B^{12})_{(.08)}\hat{a}_t, \quad \hat{\sigma}_a = 3.67\%. \quad (7.41)$$

Table 7.4 *Regular system order estimation for \hat{a}_{2t} .*

Results for different methods and values of n						
n	AIC	SBC	HQ	SVC	NIDC	P-value χ^2
0	-3.766	-3.744	-3.757	0.108	0.120	-
1	-3.757	-3.691	-3.730	0.134	0.107	0.365
2	-3.729	-3.619	-3.684	0.175	0.114	0.234
3	-3.711	-3.558	-3.649	0.229	0.136	0.434
Chosen regular system order						
\hat{n}_r	0	0	0	0	1	0

Table 7.5 *Seasonal system order estimation for \hat{a}_{2t} with the seasonal Block Hankel matrix and $s = 12$.*

Results for different methods and values of n						
n	AIC	SBC	HQ	SVC	NIDC	P-value χ^2
0	-3.766	-3.744	-3.757	0.043	0.044	-
1	-3.736	-3.670	-3.709	0.113	0.083	0.292
2	-3.724	-3.614	-3.679	0.161	0.099	0.660
3	-3.693	-3.539	-3.631	0.233	0.140	0.762
Chosen seasonal system order						
\hat{n}_s	0	0	0	0	0	0

As expected, the ML estimates are very close to those obtained from SUBML. Obviously, in univariate systems the advantage of using SUBML is not as large as in multivariate systems, but it is interesting to see how the mechanical application of these methods simplifies the identification process (specification and estimation), and, in particular, how they allow one to perform the univariate analysis of massive amounts of time series with a relatively small computational and human cost.

The code required to replicate this example can be found in the file `Example0751.m`; see Appendix D.

We now present the results of applying the same analysis to nine univariate time series. Data analyzed is briefly described in Table 7.6, while Tables 7.7, 7.8, and 7.9 compare the results reported by the literature with those obtained by sequentially applying a subspace method (SM), the SUBML algorithm, and ML estimation. The only significant difference seen is in the series H, where the results produced by SUBML are close to those of Wei [221]. However, when estimating the model by ML, the moving average parameter converges to unity. As the sample is very short ($T = 32$), this could be an example of what is called in the literature the “Pile-up Ef-

fect”⁶. Regarding this matter, Garcia–Hiernaux, Jerez, and Casals [104] shows that: (i) the algorithm SUBML described in Section 7.3 is less sensitive than ML to the pile-up effect, and (ii) discarding this effect, estimates of the moving average parameters are very similar for SUBML and ML.

Hence, the sequential use of SM, SUBML, and ML offers flexible alternatives for model estimation, as it allows one to attain final results that are consistently similar, depending on the modeling goals and the resources available. For example, even though we may sacrifice using SM in lieu of ML because of computational considerations, we can have some assurance that the results will not be too different.

7.5.2 A multivariate model for the interest rates

This example shows the performance of the previously discussed methods in a multivariate framework. Consider now four short-term interest rates of the US economy. Specifically, we consider the Federal Funds Target (FT_t), the Federal Funds effective rate (ER_t), and the 3-month and 6-month T-bills in the secondary market ($T3m_t$ and $T6m_t$, respectively), in monthly averages of yields for the period August 1985 to January 2003. The four series are depicted in Figure 7.1 (bottom).

We assume ER_t , $T3m_t$, and $T6m_t$ to be I(1) and CI(1,1) cointegrated with FT_t . Although the unit root, cointegrating rank, and cointegrating parameter estimation can be performed using subspace-based methods, see Bauer and Wagner [18], Garcia–Hiernaux [102], Garcia–Hiernaux, Jerez, and Casals [105], this is not the purpose of this book and therefore it is not shown here. We will therefore analyze a stationary multivariate process representing the spread of some short-term interest rates defined as⁷:

$$\mathbf{Y}_{1t} = \begin{bmatrix} \nabla & 0 & 0 & 0 \\ -- 1.01 & 1 & 0 & 0 \\ -- 0.92 & 0 & 1 & 0 \\ -- 0.93 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} FT_t \\ ER_t \\ T3m_t \\ T6m_t \end{bmatrix}. \quad (7.42)$$

When studying the dynamics of the multivariate process \mathbf{Y}_{1t} , most of the procedures yield an estimated system order of 4; see Table 7.10. Therefore, we proceed to estimate, via subspace methods, a model for \mathbf{Y}_{1t} with $\hat{n} = 4$ (the mode of the chosen system order), which implies an unrestricted VARMA(1,1) structure. In this first estimation we found ten coefficients that are not significantly different from zero. Those coefficients were then removed from the model, which was then re-estimated using the SUBML algorithm. Table 7.11 summarizes the results produced by these two steps. The values of the parameters vary slightly, getting closer (as we will see later) to the ML estimates. This second step yields adequate initial conditions for an ML estimation. Note that the relevance of adequate initial conditions grows with

⁶This effect means that the distribution of ML estimates for moving average coefficients is multimodal, with one of the modes on the invertibility boundary; see Ansley and Newbold [10], Davis and Song [65], and Paige, Trindade, and Wickramasinghe [174].

⁷See Garcia–Hiernaux [102] for details about the estimated values of the cointegration matrix.

Table 7.6 Datasets used to compare the SM, SUBML and ML methods in Tables 7.7, 7.8, and 7.9.

Time series ID	# obs.	Reference and description
A	45	[35]: Daily Average Number of Truck Manufacturing Defects.
B*	150	[221]: Simulated Quarterly Time Series
C	35	[221]: Monthly Unemployed Females between 16 and 19 in the USA, 01/1961-12/1985.
D*	200	[227]: Simulated Time Series # 2.
E	226	[30]: Series C - Chemical Process Temperature, readings every minute.
F	197	[30]: Series A - Chemical Process Concentration, readings every two hours.
G	178	[170]: Wisconsin Monthly Employment Time Series, 01/1961-10/1975.
H	32	[221]: Quarterly USA Beer Production, 01/1975-04/1982.
I*	36	[160]: Simulated Series.

* These are simulated series. However, while in Tables 7.7 and 7.8, series B and D present the data generating processes, series I presents the specification and estimation performed by the author.

Table 7.7 Univariate models. Comparison of models estimated for the data described in Table 7.6 - Part I. Standard errors are in parenthesis. $\hat{\sigma}_a^2$ denotes the residual variance. SM is a subspace methods estimation with weights corresponding to the CVA algorithm (see Table 7.1) and using an extended observability matrix approach (see Section 7.1).

Series	Obs	Specification and estimation by the literature	SM
A	45	$(1 - .43B)(Z_t - 1.79) = a_t$ $\hat{\sigma}_a^2 = .21$	$(1 - .61B)(Z_t - 3.94) = (1 - .24B)a_t$ $\hat{\sigma}_a^2 = .21$
B	150	$\nabla \nabla_4 Z_t = (1 - .80B)(1 - .60B^4)a_t$ $\hat{\sigma}_a^2 = 1.00$	$(1 + .22B)(1 + .21B^4)\nabla \nabla_4 Z_t = (1 - .56B)(1 - .49B^4)a_t$ $\hat{\sigma}_a^2 = .97$
C	35	$\nabla Z_t = (1 - .51B)a_t$ $\hat{\sigma}_a^2 = 1397.27$	$(1 - .00B)\nabla Z_t = (1 - .51B)a_t$ $\hat{\sigma}_a^2 = 1380.35$
D	200	$(1 - .70B + .49B^2)Z_t = a_t$ $\hat{\sigma}_a^2 = .60$	$(1 - .84B + .51B^2)Z_t = (1 - .16B + .05B^2)a_t$ $\hat{\sigma}_a^2 = .66$
E	226	$(1 - .82B)\nabla Z_t = a_t$ $\hat{\sigma}_a^2 = .02$	$(1 - .82B)\nabla Z_t = (1 - .02B)a_t$ $\hat{\sigma}_a^2 = .02$
F	197	$(1 - .92B)Z_t = 1.45 + (1 - .58B)a_t$ $\hat{\sigma}_a^2 = .10$	$(1 - .88B)Z_t = 2.08 + (1 - .50B)a_t$ $\hat{\sigma}_a^2 = .10$
G	178	$(1 - .79B)\nabla_{12}Z_t = (1 - .76B^{12})a_t$ $\hat{\sigma}_a^2 = 1.24$	$(1 - .56B)(1 - .05B^{12})\nabla_{12}Z_t = -2.11 + (1 - .18B)(1 - .35B^{12})a_t$ $\hat{\sigma}_a^2 = 1.54$
H	32	$\nabla_4 Z_t = 1.49 + (1 - .87B^4)a_t$ $\hat{\sigma}_a^2 = 2.39$	$(1 - \frac{14}{(26)}B^4)\nabla_4 Z_t = 1.21 + (1 - \frac{71}{(23)}B^4)a_t$ $\hat{\sigma}_a^2 = 2.03$
I	36	$Z_t = 51.03 + a_t$ $\hat{\sigma}_a^2 = 818.81$	$Z_t = 47.31 + a_t$ $\hat{\sigma}_a^2 = 790.54$

Table 7.8 Univariate models. Comparison of models estimated for the data described in Table 7.6 - Part 2. Standard errors are in parenthesis. $\hat{\sigma}_u^2$ denotes the residual variance. SUBML values were obtained with the algorithm described in Section 7.3, by pruning the non-significant parameters found in Table 7.7.

Series	Obs	Specification and estimation by the literature	SIM + SUBML
A	45	$(1 - \frac{43}{13}B)(Z_t - 1.79) = a_t$ $\hat{\sigma}_u^2 = .21$	$(1 - .35B)(Z_t - 1.87) = a_t$ $\hat{\sigma}_u^2 = .21$
B	150	$\nabla \nabla_4 Z_t = (1 - .80B)(1 - .60B^4)a_t$ $\hat{\sigma}_u^2 = 1.00$	$\nabla \nabla_4 Z_t = (1 - .74B)(1 - .66B^4)a_t$ $\hat{\sigma}_u^2 = .88$
C	35	$\nabla Z_t = (1 - .51B) + a_t$ $\hat{\sigma}_u^2 = 1397.27$	$\nabla Z_t = (1 - .50B) + a_t$ $\hat{\sigma}_u^2 = 1393.54$
D	200	$(1 - .70B + .49B^2)Z_t = a_t$ $\hat{\sigma}_u^2 = .60$	$(1 - .71B + .42B^2)Z_t = a_t$ $\hat{\sigma}_u^2 = .66$
E	226	$(1 - .82B)\nabla Z_t = a_t$ $\hat{\sigma}_u^2 = .02$	$(1 - .81B)\nabla Z_t = a_t$ $\hat{\sigma}_u^2 = .02$
F	197	$(1 - .92B)Z_t = 1.45 + (1 - .58B)a_t$ $\hat{\sigma}_u^2 = .10$	$(1 - .88B)Z_t = 2.07 + (1 - .51B)a_t$ $\hat{\sigma}_u^2 = .10$
G	178	$(1 - .79B)\nabla_{12}Z_t = (1 - .76B^{12})a_t$ $\hat{\sigma}_u^2 = 1.24$	$(1 - .68B)\nabla_{12}Z_t = .07 + (1 - .56B^{12})a_t$ $\hat{\sigma}_u^2 = 1.37$
H	32	$\nabla_4 Z_t = 1.49 + (1 - .87B^4)a_t$ $\hat{\sigma}_u^2 = 2.39$	$\nabla_4 Z_t = 1.36 + (1 - .59B^4)a_t$ $\hat{\sigma}_u^2 = 2.15$
I	36	$Z_t = 51.03 + a_t$ $\hat{\sigma}_u^2 = 818.81$	$Z_t = 51.03 + a_t$ $\hat{\sigma}_u^2 = 796.68$

Table 7.9 *Univariate models. Comparison of models estimated for the data described in Table 7.6 - Part 3. Standard errors are in parenthesis. $\hat{\sigma}_a^2$ denotes the residual variance. ML estimates were initiated with the parameters obtained by SUBML shown in Table 7.8.*

Series	Obs	Specification and estimation by the literature	SM + SUBML + ML
A	45	$(1 - .43B)(Z_t - .79) = a_t$ $(.13)$ $\hat{\sigma}_a^2 = .21$	$(1 - .42B)(Z_t - 1.8) = a_t$ $(.15)$ $\hat{\sigma}_a^2 = .21$
B	150	$\nabla \nabla_4 Z_t = (1 - .80B)(1 - .60B^4)a_t$	$\nabla \nabla_4 Z_t = (1 - .75B)(1 - .63B^4)a_t$ $(.06)$ $\hat{\sigma}_a^2 = .91$
C	35	$\nabla Z_t = (1 - .51B)a_t$ $(.05)$ $\hat{\sigma}_a^2 = 1397.27$	$\nabla \nabla_4 Z_t = (1 - .51B)a_t$ $(.05)$ $\hat{\sigma}_a^2 = 1393.12$
D	200	$(1 - .70B + .49B^2)Z_t = a_t$	$(1 - .72B + .41B^2)Z_t = a_t$ $(.06)$ $\hat{\sigma}_a^2 = .66$
E	226	$(1 - .82B)\nabla Z_t = a_t$ $(.04)$ $\hat{\sigma}_a^2 = .02$	$(1 - .82B)\nabla Z_t = a_t$ $(.04)$ $\hat{\sigma}_a^2 = .02$
F	197	$(1 - .92B)Z_t = 1.45 + (1 - .58B)a_t$ $(.04)$ $\hat{\sigma}_a^2 = .10$	$(1 - .92B)Z_t = 1.45 + (1 - .60B)a_t$ $(.04)$ $\hat{\sigma}_a^2 = .10$
G	178	$(1 - .79B)\nabla_{12}Z_t = (1 - .76B^{12})a_t$ $(.05)$ $\hat{\sigma}_a^2 = 1.24$	$(1 - .77B)\nabla_{12}Z_t = (1 - .61B^{12})a_t$ $(.05)$ $\hat{\sigma}_a^2 = 1.31$
H	32	$\nabla_4 Z_t = 1.49 + (1 - .87B^4)a_t$ $(.09)$ $\hat{\sigma}_a^2 = 2.39$	$\nabla_4 Z_t = 1.55 + \nabla_4 a_t$ $(.13)$ $\hat{\sigma}_a^2 = 1.64$
I	36	$Z_t = 51.03 + a_t$ (4.84) $\hat{\sigma}_a^2 = 818.81$	$Z_t = 51.03 + a_t$ (4.84) $\hat{\sigma}_a^2 = 818.86$

the complexity of the model. In the third step, the calculation of the exact information matrix and the ML estimates allows the user to conduct more accurate tests, and thereby remove some parameters whose lack of significance was not previously detected.

Table 7.10 *System order estimation for \mathbf{Y}_{1t} .*

Results for different methods and values of n						
n	AIC	SBC	HQ	SVC	NIDC	P-value χ^2
2	-4.036	-3.621	-3.868	1.076	2.305	0.000
3	-4.636	-4.093	-4.416	1.117	1.474	0.000
4	-4.877	-4.205	-4.605	1.230	1.310	0.000
5	-4.811	-4.011	-4.488	1.379	1.316	0.304

Chosen system order						
\hat{n}	4	4	4	2	4	4

In the ensuing diagnostic analysis, three new parameters are specified and the final model results in a restricted VARMA (3,1). The sequence of estimations is shown in Table 7.11. The covariance matrix of the residuals is found to be:

$$\hat{\Sigma}_{\hat{a}_1}(\%) = \begin{bmatrix} 3.33 & - & - & - \\ (.33) & & & \\ -1.59 & 2.30 & - & - \\ (.22) & (.23) & & \\ -1.33 & 1.35 & 2.84 & - \\ (.23) & (.20) & (.28) & \\ -.83 & 0.98 & 2.43 & 2.87 \\ (.22) & (.19) & (.26) & (.28) \end{bmatrix},$$

where \hat{a}_{1t} are the residuals obtained from the ML estimates. The Ljung–Box [151] statistics are given by:

$$Q_{(7)}(\%) = \begin{bmatrix} 5.8 & 10.3 & 5.4 & 8.9 \\ 8.6 & 7.7 & 4.3 & 9.9 \\ 10.6 & 17.0 & 11.1 & 12.6 \\ 6.5 & 15.5 & 8.3 & 10.1 \end{bmatrix}.$$

This example illustrates the ability to identify a multivariate process using the procedures presented in this chapter. The initial specification deviates slightly from that of the final model, likely due to the (barely significant) persistence of the second and third order autoregressive components in the $\hat{\Phi}(1, 1)$ element. However, note that application of the first two steps with SM and SUBML already yields a reasonable approximation to the final model; these two steps being obtained at low computational and human cost.

The code required to replicate this example can be found in the file Example0752.m; see Appendix D.

Table 7.11 Sequence of identified models for the process \mathbf{Y}_{1t} . The figures in parentheses are the standard errors of the parameters and the symbol “_” denotes that the corresponding parameter has been constrained to its current value, so it does not have a standard error.

Method	$\hat{\Phi}(\mathbf{B})$	$\hat{\Theta}(\mathbf{B})$
SM	$\begin{bmatrix} 1 - .49\mathbf{B} & .19\mathbf{B} & .45\mathbf{B} & -.35\mathbf{B} \\ (.08) & (.18) & (.15) & (.12) \\ .20\mathbf{B} & 1 - .29\mathbf{B} & .18\mathbf{B} & .11\mathbf{B} \\ (.04) & (.10) & (.08) & (.06) \\ .32\mathbf{B} & .13\mathbf{B} & 1 - .80\mathbf{B} & -.10\mathbf{B} \\ (.06) & (.14) & (.11) & (.09) \\ .23\mathbf{B} & .0\mathbf{B} & -.02\mathbf{B} & 1 - .94\mathbf{B} \\ (.08) & (.18) & (.15) & (.12) \end{bmatrix}$	$\begin{bmatrix} 1 + .04\mathbf{B} & .36\mathbf{B} & 1.36\mathbf{B} & -.53\mathbf{B} \\ (.08) & (.15) & (.16) & (.13) \\ .20\mathbf{B} & 1 + .09\mathbf{B} & .10\mathbf{B} & .20\mathbf{B} \\ (.06) & (.10) & (.10) & (.09) \\ .40\mathbf{B} & -.06\mathbf{B} & 1 - .29\mathbf{B} & -.06\mathbf{B} \\ (.07) & (.12) & (.12) & (.11) \\ .35\mathbf{B} & -.30\mathbf{B} & -.02\mathbf{B} & 1 - .15\mathbf{B} \\ (.09) & (.16) & (.17) & (.14) \end{bmatrix}$
SUBML	$\begin{bmatrix} 1 - .50\mathbf{B} & 0 & .40\mathbf{B} & -.30\mathbf{B} \\ .12\mathbf{B} & 1 - .54\mathbf{B} & -.03\mathbf{B} & 0 \\ .21\mathbf{B} & 0 & 1 - .85\mathbf{B} & 0 \\ .18\mathbf{B} & 0 & 0 & 1 - .86\mathbf{B} \end{bmatrix}$	$\begin{bmatrix} 1 & .20\mathbf{B} & 1.10\mathbf{B} & -.28\mathbf{B} \\ .21\mathbf{B} & 1 & -.15\mathbf{B} & -.01\mathbf{B} \\ .34\mathbf{B} & 0 & 1 - .18\mathbf{B} & 0 \\ .38\mathbf{B} & 0 & 0 & 1 \end{bmatrix}$
ML	$\begin{bmatrix} 1 - .46\mathbf{B} - .04\mathbf{B}^2 - .19\mathbf{B}^3 & 0 & .19\mathbf{B} & -.10\mathbf{B} \\ (.06) & (.05) & (.06) & (.05) \\ .16\mathbf{B} & 1 - .66\mathbf{B} & 0 & 0 \\ (.05) & (.08) & (-) & (-) \\ .19\mathbf{B} & 0 & 1 - .87\mathbf{B} & 0 \\ (.05) & (-) & (.03) & (-) \\ .23\mathbf{B} & 0 & 0 & 1 - .89\mathbf{B} \\ (.06) & (-) & (-) & (.03) \end{bmatrix}$	$\begin{bmatrix} 1 & .30\mathbf{B} & .67\mathbf{B} & 0 \\ (.08) & (.08) & (.09) & (-) \\ .28\mathbf{B} & 1 - .23\mathbf{B} & 0 & 0 \\ (.08) & (.11) & (-) & (-) \\ .28\mathbf{B} & 0 & 1 - .26\mathbf{B} & 0 \\ (.07) & (-) & (.06) & (-) \\ .43\mathbf{B} & 0 & 0 & 1 - .19\mathbf{B} \\ (.08) & (-) & (-) & (.06) \end{bmatrix}$

This page intentionally left blank

Chapter 8

Signal extraction

8.1	Input and error-related components	108
8.1.1	The deterministic and stochastic subsystems	108
8.1.2	Enforcing minimality	109
8.2	Estimation of the deterministic components	110
8.2.1	Estimating the total effect of the inputs	111
8.2.2	Estimating the individual effect of each input	113
8.3	Decomposition of the stochastic component	114
8.3.1	Characterization of the structural components	114
8.3.2	Estimation of the structural components	116
8.4	Structure of the method	116
8.5	Examples	116
8.5.1	Comparing different methods with simulated data	116
8.5.2	Common features in wheat prices	120
8.5.3	The effect of advertising on sales	123

The term “signal extraction” refers to the definition and estimation of meaningful signals that are buried in a vector of time series. In the time series literature, it often refers to the decomposition of a single time series (z_t) into the sum of “structural components” named as “trend” (t_t), “cycle” (c_t), “seasonal” (s_t) and “irregular” (ε_t):

$$z_t = t_t + c_t + s_t + \varepsilon_t. \quad (8.1)$$

In this chapter we describe a signal extraction method for a vector of time series, z_t , which is the output of a SEM (2.19)–(2.20). Building on an SS model as the starting point has a clear advantage in terms of generality, as it includes as particular cases common specifications; see Chapter 2.

On the other hand, the methodology described here has some convenient features. In particular it:

1. . . allows for both, single and multiple time series,

2. ... considers a component not included in (8.1) which is driven by exogenous variables
3. ... is “revision-free,” meaning that the estimates for the components do not change when the sample size increases, as happens with most alternative procedures including those of Burman [34], Hillmer and Tiao [125], or Harvey [118]
4. ... does nor require specifying *a priori* models for the components.

8.1 Input and error-related components

Let \mathbf{z}_t be an $m \times 1$ random vector including all the series to be decomposed and assume that it is realized by the minimal SEM model:

$$\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{K} \mathbf{a}_t \quad (8.2)$$

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_t + \mathbf{D} \mathbf{u}_t + \mathbf{a}_t, \quad (8.3)$$

which coincides with the formulation (2.19)–(2.20) that was already defined in Chapter 2. We will further consider that the model (8.2)–(8.3) has been transformed so that the matrix Φ is block-diagonal, see Subsection 3.1.3 and Appendix A.

8.1.1 The deterministic and stochastic subsystems

Following the procedure described in Subsection 3.1.1, we can decompose model (8.2)–(8.3) into the corresponding deterministic and stochastic subsystems:

$$\mathbf{x}_{t+1}^d = \Phi \mathbf{x}_t^d + \Gamma \mathbf{u}_t \quad (8.4)$$

$$\mathbf{z}_t^d = \mathbf{H} \mathbf{x}_t^d + \mathbf{D} \mathbf{u}_t \quad (8.5)$$

$$\mathbf{x}_{t+1}^s = \Phi \mathbf{x}_t^s + \mathbf{K} \mathbf{a}_t \quad (8.6)$$

$$\mathbf{z}_t^s = \mathbf{H} \mathbf{x}_t^s + \mathbf{a}_t. \quad (8.7)$$

Note that this decomposition preserves the block-diagonal structure of the system as the transition matrices in (8.4) and (8.6) coincide with that in (8.2).

Building on the deterministic-stochastic decomposition, Section 8.2 discusses the calculation of the deterministic component \mathbf{z}_t^d , which is the output of (8.4)–(8.5).

Estimating the deterministic component is useful in many situations. For example, if the inputs are control variables, their effect on the output provides a measure of the influence of past controls. On the other hand, if the inputs are leading indicators, then their individual influence shows how each one affects the forecasts for \mathbf{z}_t . Last, if the variables in \mathbf{u}_t are intervention terms (Box and Tiao [31]) or calendar variables (Cleveland and Devlin [54]), such a decomposition provides estimates for the individual effect of each input, and, if required, the means to eliminate them.

After determining the deterministic component, the stochastic term is trivially given by: $\mathbf{z}_t^s = \mathbf{z}_t - \mathbf{z}_t^d$. In Section 8.3 we will show how to further decompose this component into the trend, cycle, seasonal, and irregular components implied by model (8.6)–(8.7).

Putting all these ideas together, the decomposition under consideration can be written as:

$$z_t = z_t^d + z_t^s, \quad (8.8)$$

$$z_t^s = t_t + c_t + s_t + \varepsilon_t, \quad (8.9)$$

where t_t , c_t , s_t and ε_t are $m \times 1$ vectors of trend, cycle, seasonal and irregular components, respectively.

8.1.2 Enforcing minimality

It is easy to show that, if the SEM model (8.2)–(8.3) is minimal, then the deterministic and stochastic components, z_t^d and z_t^s , are observable. On the other hand, decomposing the SEM into its deterministic and stochastic subsystems replicates the state equation because expressions (8.4) and (8.6) can be jointly written as:

$$\begin{bmatrix} x_{t+1}^d \\ x_{t+1}^s \end{bmatrix} = \begin{bmatrix} \Phi & 0 \\ 0 & \Phi \end{bmatrix} \begin{bmatrix} x_t^d \\ x_t^s \end{bmatrix} + \begin{bmatrix} \Gamma \\ 0 \end{bmatrix} u_t + \begin{bmatrix} 0 \\ K \end{bmatrix} a_t, \quad (8.10)$$

and adding (8.5) and (8.7) yields, bearing in mind that $z_t = z_t^d + z_t^s$, the following observation equation:

$$z_t = [H \quad H] \begin{bmatrix} x_t^d \\ x_t^s \end{bmatrix} + D u_t + a_t. \quad (8.11)$$

Note that this model realizes z_t using $2n$ states, while (8.2)–(8.3) only requires n . Therefore (8.10)–(8.11) is a non-minimal system whose output is the sum of two components with identical dynamics. In most cases these components cannot be distinguished. However, such a distinction is possible in this case because the input of the deterministic subsystem, u_t , is exactly known. As we will see in Section 8.2, to estimate z_t^d we will only need adequate initial conditions for the deterministic subsystem.

On the other hand, minimality of both subsystems is not assured, even if their outputs are observable. This happens because controllability of (8.2)–(8.3) does not assure controllability of (8.4)–(8.5) and (8.6)–(8.7)¹.

Uncontrollable modes are redundant and wreak havoc on standard signal extraction algorithms. Fortunately, it is easy to detect and eliminate them by applying the Staircase algorithm (Rosenbrock [190]) to each subsystem. The following example illustrates the previous discussion and the elimination of uncontrollable modes.

Example 8.1.1 Elimination of uncontrollable modes

Consider the SISO transfer function:

$$z_t = \frac{.5}{1 - .6B} u_t + \frac{1}{1 - B} a_t, \quad (8.12)$$

¹For example, if part of the model dynamics is uniquely associated with the inputs, as happens for example in a transfer function, the states describing these dynamics are not excited in the stochastic subsystem, and, as a consequence, this subsystem is uncontrollable.

which can be written as a block-diagonal SEM (see Subsection 3.1.3 and Appendix A) as:

$$\begin{bmatrix} x_{t+1}^1 \\ x_{t+1}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & .6 \end{bmatrix} \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + \begin{bmatrix} 0 \\ -.103 \end{bmatrix} u_t + \begin{bmatrix} 1.166 \\ 0 \end{bmatrix} a_t, \quad (8.13)$$

$$z_t = [.858 \quad -2.916] \begin{bmatrix} x_t^1 \\ x_t^2 \end{bmatrix} + .5 u_t + a_t. \quad (8.14)$$

According to (8.4)–(8.5) and (8.6)–(8.7), the deterministic and stochastic subsystems are given respectively by:

$$\begin{bmatrix} x_{t+1}^{d1} \\ x_{t+1}^{d2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & .6 \end{bmatrix} \begin{bmatrix} x_t^{d1} \\ x_t^{d2} \end{bmatrix} + \begin{bmatrix} 0 \\ -.103 \end{bmatrix} u_t \quad (8.15)$$

$$z_t^d = [.858 \quad -2.916] \begin{bmatrix} x_t^{d1} \\ x_t^{d2} \end{bmatrix} + .5 u_t \quad (8.16)$$

$$\begin{bmatrix} x_{t+1}^{s1} \\ x_{t+1}^{s2} \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & .6 \end{bmatrix} \begin{bmatrix} x_t^{s1} \\ x_t^{s2} \end{bmatrix} + \begin{bmatrix} 1.166 \\ 0 \end{bmatrix} a_t \quad (8.17)$$

$$z_t^s = [.858 \quad -2.916] \begin{bmatrix} x_t^{s1} \\ x_t^{s2} \end{bmatrix} + a_t. \quad (8.18)$$

Note that the first state in (8.15) and the second state in (8.17) will never be excited, as they do not depend on the values of the input and the error, respectively. Therefore, they are uncontrollable and the overall system is not minimal. Applying the Staircase algorithm to (8.15)–(8.16) and (8.17)–(8.18), yields two minimal equivalent representations:

$$x_{t+1}^d = .6 x_t^d - .103 u_t \quad (8.19)$$

$$z_t^d = -2.916 x_t^d + .5 u_t, \quad (8.20)$$

and

$$x_{t+1}^s = x_t^s + 1.166 a_t \quad (8.21)$$

$$z_t^s = .858 x_t^s + a_t. \quad (8.22)$$

8.2 Estimation of the deterministic components

Assuming that the matrices for the minimal deterministic and stochastic subsystems have been computed, we will now address the problem of estimating the sequences z_t^s and z_t^d , for a given sample $\{z_t, u_t\}_{t=1}^N$. To this end, note that the solution for x_t^d in (8.4) is given by:

$$x_t^d = \Phi^{t-1} x_1^d + \sum_{i=1}^{t-1} \Phi^{t-1-i} \Gamma u_i. \quad (8.23)$$

If we had an adequate initial condition for the deterministic subsystem $\hat{x}_{1|N}^d$, computation of the decomposition would then reduce to:

1. propagating (8.23) to obtain the sequence $\{\hat{x}_{t|N}^d\}_{t=2}^N$;
2. estimating the output of the deterministic subsystem which, by (8.5), would be

$$\hat{z}_{t|N}^s = \mathbf{H} \hat{x}_{t|N}^d + \mathbf{D} u_t; \quad (8.24)$$

3. estimating the output of the stochastic subsystem using (8.8)

$$\hat{z}_{t|N}^s = z_t - \hat{z}_{t|N}^d. \quad (8.25)$$

De Jong [66] shows that one way to isolate the effect of the initial state x_1 on the innovations is computing the sequence:

$$\tilde{z}_t = \mathbf{H} \bar{\Phi}_{t-1} x_1 + \tilde{z}_t^*, \quad (8.26)$$

where \tilde{z}_t are the innovations resulting from an arbitrary KF($\mathbf{0}, \mathbf{0}$), \tilde{z}_t^* are the KF innovations that would result if the true initial conditions were null, and $\bar{\Phi}_t$ is the matrix sequence resulting from $\bar{\Phi}_t = (\Phi - K_t H) \bar{\Phi}_{t-1}$ with $\bar{\Phi}_1 = I$.

Expression (8.26) can be written in matrix form as:

$$\tilde{Z} = \mathbf{X} x_1 + \tilde{Z}^*, \quad (8.27)$$

where \tilde{Z} and \tilde{Z}^* are $(m \times N) \times 1$ vectors containing the innovations \tilde{z}_t and \tilde{z}_t^* , respectively, and \mathbf{X} is a matrix whose row-blocks contain the values $H \bar{\Phi}_{t-1}$.

Taking into account that $x_1 = x_1^d + x_1^s$, it would be natural to write (8.27) as $\tilde{Z} = \mathbf{X} (x_1^d + x_1^s) + \tilde{Z}^* = \mathbf{X} x_1^d + \mathbf{X} x_1^s + \tilde{Z}^*$. However, if we only take into account the controllable modes of the subsystems (see Section 8.1) the matrices affecting x_1^d and x_1^s in this expression would be different. Accordingly, expression (8.27) can be written as:

$$\tilde{Z} = \mathbf{X}^d x_1^d + \mathbf{X}^s x_1^s + \tilde{Z}^*, \quad (8.28)$$

and the problem of estimating the deterministic component reduces to determining the initial conditions x_1^d , as we have seen above.

8.2.1 Estimating the total effect of the inputs

The initial conditions for the deterministic subsystem can be determined either by assuming that they are diffuse, or by computing the Generalized Least Squares (GLS) estimate of a non-zero initial state, x_1^d , with a null covariance. As we will see, both of them are equivalent.

Diffuse initialization implies assuming that $x_1^d \sim (\bar{x}_1^d, P_1^d)$, with $(P_1^d)^{-1} = \mathbf{0}$. In this case (De Jong [66]) the initial conditions are:

$$\hat{x}_{1|N}^d = \left[(\mathbf{P}_1^d)^{-1} + (\mathbf{X}^d)^T \bar{\mathbf{B}}^{-1} \mathbf{X}^d \right]^{-1} \left[(\mathbf{P}_1^d)^{-1} \bar{x}_1^d + (\mathbf{X}^d)^T \bar{\mathbf{B}}^{-1} \tilde{Z} \right], \quad (8.29)$$

where $\bar{\mathbf{B}} = \mathbf{X}^s \mathbf{P}_1^s (\mathbf{X}^s)^T + \mathbf{B}$, and $\mathbf{B} = \text{cov}(\tilde{\mathbf{Z}}^*) = \text{diag}(\mathbf{B}_1, \mathbf{B}_2, \dots, \mathbf{B}_m)$, with $\{\mathbf{B}_t\}_{t=1,2,\dots,n}$ the sequence of covariances resulting from a KF(0,0). Taking into account that $(\mathbf{P}_1^d)^{-1} = \mathbf{0}$, expression (8.29) simplifies to:

$$\hat{\mathbf{x}}_{1|N}^d = \left[(\mathbf{X}^d)^T \bar{\mathbf{B}}^{-1} \mathbf{X}^d \right]^{-1} \left[(\mathbf{X}^d)^T \bar{\mathbf{B}}^{-1} \tilde{\mathbf{Z}} \right], \quad (8.30)$$

which is the GLS estimate of the non-null initial state, \mathbf{x}_1^d ; see De Jong [66] and Casals, Jerez, and Sotoca [40].

Note that (8.30) depends on $\bar{\mathbf{B}}^{-1} = [\mathbf{X}^s \mathbf{P}_1^s (\mathbf{X}^s)^T + \mathbf{B}]^{-1}$, and thus on the covariance \mathbf{P}_1^s . If the stochastic subsystem has unit roots, this matrix will have non-finite elements. Following De Jong and Chu-Chun-Lin [69], this inconvenience can be overcome by applying the matrix inversion lemma to $\bar{\mathbf{B}}^{-1}$, which yields:

$$\bar{\mathbf{B}}^{-1} = \mathbf{B}^{-1} - \mathbf{B}^{-1} \mathbf{X}^s \mathbf{R} (\mathbf{X}^s)^T \mathbf{B}^{-1}, \quad (8.31)$$

where $\mathbf{R} = \left[(\mathbf{P}_1^s)^{-1} + (\mathbf{X}^s)^T \mathbf{B}^{-1} \mathbf{X}^s \right]^{-1}$ and, substituting (8.31) in (8.30) yields:

$$\begin{aligned} \hat{\mathbf{x}}_{1|N}^d &= \left\{ (\mathbf{X}^d)^T \mathbf{B}^{-1} \mathbf{X}^d - (\mathbf{X}^d)^T \mathbf{B}^{-1} \mathbf{X}^s \mathbf{R} (\mathbf{X}^s)^T \mathbf{B}^{-1} \mathbf{X}^d \right\}^{-1} \\ &\quad \cdot \left\{ (\mathbf{X}^d)^T \mathbf{B}^{-1} \tilde{\mathbf{Z}} - (\mathbf{X}^d)^T \mathbf{B}^{-1} \mathbf{X}^s \mathbf{R} (\mathbf{X}^s)^T \mathbf{B}^{-1} \tilde{\mathbf{Z}} \right\}. \end{aligned} \quad (8.32)$$

The initial condition (8.32) allows for both stationary and unit roots because it depends, not on \mathbf{P}_1^s , but on its inverse $(\mathbf{P}_1^s)^{-1}$. When the system has both types of roots, the matrix $(\mathbf{P}_1^s)^{-1}$ converges to a finite value with as many null eigenvalues as nonstationary states; see Chapter 5.

Depending on the system dynamics, expression (8.32) can adopt different forms.

Case 1. The stochastic subsystem is static, so that $\mathbf{X}^s = \mathbf{0}$, and (8.32) therefore simplifies to:

$$\hat{\mathbf{x}}_{1|N}^d = \left[(\mathbf{X}^d)^T \mathbf{B}^{-1} \mathbf{X}^d \right]^{-1} (\mathbf{X}^d)^T \mathbf{B}^{-1} \tilde{\mathbf{Z}}. \quad (8.33)$$

Case 2. The deterministic and stochastic subsystems share the same dynamics, so that $\mathbf{X}^d = \mathbf{X}^s$, and (8.32) thus collapses to (8.33).

Case 3. The deterministic and stochastic subsystems share at least one nonstationary mode. This situation creates a severe identifiability problem because the matrix

$$(\mathbf{X}^d)^T \mathbf{B}^{-1} \mathbf{X}^d - (\mathbf{X}^d)^T \mathbf{B}^{-1} \mathbf{X}^s \mathbf{R} (\mathbf{X}^s)^T \mathbf{B}^{-1} \mathbf{X}^d, \quad (8.34)$$

whose inverse is required to compute (8.32), is singular. The only option in this case consists of using the pseudoinverse of (8.34) in order to compute the initial conditions.

8.2.2 Estimating the individual effect of each input

The procedure previously described is useful when one is interested in the total effect of all the exogenous variables. On the other hand, in models with several inputs one may want to estimate the individual effect of each one. This separation is easier if we write the deterministic subsystem (8.4)–(8.5) in its equivalent Luenberger Canonical Form (hereafter LCF; see Subsection 9.2.2), denoted by:

$$\bar{x}_{t+1}^d = \bar{\Phi} \bar{x}_t^d + \bar{\Gamma} u_t \quad (8.35)$$

$$z_t^d = \bar{H} \bar{x}_t^d + D u_t. \quad (8.36)$$

Model (8.35)–(8.36) can be obtained by a similar transformation of (8.4)–(8.5), given by $\bar{x}_t^d = T^{-1} x_t^d$, see Subsection 2.1.2. Accordingly, the initial state of the LCF system corresponding to the vector $\hat{x}_{1|N}^d$ defined in (8.32) would be:

$$\hat{x}_{1|N}^d = T^{-1} \hat{x}_{1|N}^d. \quad (8.37)$$

The controllability matrix of the LCF separates the dynamics associated with the individual inputs, therefore permitting one to discern whether a given input affects (or not) each system state. To see this, note that propagating the LCF according to (8.23) yields:

$$\bar{x}_t^d = \bar{\Phi}^{t-1} \bar{x}_1^d + \begin{bmatrix} \bar{\Phi}^{t-2} \bar{\Gamma} & \bar{\Phi}^{t-3} \bar{\Gamma} & \dots & \bar{\Gamma} \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_{t-1} \end{bmatrix}, \quad (8.38)$$

where $C = [\bar{\Phi}^{t-2} \bar{\Gamma} \quad \bar{\Phi}^{t-3} \bar{\Gamma} \quad \dots \quad \bar{\Gamma}]$ is the controllability matrix of the LCF system whose zero/nonzero structure characterizes which inputs affect any given state. Specifically, if the i -th state is not excited by the j -th input, then the elements $c(i, j), c(i, j+r), c(i, j+2r), \dots, c(i, j+(n-1)r)$ should be null, where r is the number of inputs.

Under these conditions, estimating the individual effect of each input requires the following two-step procedure.

Step 1: Decompose the initial condition (8.37) into r vectors, such that each one is exclusively related to one of the r system inputs.

Step 2: Propagate the r systems given by $\bar{\Phi}$, \bar{H} , and the j -th columns of $\bar{\Gamma}$ and D , for $j = 1, \dots, r$, starting from the corresponding initial conditions.

Ideally, the initial vectors $\hat{x}_{1,j}^d$ will be orthogonal, and such that $\hat{x}_{1|N}^d = \sum_{j=1}^r \hat{x}_{1,j}^d$. This orthogonality can be achieved by distributing the initial state of the LCF system according to the structure of the controllability matrix C , i.e., such that for $j = 1, \dots, r$ and $i = 1, \dots, n$, we set:

$$\tilde{x}_{1,j}^d(i) = \begin{cases} 0, & \text{if } c(i, j+kr) = 0, \text{ for } k = 0, 1, \dots, n-1, \\ \hat{x}_{1|N}^d(i), & \text{otherwise,} \end{cases} \quad (8.39)$$

where $\hat{\tilde{x}}_{1|N}^d(i)$ is the i -th element of $\hat{\tilde{x}}_{1|N}^d$. Finally, the vectors $\tilde{x}_{1,j}^d$ resulting from (8.39), should be transformed according to:

$$\hat{\tilde{x}}_{1,j}^d = \Pi_j \tilde{x}_{1,j}^d, \quad j = 1, \dots, r, \quad (8.40)$$

where Π_j is the orthogonal projector onto the null space of the matrix $[\hat{\tilde{x}}_{1,1}^d \hat{\tilde{x}}_{1,2}^d \dots \hat{\tilde{x}}_{1,j-1}^d \hat{\tilde{x}}_{1,j+1}^d \dots \hat{\tilde{x}}_{1,r}^d]$. The initial conditions (8.40) can then be interpreted as, either the part of the LCF initial state that is orthogonal to the rest of the inputs, or as the marginal contribution of each input toward the initial condition of the system.

In many models, such as transfer functions, each mode is affected only by a single input, so that the orthogonal projector Π_j is the identity, and the condition $\hat{\tilde{x}}_{1|N}^d = \sum_{j=1}^r \hat{\tilde{x}}_{1,j}^d$ is thus assured. On the other hand, if a given input affects more than one mode, then the decomposition does not assure a complete split of (8.37), so that in general there will be a remainder, $\hat{\tilde{x}}_{1|N}^d - \sum_{j=1}^r \hat{\tilde{x}}_{1,j}^d \neq 0$. This remainder can be propagated with the pair $(\bar{\Phi}, \bar{H})$, and the resulting component interpreted as the common effect of the system inputs.

8.3 Decomposition of the stochastic component

8.3.1 Characterization of the structural components

After computing the deterministic-stochastic decomposition (8.8), one may want to estimate the generalized structural components in (8.9). To this end, we first need to characterize their properties.

The literature often describes the properties of the structural components in the frequency domain and, following Burman [34]: (a) the trend is characterized by a peak at the null frequency, (b) the irregular component displays no significant peak at any frequency, and (c) the seasonal component has peaks at the basic seasonal frequency and its multiples. To complete the decomposition, (d) any reminding component should be assigned to the cyclic term.

To implement these ideas, we build on the block-diagonal form for the stochastic components (8.6)–(8.7), where the eigenvalues of the transition matrix characterize the properties of the states in both time and frequency domains; see e.g., West [223].

In particular, assume that $\lambda_{j,k} = a_j \pm b_j i$ is a pair of complex conjugate eigenvalues of the transition matrix Φ in state equation (8.6), associated with the j -th and k -th states. In this case, both states generate a peak in the pseudo-spectrum in the frequency $f_{j,k} = (2\pi)^{-1} \arctan(b_j/a_j)$, where $f_{j,k}$ is in cycles per unit time.

This general result has some particular cases that are worth discussing:

1. If the real part of $\lambda_{j,k}$ is zero, then $b_j/a_j = \infty$, implying that $f_{j,k} = 1/4$ if $b_j > 0$, or $f_{j,k} = 3/4$ if $b_j < 0$
2. If any eigenvalue is a real number, then $b_j/a_j = 0$, implying that $f_{j,k} = 0$ if $a_j > 0$, or $f_{j,k} = 1/2$ if $a_j < 0$
3. If $b_j = 0$ and a_j tends to one, then the spectral power tends to infinity at $f_{j,k} = 0$

If the output series are seasonal, it is easy to characterize the set of seasonal frequencies, defined as $\mathbf{F}_s = \{f_i = k_i/s; k_j = 1, 2, \dots, [s/2]\}$, where s denotes the seasonal period (number of observations per seasonal cycle), and $[s/2] = s/2$ if s is even, or $[s/2] = (s-1)/2$ if s is odd.

Building on these ideas, the states in the block-diagonal stochastic subsystem (8.6)–(8.7) can be naturally assigned to the structural components with the rules summarized in Table 8.1.

Table 8.1 Association between state variables and structural components according to the eigenvalues of the transition matrix and their spectral properties.

Eigenvalues of Φ		Spectral peaks	Component
Real	$\lambda_j = 1$	$f_j = 0$	Trend
	$\lambda_{j,k} = a_j, -1 < a_j < 1$	$f_j = 0, \text{if } 0 < a_j < 1$ If $a_j = 0$, then x_j is either a redundant state, which has no effect on the structural components, or a lagged uncorrelated error, which does not create a spectral peak in any frequency. In this case, it should be included in the cycle	Cycle
		$f_j = 1/2, \text{if } -1 < a_j < 0$	Seasonal if $f_{j,k} \in \mathbf{F}_s$, cycle otherwise
Complex	$\lambda_{j,k} = a_j \pm b_j i$	$f_{j,k} = (2\pi)^{-1} \arctan(b_j/a_j)$	Seasonal if $f_{j,k} \in \mathbf{F}_s$, cycle otherwise

According to this classification, we can reorganize the block-diagonal model (8.6)–(8.7) as:

$$\begin{bmatrix} \bar{x}_{t+1}^t \\ \bar{x}_{t+1}^c \\ \bar{x}_{t+1}^s \end{bmatrix} = \begin{bmatrix} \bar{\Phi}^t & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \bar{\Phi}^c & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \bar{\Phi}^s \end{bmatrix} \begin{bmatrix} \bar{x}_t^t \\ \bar{x}_t^c \\ \bar{x}_t^s \end{bmatrix} + \begin{bmatrix} \bar{K}^t \\ \bar{K}^c \\ \bar{K}^s \end{bmatrix} a_t \quad (8.41)$$

$$\mathbf{z}_t = [\bar{H}^t \quad \bar{H}^c \quad \bar{H}^s] \begin{bmatrix} \bar{x}_t^t \\ \bar{x}_t^c \\ \bar{x}_t^s \end{bmatrix} + a_t, \quad (8.42)$$

where \bar{x}_t^t is the vector of nonstationary states, \bar{x}_t^c is the vector of stationary (nonseasonal) states, and \bar{x}_t^s is the vector of seasonal states. As such, the structural components are given by the different addends in the observation equation (8.42):

$$t_t = \bar{H}^t \bar{x}_t^t \quad (8.43)$$

$$c_t = \bar{H}^c \bar{x}_t^c \quad (8.44)$$

$$s_t = \bar{H}^s \bar{x}_t^s \quad (8.45)$$

$$\varepsilon_t = a_t. \quad (8.46)$$

8.3.2 Estimation of the structural components

Given the characterization of the structural components detailed in Section 8.3.1, the problem now reduces to obtaining estimates of the state variables and combining them according to (8.43)–(8.46). The method of choice to achieve is the fixed-interval smoother; see Chapter 4.

One must also bear in mind that the stochastic subsystem (8.6)–(8.7) is a SEM, and thus the smoothed estimates of its states converge to exact values with null variances and covariances; see Section 4.6. Accordingly, the estimates for the components will likewise be exact, as they are linear combinations of these smoothed states.

Null variances and covariances are very important for signal extraction due to two reasons. First, null variances ensure that the components do not change when the sample increases². Second, zero covariances ensure that the resulting components are conditionally independent, so that any component can be analyzed and interpreted independently of any other component(s).

8.4 Structure of the method

Building on the ideas discussed in previous sections, the structural decomposition of a vector of time series, z_t , can be organized according to the following steps:

Step 1 Obtain an adequate SS model for z_t and, optionally, the equivalent SEM; see Section 2.2. Obtain the equivalent block-diagonal form; see Subsection 3.1.3.

Step 2 Obtain the implied deterministic and stochastic subsystems as shown in Section 8.1, and reduce them to their minimal dimensions by means of the Staircase algorithm; see Rosenbrock [190].

Step 3 Estimate the deterministic component, z_t^d , following the procedure described in Section 8.2.

Step 4 Determine the stochastic component as $z_t^s = z_t - z_t^d$, and compute smoothed estimates of the states in the stochastic subsystem.

Step 5 Classify the different states following the indications in Table 8.1. Compute estimates of the trend, cycle, and seasonal components by combining the estimates of the states with the corresponding coefficients in the observation equation; see (8.43)–(8.45).

Step 6 Estimate the irregular component by subtracting from z_t the estimates for the trend, cycle, and seasonal components; see (8.42).

8.5 Examples

8.5.1 Comparing different methods with simulated data

This example compares the components extracted from a SEM, according to the method described in previous sections, with those resulting from a MEM, which is

²Intuitively, if a magnitude is exactly known, then its estimate will not change, regardless of whether the sample is augmented.

the representation employed by many alternative approaches. To this end, we simulate a stochastic process, defined in Table 8.2, which can be interpreted both, as a Structural Time Series Model (STSM), see Harvey [118], as well as an ARIMA-based decomposition. The components are then extracted using the direct MEM representation and the equivalent SEM³.

Table 8.2 *Definition of the data generating process in STSM and ARIMA forms.*

Component	STSM	ARIMA representation
Trend	$\mu_{t+1} = \mu_t + \beta_t$ $\beta_{t+1} = \beta_t + \zeta_t$	$(1 - B)^2 \mu_{t+1} = \zeta_t$
Seasonal	$\gamma_{t+1} = \gamma_t + \gamma_{t-1} + \gamma_{t-2} + \omega_t$	$(1 + B + B^2 + B^3) \gamma_{t+1} = \omega_t$
Time series	$z_t = \mu_t + \gamma_t + e_t$	$(1 - B)(1 - B^4) z_t = (1 - 0.93B - 0.09B^2 - 0.04B^3 - 0.59B^4 + 0.55B^5) a_t$ $\simeq (1 - 0.93B)(1 - 0.54B^4) a_t$

Therefore the trend is an integrated random walk process and the seasonal component follows what Harvey [118] calls a “dummy variable seasonality” model, where the sum of the seasonal components over a year is a random disturbance. The terms ζ_t , ω_t , and e_t are Gaussian white noise errors with the following covariances:

$$\text{cov} \begin{pmatrix} \zeta_t \\ \omega_t \\ e_t \end{pmatrix} = \begin{bmatrix} 1/1600 & 0 & 0 \\ 0 & 0.1 & 0 \\ 0 & 0 & 1 \end{bmatrix}. \quad (8.47)$$

Note that the variance ratio, $\sigma_\zeta^2/\sigma_e^2 = 1/1600$, was chosen according to the value assumed in the Hodrick–Prescott [126] filter for quarterly data.

The second column in Table 8.2 reveals how the model can be interpreted as an ARIMA-model-based decomposition, which realizes an ARIMA process for the observable series z_t . Using the results in Chapters 2 and 3, this model can be written in the equivalent block-diagonal SEM form displayed in Table 8.3.

Note that: (a) the states in this representation are clearly separated in (1×1) and (2×2) blocks, and (b) the eigenvalues of the transition matrix are identical to the inverse AR roots of the stochastic processes in Table 8.2. According to their frequencies, the first and second states correspond to the trend, while the third, fourth, and fifth states correspond to the seasonal component.

Taking into account this classification and the coefficients in the last row of Table 8.3, the trend and seasonal components would be given by:

$$t_t = \bar{x}_t^1 \quad (8.48)$$

$$s_t = -0.707\bar{x}_t^4 + 0.577\bar{x}_t^5 \quad (8.49)$$

$$\varepsilon_t = z_t - t_t - s_t. \quad (8.50)$$

³Note that this design puts the method described here at a deliberate disadvantage, as the MEM takes into account certain variance proportionality constraints which are then lost in the SEM representation.

Table 8.3 Structure of the block-diagonal SEM representation of the data generating process and association of the state variables to structural components.

Outputs	Inputs						Eig($\bar{\Phi}$)	Freq. f_i	Compo- nent
	\bar{x}_t^1	\bar{x}_t^2	\bar{x}_t^3	\bar{x}_t^4	\bar{x}_t^5	a_t			
	$\bar{\Phi}$					E			
\bar{x}_{t+1}^1	1	1	0	0	0	0.188	1	0	Trend
\bar{x}_{t+1}^2	0	1	0	0	0	0.019	1	0	
\bar{x}_{t+1}^3	0	0	0	-0.577	0	0.135	$\pm i$	0.25	Seasonal
\bar{x}_{t+1}^4	0	0	1.732	0	0	0.005	-1	0.5	
\bar{x}_{t+1}^5	0	0	0	0	-1	-0.203			
	H					-			
z_t	1	0	0	-0.707	0.577	1			

Combining the SS model in Table 8.3 with (8.48)–(8.50), we can derive ARIMA representations for these components; see Chapter 9. Table 8.4 compares them with the corresponding models for the data generating process.

Table 8.4 Comparison between the ARIMA representations of the STSM and the block-diagonal SEM form.

Component	STSM	Block-diagonal SEM
Trend	$(1 - B)^2 \mu_{t+1} = \zeta_{t-1}$ $\zeta_{t-1} \sim NID(0, 1/1600)$	$(1 - B)^2 t_t = (1 - 0.899B) \varepsilon_{t-1}$
Seasonal	$(1 + B + B^2 + B^3) \gamma_t = \omega_{t-1}$ $\omega_{t-1} \sim NID(0, 0.1)$	$(1 + B + B^2 + B^3) s_t =$ $(1 + 1.402B + 2.347B^2) (-0.120) \varepsilon_{t-1}$
Irregular	$e_t \sim NID(0, 1)$	$\varepsilon_t \sim NID(0, 1.824)$

Note that the ARIMA representations for the STSM and the equivalent SEM in Table 8.4 have the same autoregressive structures, but different error terms. While the components of the STSM follow pure AR processes, the components implied by the SEM receive shocks from the delayed irregular component, affected by different moving average polynomials and scale parameters.

We now generate 200 random samples for the data generating process and compute the smoothed estimates of the components using both the STSM and the equivalent SEM. Figure 8.1 compares both sets of components.

Note that: (a) the trend and seasonal components obtained in both cases are very similar; (b) the trend estimate resulting from the SEM is more volatile because the smoothness of the STSM trend is a byproduct of the variance ratios in (8.47), which are explicit in the STSM but not in the SEM; and (c) the irregular components have different volatilities, as expected; see Table 8.4.

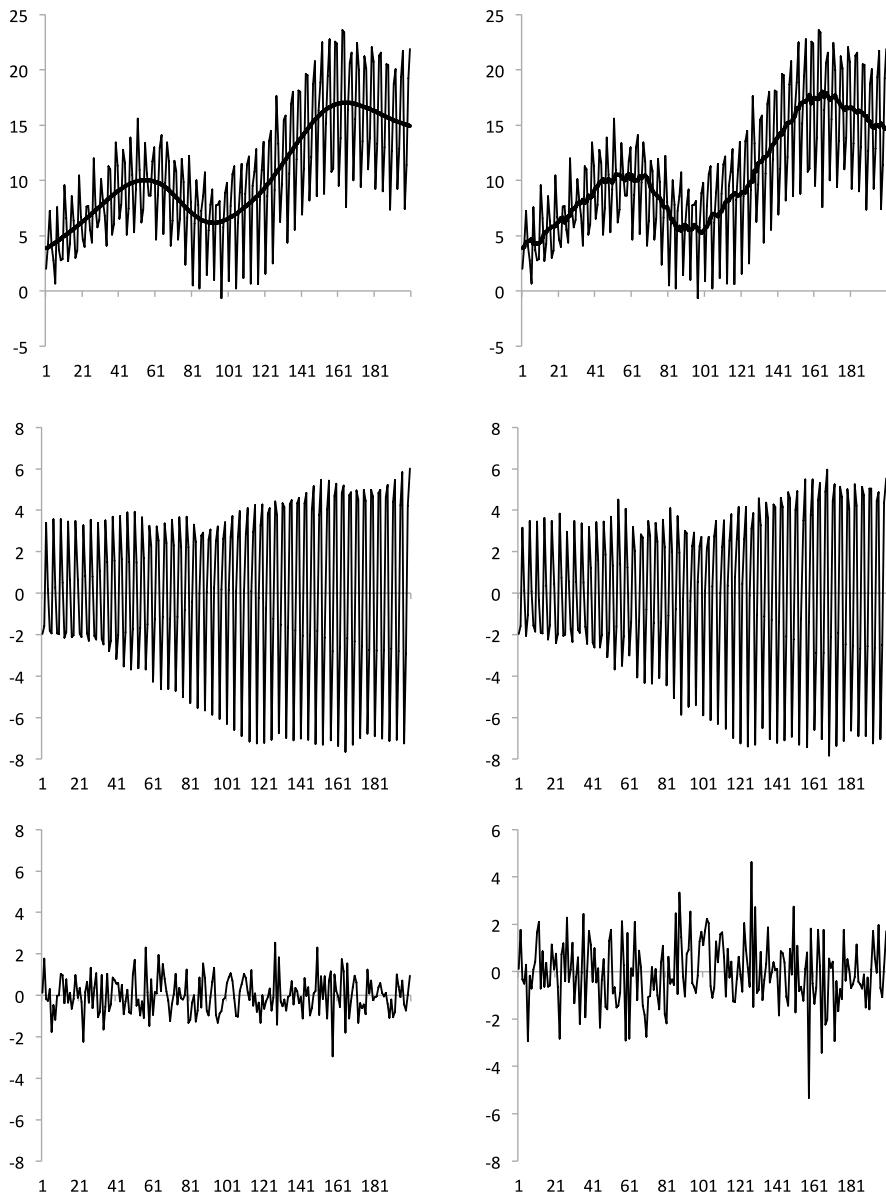


Figure 8.1 Comparison between the model components obtained from the STSM (left) and the equivalent SEM (right).

Despite their overall similarity, the uncertainty affecting both sets of components is very different. Figure 8.2 compares the trace of the covariance matrices corresponding to the STSM and SEM components. Note that the uncertainty of the components derived from the MEM display, the U-shape characteristic of symmetric filters, meaning that the estimates at the center of the sample are more precise than those at the extremes. On the other hand, the components implied by the SEM converge to exact values, so they will not be revised if the sample increases with new data.

The code required to replicate this example can be found in the file `Example0851.m`; see Appendix D.

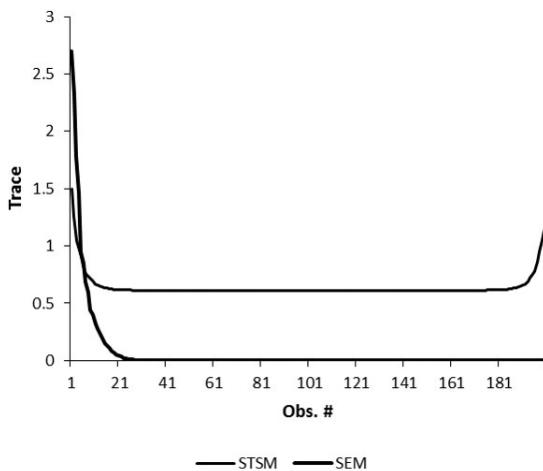


Figure 8.2 *Trace of the smoothed covariance matrices corresponding to the STSM and the equivalent SEM.*

8.5.2 Common features in wheat prices

In this example we analyze the annual average prices of the wheat sold at *Monasterio de Sandoval* (S_t) and at *Fábrica de la iglesia de Alaraz* (A_t), between 1691 and 1788 ($N = 98$), in reales per fanega (source: own compilation). The second series has six missing values due to discontinuities in the source. The goals of the analysis are testing if these series show common statistical features, see Engle and Kozicki [87], that could be interpreted as symptoms of market integration, and, if found, apply our method to estimate them.

After a standard univariate analysis, the following models were fitted to the log-transformed series and the differences in log prices:

$$(1 - 0.042B + 0.296B^2)(1 - B)\log S_t = \hat{a}_t^S, \quad \hat{\sigma}_S^2 = 0.077, \\ Q(10) = 5.29, \quad p = 4.10, \quad df = 0.54, \quad (8.51)$$

$$(1 - 0.004B + 0.342B^2)(1 - B) \log A_t = \hat{a}_t^A, \quad \hat{\sigma}_A^2 = 0.113, \\ Q(10) = 10.67, \quad p = 3.99, \quad df = 0.59, \quad (8.52)$$

$$(1 - 0.280B)(\log A_t - \log S_t) = 0.107 + \hat{a}_t^S, \quad \hat{\sigma}^2 = 0.067, \quad Q(10) = 6.02, \\ (0.102) \quad (0.031) \quad (8.53)$$

where the figures in parentheses are standard errors for the corresponding estimates, $Q(10)$ is the Ljung–Box [151] statistic computed with the first 10 lags of the residual sample autocorrelation function, and p and df are, respectively, the period and damping factor of the pseudo-cycle implied by the corresponding AR(2) factors. Because of the missing values in series A_t , the estimates in (8.52) and (8.53) were obtained using the technique described in Subsection 3.3.2.

It therefore appears that individual log prices are adequately represented by ARIMA(2,1,0) processes, implying that both series are nonstationary and have harmonic pseudo-cycles with a period of roughly four years. Also, the model for the difference of log prices is a stationary AR(1) with a constant term, meaning that both series share a common trend and a common harmonic cycle, see Engle and Kozicki [87]. The presence of these cofeatures supports the idea that both markets were substantially integrated in this historical period.

Further evidence can be found by building a multivariate model, constrained to represent these cofeatures. After a standard specification analysis (Jenkins and Alavi [131]) we obtained the model:

$$\begin{bmatrix} 1 - 0.317B & 0 \\ 0 & (1 - 0.048B + 0.260B^2)(1 - B) \end{bmatrix} \begin{bmatrix} 1 & -1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \log A_t \\ \log S_t \end{bmatrix} \\ = \begin{bmatrix} 0.098 \\ (0.026) \end{bmatrix} + \begin{bmatrix} 1 & 0 \\ 0.602B & 0 \end{bmatrix} \begin{bmatrix} \hat{a}_t^1 \\ \hat{a}_t^2 \end{bmatrix}, \quad (8.54)$$

with a noise covariance matrix of

$$\hat{\Sigma}_a = \begin{bmatrix} 0.068 & -0.003 \\ -0.003 & 0.053 \end{bmatrix}, \quad Q(10) = \begin{bmatrix} 7.54 & 11.65 \\ 11.11 & 9.76 \end{bmatrix},$$

where $Q(10)$ is a matrix of portmanteau statistics for the residual sample autocorrelations and cross-correlations.

Table 8.5 presents the block-diagonal SEM equivalent to (8.54). According to the eigenvalues of the transition matrix, the first state corresponds to the trend, the second state to a cyclical movement, and the third and fourth states to a harmonic pseudo-cycle. The observation equation in Table 8.5 implies the following decomposition of

Table 8.5 Structure of the block-diagonal SEM equivalent to model (8.54) and association of the state variables to structural components.

Outputs	Inputs				$\text{eig}(\bar{\Phi})$	Component
	\bar{x}_t^1	\bar{x}_t^2	\bar{x}_t^3	\bar{x}_t^4		
	Φ				$\bar{\Gamma}$	E
\bar{x}_{t+1}^1	1	0	0	0	0.497	0.825
\bar{x}_{t+1}^2	0	0.317	0	0	0.317	0
\bar{x}_{t+1}^3	0	0	0.024	1.001	0	0.602
\bar{x}_{t+1}^4	0	0	-0.259	0.024	0	0.056
						Trend
						Cycle
						Cycle
						(harmonic)
$\log A_t$	1	1	0.202	-0.819	\bar{H}	I
$\log S_t$	1	0	0.202	-0.819	0	0
						1

the output:

$$t_t = \bar{x}_t^1 \quad (8.55)$$

$$c_t = 0.202\bar{x}_t^3 - 0.819\bar{x}_t^4 \quad (8.56)$$

$$c_t^A = \bar{x}_t^2 + 0.098 \quad (8.57)$$

$$\varepsilon_t^A = \hat{a}_t^1 \quad (8.58)$$

$$\varepsilon_t^S = \hat{a}_t^2, \quad (8.59)$$

where t_t denotes the common trend, c_t is the common cycle, and c_t^A is a stationary component capturing all the individual cyclical features of the log price in Alaraz. Note that this component includes a constant term. Finally, the white noise terms $\varepsilon_t^A = a_t^1$ and $\varepsilon_t^S = a_t^2$, are the irregular components affecting the series of Alaraz and Sandoval, respectively.

As the sample has some missing values, the trace of the smoothed covariance matrices drops quickly to zero, but shows transitory peaks of uncertainty in the years following a missing value (Figure 8.3). The estimated cofeatures are shown in Figure 8.4.

The code required to replicate this example can be found in the file `Example0852.m`; see Appendix D.

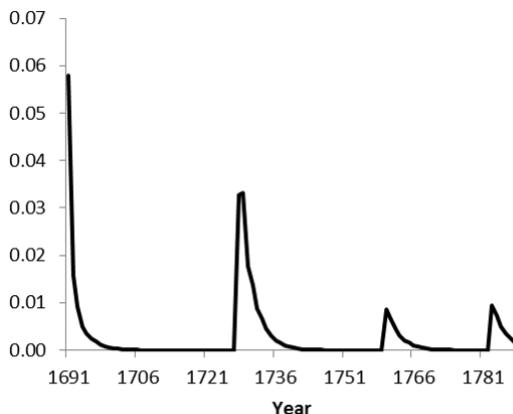


Figure 8.3 *Trace of the smoother covariance of Model (8.54).*

8.5.3 The effect of advertising on sales

This example will highlight the manner in which the decompositions proposed in this chapter can be used to gain insight into the relationship between different time series. To this end, we will use the famous monthly series of sales and advertising of the Lydia Pinkham vegetable compound, from January 1954 to June 1960; see Palda

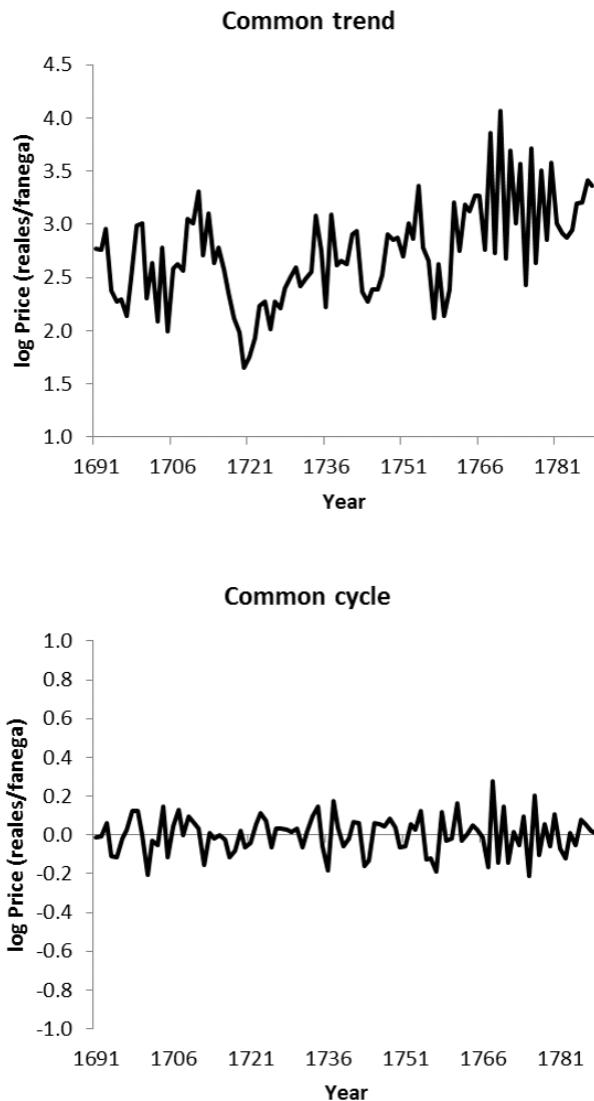


Figure 8.4 *Common features of (log) wheat prices.*

[175]. Simple inspection of the series, displayed in Figure 8.5, shows that both show seasonal fluctuations and a downward drift.

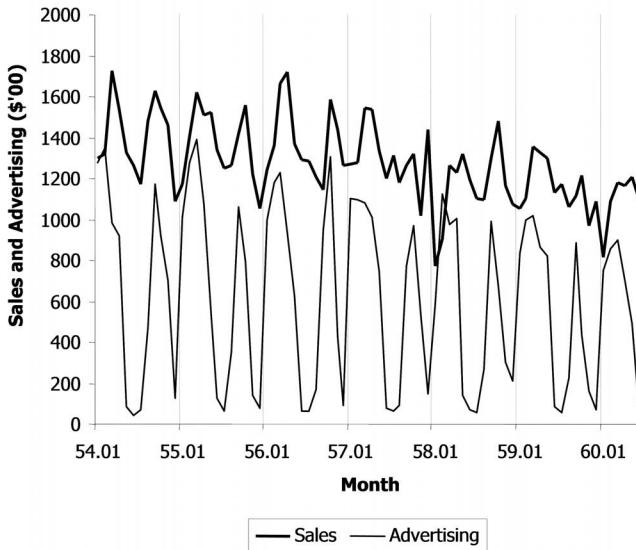


Figure 8.5 *Monthly series of sales and advertising of the Lydia Pinkham vegetable compound from January 1954 to June 1960. Source: Palda [175].*

Building on the methods described by Box, Jenkins, and Reinsel [30], we obtained the following model:

$$\log S_t \times 100 = \frac{(.022) (.022) (.026)}{.048 + .016B + .043B^2} \log A_t \times 100 \\ \frac{(1 - .713B + .751B^2)}{(1 - .127)(.094)} + \frac{(.052) (.094)}{(1 - .899B)(1 - .628B^{12})} \hat{a}_t, \quad (8.60)$$

where S_t and A_t denote, respectively, monthly sales and advertising⁴, while the values in parentheses are standard errors for the corresponding estimates. The estimated residual standard deviation is $\hat{\sigma}_a = 7.536$, and the residuals do not show evidence of misspecification, as the portmanteau Q statistics for the residual auto and cross-correlations did not reject the null of no autocorrelation.

Note that:

1. The roots of the polynomials in the numerator and denominator of the transfer function (8.60) are, respectively, $-0.186 \pm 1.040i$ and $0.475 \pm 1.052i$. Thus, even though this model may be somewhat overparametrized, there are no redundant dynamic factors in this term.

⁴Observations between November 1957 and February 1958 have been treated as missing values because they were outliers.

2. The impulse-response function implied by this specification has both positive and negative values. In a sales-advertising system this can happen when the product has a loyal customer base and the advertising accelerates the consumption, thereby changing the distribution of restocking purchases over time.
3. The value of the MA coefficient is so close to one that standard hypothesis testing would not reject the null of noninvertibility. However, one must take into account that under this null, Gaussian maximum-likelihood estimates display a so-called “pile-up” effect, which makes standard tests biased toward non-rejection. Taking into account these distortions, the testing procedure proposed in Section 3 of Davis, Chen, and Dunsmuir [63] safely rejects noninvertibility at the 5% significance level.

Using the decomposition described in Section 8.1, we: (a) estimated the additive contribution of log advertising to log sales, (b) obtained the corresponding multiplicative estimate of this effect, and (c) discounted it from the original sales series. The results are shown in Figure 8.6. Figure 8.7 plots the resulting estimate of the value added by advertising, computed as the difference between the contribution of advertising to sales in each month and the corresponding investment. Note that the return over several months is negative. In fact, the sum of the estimates in Figure 8.7 is -0.3 million dollars, and thus advertising in this case did not create value for the firm.

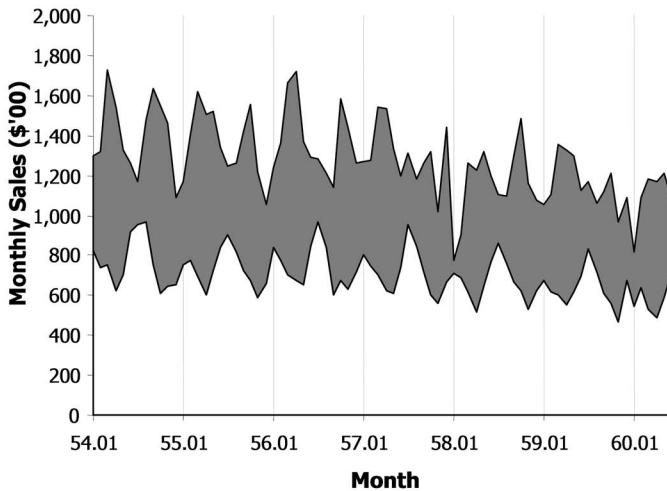


Figure 8.6 *Sales versus the exponential of the stochastic component for the Lydia Pinkham data. The distance between both series (gray area) is an estimate of the effect of advertising over sales.*

Further decomposition of the stochastic components reveals a possible cause of this lack of performance. Figure 8.8 plots the multiplicative components of sales corresponding to seasonality and the estimated effect of advertising. Note that both series have a clear negative correlation, so advertising was systematically increased

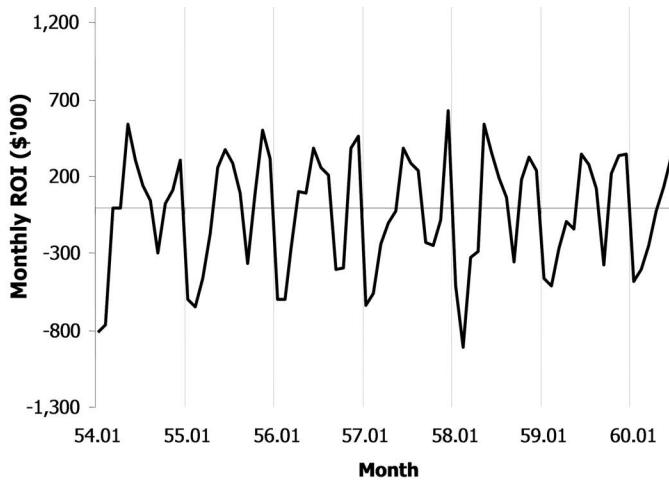


Figure 8.7 Estimated value added by advertising, computed in each period as the estimate of sales generated by advertising minus the advertising investment (Lydia Pinkham data).

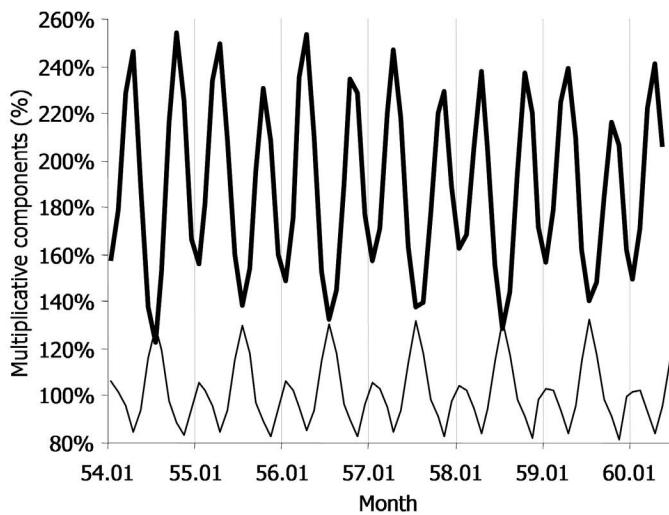


Figure 8.8 Multiplicative effect of advertising (thick line) versus multiplicative seasonality of sales (thin line) for the Lydia Pinkham data. The values between November 1957 and February 1958 are smoothed interpolations.

in the seasonal troughs, and decreased in the peaks. Taking into account that both effects are multiplicative, it is immediately apparent that this anti-cyclical management of advertising did not optimize sales.

The code required to replicate this example can be found in the file Example0853.m; see Appendix D.

This page intentionally left blank

Chapter 9

The VARMAX representation of a State-Space model

9.1	Notation and previous results	130
9.2	Obtaining the VARMAX form of a State-Space model	131
9.2.1	From State-Space to standard VARMAX	132
9.2.2	From State-Space to canonical VARMAX	133
9.3	Practical applications and examples	135
9.3.1	The VARMAX form of some common State-Space models	135
9.3.2	Identifiability and conditioning of the estimates	135
9.3.3	Fitting an errors-in-variables model to Wolf's sunspot series	139
9.3.4	"Bottom-up" modeling of quarterly US GDP trend	140

In Subsection 2.2.4 we presented an analytical procedure to write a generalized VARMAX model in SS form. In this context, "analytical" means that it can be applied to formulations where the model parameters may be generically represented by letters. In this chapter we will discuss a numerical procedure to do the inverse transformation, that is, to obtain the coefficients of the VARMAX representation for any linear SS model. By "numerical" we now mean that this is a computational procedure, which requires numerical values for the parameters.

The fact that there are procedures to go from VARMAX to SS form and vice versa has a clear and interesting consequence: both representations are equally general and choosing between them is just a matter of convenience. Beyond this theoretical implication, these procedures are also useful to solve three practical problems: discussing the identifiability of an SS model, performing diagnostic checks, and calibrating the SS model parameters to ensure that it realizes any given reduced-form VARMAX. Section 9.3 discusses these applications and provides practical examples.

9.1 Notation and previous results

Much applied work in time series analysis is based on variations of the model:

$$\bar{\mathbf{F}}(\mathbf{B})\mathbf{z}_t = \bar{\mathbf{G}}(\mathbf{B})\mathbf{u}_t + \bar{\mathbf{L}}(\mathbf{B})\mathbf{a}_t \quad (9.1)$$

which was already defined in equation (2.33). We will assume that this model is left coprime¹, but the roots of $\bar{\mathbf{F}}(\mathbf{B})$ and $\bar{\mathbf{L}}(\mathbf{B})$ are allowed to be either inside or outside the unit circle.

Left coprimeness is important, but it is not enough to ensure identifiability of (9.1), as there are still infinitely many parameter sets that realize \mathbf{z}_t . To see this, note that one can pre-multiply expression (9.1) by any scalar or conformable matrix and obtain an output-equivalent model. Accordingly, identifiability requires additional constraints over $\bar{\mathbf{F}}_0$ and $\bar{\mathbf{L}}_0$. For instance, $\bar{\mathbf{F}}_0 = \bar{\mathbf{L}}_0 = \mathbf{I}$ yields the standard VARMAX process.

An important feature of model (9.1) is its maximum dynamic order, defined as $p_{\max} = \max\{p, s, q\}$, where p , s , and q are the orders of the polynomials $\bar{\mathbf{F}}(\mathbf{B})$, $\bar{\mathbf{G}}(\mathbf{B})$, and $\bar{\mathbf{L}}(\mathbf{B})$, respectively. We will use this concept often throughout this chapter.

An interesting alternative to standard VARMAX is the widely used VARMAX *echelon* form. The system (9.1) is in echelon form if the triple $[\bar{\mathbf{F}}(\mathbf{B}) : \bar{\mathbf{G}}(\mathbf{B}) : \bar{\mathbf{L}}(\mathbf{B})]$ is in echelon canonical form, *i.e.*, denoting by $\bar{F}_{kl}(\mathbf{B})$ the kl -th element of $\bar{\mathbf{F}}(\mathbf{B})$, and similarly $\bar{G}_{kl}(\mathbf{B})$ for $\bar{\mathbf{G}}(\mathbf{B})$ and $\bar{L}_{kl}(\mathbf{B})$ for $\bar{\mathbf{L}}(\mathbf{B})$, the polynomial operators are uniquely defined by:

$$\bar{F}_{kk}(\mathbf{B}) = 1 + \sum_{i=1}^{p_k} \bar{F}_{kk}(i)\mathbf{B}^i, \quad \text{for } k = 1, \dots, m \quad (9.2)$$

$$\bar{F}_{kl}(\mathbf{B}) = \sum_{i=p_k-p_{kl}+1}^{p_k} \bar{F}_{kl}(i)\mathbf{B}^i, \quad \text{for } k \neq l \quad (9.3)$$

$$\bar{G}_{kl}(\mathbf{B}) = \sum_{i=0}^{p_k} \bar{G}_{kl}(i)\mathbf{B}^i, \quad \text{for } k = 1, \dots, m \quad (9.4)$$

$$\bar{L}_{kl}(\mathbf{B}) = \sum_{i=0}^{p_k} \bar{L}_{kl}(i)\mathbf{B}^i, \quad \text{with } \bar{L}_{kl}(0) = \bar{F}_{kl}(0) \text{ for } k, l = 1, \dots, m, \quad (9.5)$$

where the integers p_k , $k = 1, \dots, m$, are known as “Kronecker indices” or “observability indices.” They determine the structure of ones and zeros in the echelon form. Equation (9.3) uses the index p_{kl} defined as:

$$p_{kl} = \begin{cases} \min(p_k + 1, p_l), & \text{for } k \geq l, \\ \min(p_k, p_l), & \text{for } k < l, \end{cases} \quad k, l = 1, \dots, m. \quad (9.6)$$

where m is the dimension of \mathbf{z}_t .

¹For a formal definition of left coprimeness see, e.g. Reinsel [187]. Informally speaking, a process is left coprime if it does not have identical dynamic factors in its autoregressive and moving average polynomials.

Example 9.1.1 Standard VARMAX versus VARMAX echelon

Consider the standard VARMA(2,2) model:

$$(\mathbf{I} + \bar{\mathbf{F}}_1 \mathbf{B} + \bar{\mathbf{F}}_2 \mathbf{B}^2) \mathbf{z}_t = (\mathbf{I} + \bar{\mathbf{L}}_1 \mathbf{B} + \bar{\mathbf{L}}_2 \mathbf{B}^2) \mathbf{a}_t, \quad (9.7)$$

where,

$$\begin{aligned} \bar{\mathbf{F}}_1 &= \begin{bmatrix} -0.70 & 0 & 0 \\ 0.48 & -0.50 & -0.90 \\ -0.02 & 0.30 & -0.20 \end{bmatrix}, & \bar{\mathbf{F}}_2 &= \begin{bmatrix} 0.30 & -0.20 & 0.50 \\ -0.12 & 0.08 & -0.20 \\ 0.18 & -0.12 & 0.30 \end{bmatrix}, \\ \bar{\mathbf{L}}_1 &= \begin{bmatrix} -0.20 & 0.40 & 0.70 \\ 0.68 & -0.46 & -0.68 \\ 0.18 & 1.24 & -0.38 \end{bmatrix}, & \bar{\mathbf{L}}_2 &= \begin{bmatrix} 0.30 & 0.50 & -0.80 \\ -0.12 & -0.20 & 0.32 \\ 0.18 & 0.30 & -0.48 \end{bmatrix}. \end{aligned}$$

Note that this representation has 34 non-zero parameters, excluding those of the covariance matrix, and $p_{\max} = 2$.

Consider now the VARMA echelon form:

$$(\bar{\mathbf{F}}_0 + \bar{\mathbf{F}}_1 \mathbf{B} + \bar{\mathbf{F}}_2 \mathbf{B}^2) \mathbf{z}_t = (\bar{\mathbf{L}}_0 + \bar{\mathbf{L}}_1 \mathbf{B} + \bar{\mathbf{L}}_2 \mathbf{B}^2) \mathbf{a}_t, \quad (9.8)$$

with,

$$\begin{aligned} \bar{\mathbf{F}}_0 = \bar{\mathbf{L}}_0 &= \begin{bmatrix} 1.0 & 0 & 0 \\ 0.4 & 1.0 & 0 \\ -0.6 & 0 & 1.0 \end{bmatrix}, & \bar{\mathbf{F}}_1 &= \begin{bmatrix} -0.7 & 0 & 0 \\ 0.2 & -0.5 & -0.9 \\ 0.4 & 0.3 & -0.2 \end{bmatrix}, \\ \bar{\mathbf{F}}_2 &= \begin{bmatrix} 0.3 & -0.2 & 0.5 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, & \bar{\mathbf{L}}_1 &= \begin{bmatrix} -0.2 & 0.4 & 0.7 \\ 0.6 & -0.3 & -0.4 \\ 0.3 & 1.0 & -0.8 \end{bmatrix}, & \bar{\mathbf{L}}_2 &= \begin{bmatrix} 0.3 & 0.5 & -0.8 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}. \end{aligned}$$

It is easy to see that models (9.7) and (9.8) are equivalent by premultiplying (9.8) by $\bar{\mathbf{F}}_0^{-1}$. Kronecker indices in model (9.8) are $p_k = \{2, 1, 1\}$, corresponding to the maximum dynamic order of each component of \mathbf{z}_t . Obviously, p_{\max} must be the same in the standard (9.7) and echelon (9.8) representations. Finally, the VARMA echelon: (a) reduces the number of non-zero parameters from 34 to 24 and, (b) is a canonical form, meaning that there are no alternative representations with fewer parameters. These advantages have been pointed out by many authors, Hannan and Deistler [116] and Lutkepohl and Poskitt [159] among others, and more recently by Mélard, Roy, and Saidi [164].

9.2 Obtaining the VARMAX form of a State-Space model

We will now describe two separate procedures for computing the VARMAX coefficients corresponding to a given SS model.

9.2.1 From State-Space to standard VARMAX

The first algorithm requires that the dynamics of the SS model comply with two conditions:

Condition C.1: The system order, n , must be an integer multiple of the number of system outputs, m .

Condition C.2: The maximum system order, p_{\max} , must be equal to n/m , and the corresponding observability matrix,

$$\mathbf{O}_{p_{\max}} = [\mathbf{H}^T \quad (\mathbf{H}\Phi)^T \quad (\mathbf{H}\Phi^2)^T \quad \dots \quad (\mathbf{H}\Phi^{p_{\max}-1})^T]^T,$$

must be of full rank.

Two important special cases that satisfy these conditions are:

1. Every univariate minimal system, as it has $m = 1$ and $\mathbf{O}_{p_{\max}} = \mathbf{O}_n$, which is of full rank (for any minimal system), and
2. The multivariate minimal systems whose outputs ($z_{k,t}$) have identical Kronecker indices (p_k), since in this case, $n = \sum_{k=1}^m p_k$.

Under conditions C.1 and C.2, the algorithm proceeds as follows.

Step 1: If the SS model is not minimal (see Definition 2.4), reduce it to an equivalent minimal realization by the procedure discussed in Subsection 8.1.2.

Step 2: Transform the minimal SS model to the corresponding SEM form (2.19)–(2.20) by the procedure described in Subsection 2.2.4.

Step 3: Obtain the Observable Canonical Form (OCF). To this end, we must take into account that for any minimal SEM fulfilling conditions C.1 and C.2, there exists a nonsingular matrix \mathbf{T} such that the similar transformation given by $\Phi^* = \mathbf{T}^{-1}\Phi\mathbf{T}$ and $\mathbf{H}^* = \mathbf{HT}$ has the OCF structure, that is:

$$\Phi^* = \begin{bmatrix} -\bar{\mathbf{F}}_1 & \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ -\bar{\mathbf{F}}_2 & \mathbf{0} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ -\bar{\mathbf{F}}_{p_{\max}-1} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{I} \\ -\bar{\mathbf{F}}_{p_{\max}} & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \end{bmatrix}, \quad \mathbf{H}^* = [\mathbf{I} \quad \mathbf{0} \quad \dots \quad \mathbf{0}]. \quad (9.9)$$

where $\bar{\mathbf{F}}_i$, $i = 1, 2, \dots, p_{\max}$, denotes the autoregressive coefficient matrices in Model (9.1). Note that the first component in \mathbf{H}^* is the identity matrix because $\bar{\mathbf{F}}_0 = \mathbf{I}$ in standard VARMAX. A formal proof of this result and a procedure to compute the transformation matrix \mathbf{T} can be found in Casals, Garcia-Hiernaux, and Jerez [39].

Step 4: Obtain the polynomial matrices $\bar{\mathbf{G}}(\mathbf{B})$ and $\bar{\mathbf{L}}(\mathbf{B})$ in Model (9.1). Using the matrices $\bar{\mathbf{F}}_i$ in (9.9) the remaining polynomial matrices of the VARMAX repre-

sentation can be calculated as:

$$\bar{G}_0 = D, \quad \begin{bmatrix} \bar{G}_1 \\ \bar{G}_2 \\ \vdots \\ \bar{G}_{p_{\max}} \end{bmatrix} = T^{-1}\Gamma + \begin{bmatrix} \bar{F}_1\bar{G}_0 \\ \bar{F}_2\bar{G}_0 \\ \vdots \\ \bar{F}_{p_{\max}}\bar{G}_0 \end{bmatrix}, \quad \begin{bmatrix} \bar{L}_1 \\ \bar{L}_2 \\ \vdots \\ \bar{L}_{p_{\max}} \end{bmatrix} = T^{-1}E + \begin{bmatrix} \bar{F}_1 \\ \bar{F}_2 \\ \vdots \\ \bar{F}_{p_{\max}} \end{bmatrix}, \quad (9.10)$$

where D, Γ and E come from the original SEM and T is the transformation matrix.

9.2.2 From State-Space to canonical VARMAX

We will now discuss an alternative algorithm which is more general than the previous one, as it does not require any *a priori* conditions. The downside is that it is more complex. In particular, it requires one to determine the Kronecker indices of the system, which in the previous case were directly implied by conditions C.1 and C.2. The algorithm can be divided into two stages: obtaining the Luenberger Canonical Form, and deriving the corresponding VARMAX echelon coefficients.

Stage 1: Obtaining the LCF.

Step 1 (Enforcing minimality) and *Step 2* (Obtaining the equivalent SEM) are identical to those in the previous algorithm.

Step 3. Identifying the Kronecker indices. To this end, consider the observability matrix of the SEM:

$$O_n = [\mathbf{H}^T \quad (\mathbf{H}\Phi)^T \quad (\mathbf{H}\Phi^2)^T \quad \dots \quad (\mathbf{H}\Phi^{n-1})^T]^T. \quad (9.11)$$

Minimality assures that this matrix has n linearly independent rows. If these rows are chosen in descending order we can build a base, which after re-ordering can be written in block matrix form as:

$$\mathbf{M} = \left[\begin{array}{cccc} \mathbf{h}_1 & (\mathbf{h}_1^T \Phi)^T & \dots & (\mathbf{h}_1^T \Phi^{p_1-1})^T \\ & \dots & \mathbf{h}_m & (\mathbf{h}_m^T \Phi)^T & \dots & (\mathbf{h}_m^T \Phi^{p_m-1})^T \end{array} \right]^T, \quad (9.12)$$

where \mathbf{h}_k is the k -th row of \mathbf{H} , and p_k ($k = 1, \dots, m$) are the Kronecker indices, so that $\sum_{k=1}^m p_k = n$.

Step 4. Computing the LCF matrices. For any minimal SEM with Kronecker indices p_k ($k = 1, \dots, m$), there exists a nonsingular matrix T such that $\Phi^* = T^{-1}\Phi T$ and $\mathbf{H}^* = \mathbf{H}T$ have the LCF structure, that is:

$$\Phi^* = \begin{bmatrix} F_1 & Q_1 & 0 & \dots & 0 \\ F_2 & 0 & Q_2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ F_{p_{\max}-1} & 0 & 0 & \dots & Q_{p_{\max}-1} \\ F_{p_{\max}} & 0 & 0 & \dots & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{H}^* = [F_0 \quad \mathbf{0} \quad \mathbf{0}]. \quad (9.13)$$

being Φ^* a companion matrix, where each F_j block ($j = 1, \dots, p_{\max}$) has a number of rows equal to the number of Kronecker indices greater or equal to k , has $\bar{m} = \sum_{k=1}^m \min\{p_k, 1\}$ columns, and some null elements. In fact, the (k, l) -th element of F_j will be nonzero only if $j \in [p_k - p_{kl} + 1, p_k]$, where p_{kl} is defined as:

$$p_{kl} = \begin{cases} \min(p_k + 1, p_l) & \text{for } k \geq l \\ \min(p_k, p_l) & \text{for } k < l \end{cases} \quad k, l = 1, 2, \dots, m, \quad (9.14)$$

where the integers p_k , $k = 1, \dots, m$, are the Kronecker or observability indices.

Each Q_k block is a zero/one matrix, with as many columns as the number of observability indices that are greater or equal to k . If the endogenous variables are sorted according to their observability indices, the structure of Q_k will be $Q_k = [I_{k+1} \ 0]^T$, where I_{k+1} is an identity matrix with the same number of rows as F_{k+1} . With respect to H^* , F_0 is an $m \times \bar{m}$ matrix, such that the rows corresponding to components with nonzero observability indices can be organized in an $\bar{m} \times \bar{m}$ lower triangular matrix with ones in the main diagonal. A formal proof of this result and how to derive the transformation matrix T can be found in Casals, Garcia-Hiernaux, and Jerez [39].

Stage 2: Identifying the VARMAX echelon coefficients.

Transforming the SEM (2.19)–(2.20) to the LCF yields:

$$\boldsymbol{x}_{t+1}^* = \Phi^* \boldsymbol{x}_t^* + \Gamma^* \boldsymbol{u}_t + E^* \boldsymbol{a}_t \quad (9.15)$$

$$\boldsymbol{z}_t = H^* \boldsymbol{x}_t^* + D \boldsymbol{u}_t + \boldsymbol{a}_t, \quad (9.16)$$

where the matrices Φ^* and H^* have the LCF structure characterized in (9.13). In order to write this model in an equivalent polynomial form, it is convenient to augment the system dimension up to mp_{\max} by adding as many non-excited states as needed. Then, the structure of Φ^* will be as in an LCF, but with: (i) the identity matrix instead of Q_j , and (ii) an augmented dimension of matrices F_j , now $m \times m$. Note that the constraints on potentially nonzero parameters also affect these augmented matrices. Consequently, the new non-excited states require adding null columns to H^* except for the endogenous variables with a null observability matrix, so that the augmented F_j is an $m \times m$ lower triangular matrix with ones in its main diagonal. This particular structure allows us to write the observation equation as:

$$F_0^{-1} \boldsymbol{z}_t = \boldsymbol{x}_{1:m,t}^* + F_0^{-1} D \boldsymbol{u}_t + F_0^{-1} \boldsymbol{a}_t, \quad (9.17)$$

where $\boldsymbol{x}_{1:m,t}^*$ denotes the first m elements of the state vector \boldsymbol{x}_t^* . According to (9.17), one can isolate the $\boldsymbol{x}_{1:m,t}^*$ and substitute them into the state equation.

Finally, taking into account the companion structure of Φ^* , we obtain the coefficients in the VARMAX form (9.1) as:

$$\bar{\mathbf{F}}_0 = \mathbf{F}_0^{-1}, \quad \bar{\mathbf{F}}_j = -\mathbf{F}_j \mathbf{F}_0^{-1}, \quad (9.18)$$

$$\bar{\mathbf{G}}_0 = \mathbf{F}_0^{-1} \mathbf{D}, \quad \bar{\mathbf{G}}_j = \mathbf{\Gamma}^* + \bar{\mathbf{F}}_j \mathbf{D}, \quad (9.19)$$

$$\bar{\mathbf{L}}_0 = \bar{\mathbf{F}}_0, \quad \bar{\mathbf{L}}_j = \mathbf{E}^* + \bar{\mathbf{F}}_j, \quad j = 1, \dots, p_{\max}, \quad (9.20)$$

$$\bar{\mathbf{F}}(\mathbf{B}) = \sum_{j=0}^{p_{\max}} \bar{\mathbf{F}}_j \mathbf{B}^j, \quad \bar{\mathbf{G}}(\mathbf{B}) = \sum_{j=0}^{p_{\max}} \bar{\mathbf{G}}_j \mathbf{B}^j, \quad \bar{\mathbf{L}}(\mathbf{B}) = \sum_{j=0}^{p_{\max}} \bar{\mathbf{L}}_j \mathbf{B}^j. \quad (9.21)$$

Note that this representation has the characteristic structure of a canonical VARMAX echelon model; see Dickinson, Morf, and Kailath [79].

9.3 Practical applications and examples

9.3.1 The VARMAX form of some common State-Space models

SS models are adequate for many uses, such as extracting the structural components of a time series or working with nonstandard samples such as *e.g.*, those with missing values, aggregated data, or observation errors. On the other hand, VARMAX representations are more familiar in mainstream time series analysis, and they come equipped with a convenient set of diagnostic checks stemming from the fact that they produce a unique residual vector whose statistical properties are well understood; see *e.g.* Li [148]. To get the best of both worlds, one therefore needs the ability to derive the VARMAX form corresponding to a given SS model, bearing in mind that the literature describes how to write a VARMAX model in SS form; see Subsection 2.2.4.

Table 9.1 illustrates the results of the methods described in Section 9.2 by showing the ARIMAX structure corresponding to some common SS models. All the models in Table 9.1 are univariate, so these results could have been obtained via other approaches such as, *e.g.*, by identifying the autocorrelation function of the endogenous variable, see Chapter 2 of Harvey [118], or using the pseudo-spectrum implied by the unobserved components and reduced form models, see Bujosa, Garcia-Ferrer, and Young [33].

Table 9.2 shows that the methods described in the previous section can also be applied to vector models. Even in the simplest cases, it would be difficult to obtain the corresponding VARMAX form via the above mentioned autocorrelation and spectral approaches.

The code required to replicate this example can be found in the files `Example0931a.m` and `Example0931b.m`; see Appendix D.

9.3.2 Identifiability and conditioning of the estimates

The methods described in Section 9.2 are also useful to analyze two important issues: model identifiability and conditioning.

Table 9.1 Structural and ARIMAX forms for some common univariate specifications. Both representations are observationally equivalent.

Model	Structural representation		ARIMAX representation
(i) Dynamic regression with errors in variables	$x_{t+1} = .5x_t + .7u_t + w_t, \quad w_t \sim \text{IID}(0, 1.5)$ $z_t = x_t + v_t, \quad v_t \sim \text{IID}(0, 1)$		$(1 + .5B)z_t = .7u_{t-1} + (1 - .188B)a_t, \quad \sigma_a^2 = 2.660$
(ii) Random walk plus noise model	$\mu_{t+1} = \mu_t + \zeta_t, \quad \zeta_t \sim \text{IID}(0, 1/100)$ $z_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{IID}(0, 1)$		$(1 - B)z_t = (1 - .905B)a_t, \quad \sigma_a^2 = 1.105$
(iii) Integrated random walk plus noise model	$\mu_{t+1} = \mu_t + \beta_t, \quad \beta_t = \beta_{t-1} + \zeta_t, \quad \zeta_t \sim \text{IID}(0, 1/100)$ $z_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{IID}(0, 1)$		$(1 - B)^2 z_t = (1 - 1.558B + .638B^2)a_t, \quad \sigma_a^2 = 1.567$
(iv) Model (iii) with one input	$\mu_{t+1} = \mu_t + \beta_t, \quad \beta_{t+1} = \beta_t + \zeta_t, \quad \zeta_t \sim \text{IID}(0, 1/100)$ $z_t = \mu_t + .5u_t + \varepsilon_t, \quad \varepsilon_t \sim \text{IID}(0, 1)$		$(1 - B)^2 z_t = .5(1 - B)^2 u_t + (1 - 1.558B + .638B^2)a_t, \quad \sigma_a^2 = 1.567$
(v) Model (iii) with quarterly dummy variable seasonality	$\mu_{t+1} = \mu_t + \beta_t, \quad \beta_{t+1} = \beta_t + \zeta_t, \quad \zeta_t \sim \text{IID}(0, 1/100)$ $\gamma_{t+1} = -\gamma - \gamma_{t-1} - \gamma_{t-2} + w_t, \quad w_t \sim \text{IID}(0, 1/10)$ $z_t = \mu_t + \varepsilon_t, \quad \varepsilon_t \sim \text{IID}(0, 1)$		$(1 - B)(1 - B^4)z_t = (1 - .714B + .114B^2 - .010B^3 - .563B^4 + .438B^5)a_t, \quad \sigma_a^2 = 2.283$

Note: In all cases the errors affecting the state and observation equations are assumed to be independent.

Table 9.2 Structural and VARMAX forms for some common bivariate specifications. Both representations are observationally equivalent.

Model	Representations: Structural and VARMAX
(vi) Seemingly unrelated time series equations with inputs	$\begin{bmatrix} \mu_{t+1}^1 \\ \mu_{t+1}^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t^1 \\ \mu_t^2 \end{bmatrix} + \begin{bmatrix} \zeta_t^1 \\ \zeta_t^2 \end{bmatrix}, \quad \begin{bmatrix} \zeta_t^1 \\ \zeta_t^2 \end{bmatrix} \sim \text{IID} \left(\begin{bmatrix} 0 & 1/100 \\ 0 & 5/1000 \end{bmatrix}, \begin{bmatrix} 1/100 & 5/1000 \\ 5/1000 & 2/100 \end{bmatrix} \right);$ $\begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t^1 \\ \mu_t^2 \end{bmatrix} + \begin{bmatrix} .5 \\ .7 \end{bmatrix} u_t + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix} \sim \text{IID} \left(\begin{bmatrix} 0 & 1 \\ 0 & .2 \\ 0 & .5 \end{bmatrix}, \begin{bmatrix} 1 & .2 & .5 \end{bmatrix} \right).$ $\begin{bmatrix} 1-B & 0 \\ 0 & 1-B \end{bmatrix} \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} = \begin{bmatrix} .5 \\ .7 \end{bmatrix} (1-B)u_t + \begin{bmatrix} 1-.910B & -.012B \\ -.009B & 1-.816B \end{bmatrix} \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix},$ $\begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix} \sim \text{IID} \left(\begin{bmatrix} 0 & 1.104 \\ 0 & .232 \end{bmatrix}, \begin{bmatrix} 232 \\ 610 \end{bmatrix} \right).$
(vii) Two series with a common random walk trend	$\begin{bmatrix} \mu_{t+1} \\ z_t^1 \end{bmatrix} = \mu_t + \zeta_t, \quad \zeta_t \sim \text{IID}(0, 1/100);$ $\begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \end{bmatrix} \mu_t + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix} \sim \text{IID} \left(\begin{bmatrix} 0 & 1 \\ 0 & .2 \\ 0 & .5 \end{bmatrix}, \begin{bmatrix} 1 & .2 & .5 \end{bmatrix} \right).$ $\begin{bmatrix} 1-B & 0 \\ -1 & 1 \end{bmatrix} \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} = \begin{bmatrix} 1-.961B & .104B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix}, \quad \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix} \sim \text{IID} \left(\begin{bmatrix} 0 & 1.070 \\ 0 & .270 \end{bmatrix}, \begin{bmatrix} 270 \\ 570 \end{bmatrix} \right).$
(viii) Two series with a common trend and an autonomous random walk component	$\begin{bmatrix} \mu_{t+1}^1 \\ \beta_{t+1}^1 \\ \mu_{t+1}^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_t^1 \\ \beta_t^1 \\ \mu_t^2 \end{bmatrix} + \begin{bmatrix} 0 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \zeta_t^1 \\ \zeta_t^2 \end{bmatrix}, \quad \begin{bmatrix} \zeta_t^1 \\ \zeta_t^2 \end{bmatrix} \sim \text{IID} \left(\begin{bmatrix} 0 & 1/100 \\ 0 & 5/1000 \end{bmatrix}, \begin{bmatrix} 1/100 & 5/1000 \\ 5/1000 & 2/100 \end{bmatrix} \right);$ $\begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 1 & 1 \end{bmatrix} \begin{bmatrix} \mu_t^1 \\ \beta_t^1 \\ \mu_t^2 \end{bmatrix} + \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix}, \quad \begin{bmatrix} \varepsilon_t^1 \\ \varepsilon_t^2 \end{bmatrix} \sim \text{IID} \left(\begin{bmatrix} 0 & 1 \\ 0 & .2 \\ 0 & .5 \end{bmatrix}, \begin{bmatrix} 1 & .2 & .5 \end{bmatrix} \right).$ $\begin{bmatrix} (1-B)^2 & 0 \\ 0 & 1-B \end{bmatrix} \begin{bmatrix} z_t^1 \\ z_t^2 \end{bmatrix} = \begin{bmatrix} 1-1.582B+.650B^2 & .069B-.021B^2 \\ .036B & 1-.777B \end{bmatrix} \begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix},$ $\begin{bmatrix} a_t^1 \\ a_t^2 \end{bmatrix} \sim \text{IID} \left(\begin{bmatrix} 0 & 1.549 \\ 0 & .329 \end{bmatrix}, \begin{bmatrix} .329 \\ .659 \end{bmatrix} \right).$

Note: In all these models the errors affecting the state and observation equations are assumed to be independent.

A parametric model is said to be *identifiable* if no two parameter settings yield the same distribution of observations. By definition, canonical VARMAX models are always identifiable, while, in general, there are infinitely many SS models realizing the same reduced-form VARMAX (see discussion on similar transformations in Subsection 2.1.2). Therefore, SS models are unidentifiable unless one imposes an adequate structure on its parameter matrices.

Glover and Willem [108] provide analytical conditions for local and global identifiability of a linear SS model. These conditions are scientifically valuable, but very difficult to apply in practice. On the other hand, the algorithms described in Section 9.2 can be applied to discuss the local identifiability of a given SS model via the following numerical procedure.

Step 1: Compute the response of VARMAX coefficients to small perturbations in the SS parameters. These values would be finite-difference approximations to the corresponding partial derivatives.

Step 2: Organize these derivatives into a Jacobian matrix, \mathbf{J} , with as many columns as the number of free parameters in the SS model (N_{SS}), and as many rows as the number of parameters in the VARMAX form (N_V).

Step 3: Compute the rank of \mathbf{J} , denoted by $\text{rank}(\mathbf{J})$.

Step 4: Characterize the identifiability of the system according to the procedure described in Table 9.3.

Table 9.3 *Characterization of the identifiability of an SS model using the rank of the Jacobian (\mathbf{J}) measuring how a change in the structural model parameters affects the parameters in the reduced form.*

Condition	Identifiability
$\text{rank}(\mathbf{J}) < N_{SS}$	The SS model parameters are underidentified
$\text{rank}(\mathbf{J}) = N_{SS} = N_V$	The SS model parameters are exactly identified
$\text{rank}(\mathbf{J}) = N_{SS} < N_V$	The SS model parameters are overidentified

Note: N_{SS} and N_V are, respectively, the number of parameters in the SS and VARMAX representation.

An efficient way to perform Step 3 above consists of computing the singular-value decomposition of \mathbf{J} in Step 2; see Appendix A.4. Note that a null singular value corresponds to a linear combination of the VARMAX parameters that is not affected by perturbations in the SS model parameters, thereby pointing to a specific source of non-identifiability. Accordingly, the rank of \mathbf{J} is the number of non-zero singular values.

Moreover, by defining the transformation as the function $P_V = f(P_{SS})$, being P_V and P_{SS} vectorizations of the set of VARMAX and SS parameters, respectively, one can use the Jacobian \mathbf{J} to compute its condition number as $c[f(P_{SS})] = \|\mathbf{J}\| \cdot \|P_{SS}\| / \|P_V\|$. This value measures the sensitivity of the VARMAX coefficients with respect to changes in the SS model parameters, so that a small conditioning

number means that the SS coefficients are ill-identified because they are insensitive to changes in the VARMAX parameters².

Consider for example the following SS model:

$$x_{t+1} = -\phi x_t + w_t, \quad w_t \sim \text{IID}(0, 0.1) \quad (9.22)$$

$$z_t = x_t + v_t, \quad v_t \sim \text{IID}(0, 1), \quad (9.23)$$

with $E(w_t, v_t) = 0$, which has an ARMA(1,1) reduced form:

$$(1 + \phi B)z_t = (1 + \theta B)a_t, \quad a_t \sim \text{IID}(0, \sigma_a^2) \quad (9.24)$$

where the AR parameter in (9.24) coincides with the opposite of the transition scalar in (9.22). Figure 9.1 displays the values of the θ and σ_a^2 parameters associated with different values of ϕ , as well as the smallest singular value of the Jacobian defined above in each case. Note that, when $\phi = 0$, the corresponding singular value is null. In this case, the structural model degenerates to the sum of two white noise processes, and is therefore unidentifiable.

The code required to replicate this example can be found in the file `Example0932.m`; see Appendix D.

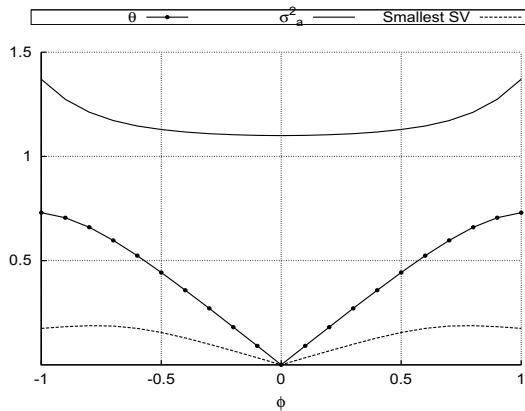


Figure 9.1 *Parameters of the ARMA model (9.24) and smallest singular value (SV) of the Jacobian for several values of ϕ in the AR(1) plus error model.*

9.3.3 Fitting an errors-in-variables model to Wolf's sunspot series

This example illustrates the use of the procedures described in Section 9.2 to perform diagnostic checking on a given estimated SS model.

²The term “ill-identification” refers here to a relaxation of the idea of strict non-identification, this being a situation where a statistical model has more than one set of parameters generating the same distribution of observations, implying that multiple parameterizations are observationally equivalent.

Consider the annual series of Wolf's Sunspot Numbers 1700–1988, taken from [210]. Several prior analyses of this series conclude that it has a harmonic cycle with a period of about 11 years. On the other hand, the dataset is composed of records compiled by human observers using optical devices of diverse qualities, so it seems natural to infer that the recorded values are affected by observation errors.

Building on these two ideas, we fitted an AR(2) plus white noise errors model to the original data divided by 10, to improve the scaling of the parameters. The resulting estimates are:

$$(1 - 1.444B + .743B^2)\hat{z}_t^* = 1.476 + \hat{w}_t, \quad \hat{\sigma}_w^2 = 2.205, \\ z_t = \hat{z}_t^* + \hat{v}_t, \quad \hat{\sigma}_v^2 = 0.147, \quad \ell^*(\hat{\theta}) = 554.246, \quad (9.25)$$

where z_t and \hat{z}_t^* are, respectively, the scaled Wolf number at year t , and the underlying “error free” value. $\ell^*(\hat{\theta})$ denotes the minus log-likelihood value corresponding to the estimates. Note that the latent AR(2) structure has complex roots, which implies that the data follows a damped cycle with a period of 10.87 years.

Using the algorithm described in Subsection 9.2.1, the SS representation of (9.25) can be written as the ARMA(2,2) model:

$$(1 + 1.444B + .743B^2)z_t = 1.476 + (1 - .133B + .041B^2)\hat{a}_t, \\ \hat{\sigma}_{\hat{a}}^2 = 2.689, \quad Q(8) = 10.59, \quad \ell^*(\hat{\theta}) = 554.246, \quad (9.26)$$

where $Q(8)$ is the residual portmanteau Q statistic computed with 8 lags. Note that the likelihood of this model coincides with that of (9.25).

Model (9.26) has six parameters while (9.25) has only five. Therefore, the latter is an overidentified structural form. We can check that the structural model's (overidentification) constraint is consistent with the data, by freely estimating the parameters in (9.26):

$$(1 + 1.428B + .733B^2)z_t = 1.509 + (1 - .112B + .064B^2)\hat{a}_t, \\ \hat{\sigma}_{\hat{a}}^2 = 2.688, \quad \ell^*(\hat{\theta}) = 554.169. \quad (9.27)$$

Therefore, models (9.26) and (9.27) are almost identical. Their statistical consistency can be formally tested by computing a likelihood ratio statistic; for which the value of 0.154 confirms that the structural constraint is consistent with the data.

The code required to replicate this example can be found in the file `Example0933.m`; see Appendix D.

9.3.4 “Bottom-up” modeling of quarterly US GDP trend

The model-building sequence adopted in the above Wolf's sunspot series example can be described as “top-down,” meaning that we first fit a structural (“top”) model and then obtain the corresponding ARIMA (“down”) reduced form.

We will now demonstrate that the methods presented in Section 9.2 can be used to implement a “bottom-up” modeling strategy, whereby one first fits a reduced-form

model to the data, and then computes the structural model parameters which (either exactly or approximately) realize the reduced form. This approach, originally proposed by Nerlove, Grether, and Carvalho [166], is justified if one wants to combine the advantages of a structural SS model with the ability of reduced-form models to capture the data sample properties.

Note also that this idea is closely related to the notion of ARIMA-based time series decomposition, originally suggested by Hillmer and Tiao [125], where a reduced form ARIMA model is decomposed into the sum of several structural components; see Chapter 8.

Now consider the quarterly and seasonally adjusted series of US Gross Domestic Product (GDP_t), from the 1st quarter of 1947 to the 3rd quarter of 2008, in constant 2000 US Dollars. The GDP trend is often extracted with the filter proposed by Hodrick and Prescott [126], which is equivalent to the smoothed trend obtained from the SS model:

$$\begin{aligned}\mu_{t+1} &= \mu_t + \beta_t \\ \beta_{t+1} &= \beta_t + \zeta_t \\ z_t &= \mu_t + \varepsilon_t \\ \begin{bmatrix} \zeta_t \\ \varepsilon_t \end{bmatrix} &\sim \text{IID} \left(\begin{bmatrix} \zeta_t \\ \varepsilon_t \end{bmatrix}, \begin{bmatrix} \sigma_\zeta^2 & 0 \\ 0 & \sigma_\varepsilon^2 \end{bmatrix} \right),\end{aligned}\tag{9.28}$$

with a signal-to-noise variance ratio such that $\sigma_\zeta^2/\sigma_\varepsilon^2 = 1/1600$; see Harvey and Trimbur [120].

While the Hodrick–Prescott filter is a simple and effective tool to extract a smooth long-term trend component, it does not capture the dynamics of the data well. In this case, if we fit model (9.28) to the series $z_t = \log(GDP_t) \times 100$, we obtain the ML variance estimates $\sigma_\varepsilon^2 = 1.359$ and $\sigma_\zeta^2 = 1.359/1600$. Applying our method to this model yields the reduced form ARIMA model:

$$(1 - B)^2 \log(GDP_t) \times 100 = (1 - 1.777B + 0.799B^2)\hat{\alpha}_t, \\ \sigma_{\hat{\alpha}}^2 = 1.699, \quad Q(15) = 239.82, \quad \ell^*(\hat{\theta}) = 174.8,\tag{9.29}$$

where the large value of the residual Q -statistic suggests that a strict Hodrick–Prescott specification does not capture all the autocorrelation in this series. Therefore, we may want to adjust a trend model with the dynamic structure of (9.28) so that it realizes a previously fitted ARIMA model. This modeling strategy can be implemented with the following procedure:

Step (1) Fit an ARIMA (or VARMAX, in a multiple time series framework) model to the dataset.

Step (2) Compute the SS model parameters which realize more closely the model previously fitted. This requires a non-linear iterative procedure to minimize a given loss function. In this example we specify this function as the square root of the difference between the parameters of:

Step (2.a) ...the model fitted in Step (1), and

Step (2.b) ...those of the reduced-form corresponding to the SS model³.

Table 9.4 summarizes the results of the bottom-up sequence applied to the GDP data. In Step (1) we fitted an ARIMA model to $z_t = \log(GDP_t) \times 100$. Note that its parameters are very different from those of model (9.29).

In Step (2.a) we estimated the two variances of an integrated random-walk model by minimizing the loss function defined above and the corresponding reduced-form model. Note that the latter is similar but not identical to the model in Step (1), so both models are not strictly exactly equivalent.

On the other hand, comparing the models in Steps (1) and (2.a) we note immediately that the latter is overidentified, as it only has two free parameters. In Step (2.b) we freed the null constraint imposed on the model covariance, in order to improve the fit between it and the reduced-form. The results show clearly that both are equivalent so we can conclude that, without the zero-correlation overidentifying constraint, the dynamic structure underlying the HP filter model could be flexible enough to capture most of the data autocorrelation.

The code required to replicate this example can be found in the file `Example0934.m`; see Appendix D.

³Note that there are many valid specifications for the loss function employed in Step (2). For example, one could minimize the squared sum of the difference between the log-likelihood of both models, or the residual series generated by both models. These alternative functions would be particularly useful if the SS model cannot realize exactly the reduced form model.

Table 9.4 Step-by-step results from a bottom-up modeling of the US GDP series.

Step	Structural model	ARIMAX model
(1) Fit a reduced-form model to the series	NONE	$(1 - B)^2 \log(GDP_t) \times 100 = (1 - .661B + .000^* B^2)\hat{a}_t;$ $\sigma_{\hat{a}}^2 = .250, Q(15) = 16.54, \ell^*(\hat{\theta}) = 75.371$
(2.a) Estimate the integrated random walk constraining the covariance to zero	$\text{cov} \begin{bmatrix} \zeta_t \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} .083 & 0^* \\ 0^* & .044 \end{bmatrix}$	$(1 - B)^2 \log(GDP_t) \times 100 = (1 - .600B + .176B^2)\hat{a}_t;$ $\sigma_{\hat{a}}^2 = .250, Q(15) = 22.64, \ell^*(\hat{\theta}) = 77.173$
(2.b) Estimate the integrated random walk freeing the covariance constraint	$\hat{\text{cov}} \begin{bmatrix} \zeta_t \\ \varepsilon_t \end{bmatrix} = \begin{bmatrix} .029 & -.083 \\ -.083 & .083 \end{bmatrix}$	$(1 - B)^2 \log(GDP_t) \times 100 = (1 - .661B + .000B^2)\hat{a}_t;$ $\sigma_{\hat{a}}^2 = .250, Q(15) = 16.54, \ell^*(\hat{\theta}) = 75.371$

Notes: The figure in parenthesis is the standard error of the estimate; some ARIMAX parameters do not have standard errors because they have been derived from the structural SS model estimates. The values $Q(15)$ and $\ell^*(\hat{\theta})$ represent, respectively, the Ljung–Box [15,1] Q statistic computed with 15 lags of the residual autocorrelation function, and the minus log (Gaussian) likelihood upon convergence. The figures with an asterisk correspond to constrained parameters.

This page intentionally left blank

Chapter 10

Aggregation and disaggregation of time series

10.1	The effect of aggregation on an SS model	146
10.1.1	The high-frequency model in stacked form	146
10.1.2	Aggregation relationships	148
10.1.3	Relationships between the models for high, low, and mixed-frequency data	149
10.1.4	The effect of aggregation on predictive accuracy	150
10.2	Observability in the aggregated model	150
10.2.1	Unobservable modes	150
10.2.2	Observability and fixed-interval smoothing	152
10.2.3	An algorithm to aggregate a linear model: theory and examples	153
10.3	Specification of the high-frequency model	154
10.3.1	Enforcing approximate consistency	154
10.3.2	“Bottom-up” determination of the quarterly model	158
10.4	Empirical example	158
10.4.1	Annual model	159
10.4.2	Decomposition of the quarterly indicator	159
10.4.3	Specification and estimation of the quarterly model	160
10.4.4	Diagnostics	160
10.4.5	Forecast accuracy and non-conformable samples	162
10.4.6	Comparison with alternative methods	163

This chapter discusses how to model, interpolate, and forecast a vector of time series sampled at different frequencies. To avoid cumbersome wording, we will refer to the low-frequency variables as “annual” and to the high-frequency variables as “quarterly.” The results, however, are valid for any combination of sampling frequencies.

Our starting point consists of breaking the global problem into four basic questions: How could one specify a quarterly model from a sample combining quarterly and annual data? How could such a model be estimated? How can we compute the within-sample estimates of the unobserved quarterly values? How can we forecast the annual variables, exploiting the quarterly information available? All these issues, except the first, have been addressed in previous chapters. In particular, Chapter 5 discussed the computation of the Gaussian likelihood for any fixed parameter SS model, thus solving the problem of estimation, and Chapter 4 presented the fixed-interval smoothing algorithm, which can be used after estimation to estimate and forecast the quarterly values.

Therefore, in this chapter we concentrate on the specification problem. To this end we study how aggregation over time affects the dynamic structure of a vector of time series and its observability. We find that the basic dynamic components remain unchanged but some of them, in particular those related to seasonality, become unobservable. Building on these results, we present a structured specification method based on the idea that the models relating the variables in high and low sampling frequencies should be mutually consistent. Once a consistent and observable high-frequency model has been specified, estimation, diagnostic checking, interpolation, and forecasting can be addressed by standard SS methods. An example with national accounting data illustrates the practical application of these ideas.

10.1 The effect of aggregation on an SS model

Let \mathbf{z}_t be an $m \times 1$ random vector of quarterly values, realized by the minimal SEM:

$$\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \Gamma \mathbf{u}_t + \mathbf{E} \mathbf{a}_t \quad (10.1)$$

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_t + \mathbf{D} \mathbf{u}_t + \mathbf{a}_t. \quad (10.2)$$

10.1.1 The high-frequency model in stacked form

Let S be the seasonal frequency, defined as the number of high-frequency sampling periods (quarters) corresponding to a single low-frequency (annual) observation. It is difficult to discuss aggregation over time using the quarterly representation (10.1)–(10.2). To this end, it is more convenient to use a representation where the values of \mathbf{z}_t , \mathbf{u}_t , and \mathbf{a}_t corresponding to a whole seasonal cycle are “stacked” into the following vectors:

$$\mathbf{z}_{t:t+S-1} = \begin{bmatrix} \mathbf{z}_t \\ \mathbf{z}_{t+1} \\ \vdots \\ \mathbf{z}_{t+S-1} \end{bmatrix}, \quad \mathbf{u}_{t:t+S-1} = \begin{bmatrix} \mathbf{u}_t \\ \mathbf{u}_{t+1} \\ \vdots \\ \mathbf{u}_{t+S-1} \end{bmatrix}, \quad \mathbf{a}_{t:t+S-1} = \begin{bmatrix} \mathbf{a}_t \\ \mathbf{a}_{t+1} \\ \vdots \\ \mathbf{a}_{t+S-1} \end{bmatrix}. \quad (10.3)$$

Without loss of generality we assume that the aggregation period coincides with the seasonal frequency, so the stacked vectors in (10.3) include all the values subject to aggregation. Under these conditions, the following proposition holds.

Proposition 10.1 *Model (10.1)–(10.2) can be written equivalently as:*

$$\boldsymbol{x}_{T+1} = \bar{\Phi} \boldsymbol{x}_T + \bar{\Gamma} \boldsymbol{u}_{t:t+S-1} + \bar{\mathbf{E}} \boldsymbol{a}_{t:t+S-1} \quad (10.4)$$

$$\boldsymbol{z}_{t:t+S-1} = \bar{\mathbf{H}} \boldsymbol{x}_T + \bar{\mathbf{D}} \boldsymbol{u}_{t:t+S-1} + \bar{\mathbf{C}} \boldsymbol{a}_{t:t+S-1} \quad (10.5)$$

where the time indexes T and t refer respectively to the annual and quarterly time units so that, if $\boldsymbol{x}_T = \boldsymbol{x}_t$, then $\boldsymbol{x}_{T+1} = \boldsymbol{x}_{t+S}$, and the matrices in (10.4)–(10.5) are related to those in (10.1)–(10.2) by the expressions:

$$\bar{\Phi} = \Phi^S \quad (10.6)$$

$$\bar{\Gamma} = [\Phi^{S-1} \Gamma, \Phi^{S-2} \Gamma, \dots, \Gamma] \quad (10.7)$$

$$\bar{\mathbf{E}} = [\Phi^{S-1} \mathbf{E}, \Phi^{S-2} \mathbf{E}, \dots, \mathbf{E}] \quad (10.8)$$

$$\bar{\mathbf{H}} = \begin{bmatrix} \mathbf{H} \\ \mathbf{H}\Phi \\ \vdots \\ \mathbf{H}\Phi^{S-1} \end{bmatrix} \quad (10.9)$$

$$\bar{\mathbf{D}} = \begin{bmatrix} \mathbf{D} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\Gamma & \mathbf{D} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}\Phi^{S-2}\Gamma & \mathbf{H}\Phi^{S-3}\Gamma & \dots & \mathbf{D} \end{bmatrix} \quad (10.10)$$

$$\bar{\mathbf{C}} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{H}\mathbf{E} & \mathbf{I} & \dots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{H}\Phi^{S-2}\mathbf{E} & \mathbf{H}\Phi^{S-3}\mathbf{E} & \dots & \mathbf{I} \end{bmatrix} \quad (10.11)$$

$$\bar{\mathbf{Q}} = \mathbf{I} \otimes \mathbf{Q}. \quad (10.12)$$

Proof. Equation (10.5) is obtained by successively substituting (10.1) in (10.2) and writing the resulting system in matrix form. On the other hand (10.4) is immediately obtained by propagating (10.1) from t to $t+S$. ■

Note that, according to expression (10.6), the size of the state vector in the stacked model (10.4)–(10.5) is the same as that in the SEM (10.1)–(10.2). This is a critical feature of the stacked model, as any increase in the size of the state vector would imply that the representation (10.4)–(10.5) has either redundant states or spurious dynamic structures, not present in the original high-frequency SEM.

10.1.2 Aggregation relationships

Now consider the aggregation relationships:

$$\mathbf{z}_T^A = \mathbf{J}^A \mathbf{z}_{t:t+S-1}, \quad (10.13)$$

$$\mathbf{z}_T^P = \mathbf{J}^P \mathbf{z}_{t:t+S-1}, \quad (10.14)$$

where \mathbf{z}_T^A and \mathbf{z}_T^P denote, respectively, the vectors of annual and partially aggregated data observed in year T , including the quarterly values at time $t, t+1, \dots, t+S-1$. By “partially aggregated data” we mean a mixed-frequency sample combining all of the observed annual and quarterly values; see Examples 10.1.2.1 and 10.1.2.2 below.

Note also that there always exists a relationship between the annual and partially aggregated series:

$$\mathbf{z}_T^A = \mathbf{J}^* \mathbf{z}_T^P, \quad (10.15)$$

where (10.13)–(10.14) imply $\mathbf{J}^* \mathbf{J}^P = \mathbf{J}^A$; see Example 10.1.2.3 below.

The aggregation matrices in (10.13), (10.14) and (10.15) depend on the aggregation constraints and the nature of the variables. Following Di Fonzo [76], there are four basic types of quarterly variables: flows, indices, beginning-of-period stocks, or end-of-period stocks. Accordingly, the annual samples will be sums of quarterly values, averages, or discrete beginning-of-period/end-of-period values. The following examples illustrate some common aggregation structures:

Example 10.1.2.1

If all the variables considered are flows then $\mathbf{J}^A = [\mathbf{I}, \mathbf{I}, \dots, \mathbf{I}]$, meaning that each annual figure is the sum of the corresponding quarterly values. On the other hand, $\mathbf{J}^A = [\mathbf{0}, \mathbf{0}, \dots, \mathbf{I}]$ would imply that the variables are end-of-period stocks, such that the value \mathbf{z}_{t+S-1} is observed while the values at $t, t+1, \dots, t+S-2$ are not.

Example 10.1.2.2

Assume that the first m_1 variables in the vector $\mathbf{z}_{t:t+S-1}$ of (10.3) are annual flows, while the remaining m_2 variables ($m_1 + m_2 = m$) are quarterly values. In this case the partial aggregation matrix would be:

$$\mathbf{J}^P = \begin{bmatrix} \mathbf{I}_{m_1} & \mathbf{0} & \mathbf{I}_{m_1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{I}_{m_2} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I}_{m_2} & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{I}_{m_2} \end{bmatrix}_{(m_1+S\cdot m_2) \times (S \cdot m_1 + S \cdot m_2)}, \quad (10.16)$$

where I_m is the $m \times m$ identity and the vectors of quarterly and partially observed values have the following structures:

$$\mathbf{z}_{t:t+S-1} = \begin{bmatrix} z_t^{m_1} \\ z_t^{m_2} \\ z_{t+1}^{m_1} \\ z_{t+1}^{m_2} \\ \vdots \\ z_{t+S-1}^{m_1} \\ z_{t+S-1}^{m_2} \end{bmatrix}_{(S \cdot m_1 + S \cdot m_2) \times 1} \quad (10.17)$$

$$\mathbf{z}_T^P = \begin{bmatrix} z_t^{m_1} + z_{t+1}^{m_1} + \dots + z_{t+S-1}^{m_1} \\ z_t^{m_2} \\ z_{t+1}^{m_2} \\ \vdots \\ z_{t+S-1}^{m_2} \end{bmatrix}_{(m_1 + S \cdot m_2) \times 1}. \quad (10.18)$$

Example 10.1.2.3

Assuming again that all the variables are flows, the matrix \mathbf{J}^* in (10.15), transforming partially aggregated data into annual values, would be:

$$\mathbf{J}^* = \begin{bmatrix} I_{m_1} & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & I_{m_2} & I_{m_2} & \cdots & I_{m_2} \end{bmatrix}_{(m_1 + m_2) \times (m_1 + S \cdot m_2)}. \quad (10.19)$$

10.1.3 Relationships between the models for high, low, and mixed-frequency data

Under the previous conditions, the following result formalizes the relationship between the quarterly data model in the form (10.4)–(10.5) and the corresponding models for the annual and partially observed data:

Proposition 10.2 *The models for \mathbf{z}_T^A and \mathbf{z}_T^P have the state equation (10.4) with the observation equations:*

$$\mathbf{z}_T^A = \mathbf{H}^A \mathbf{x}_T + \mathbf{D}^A \mathbf{u}_{t:t+S-1} + \mathbf{C}^A \mathbf{a}_{t:t+S-1} \quad (10.20)$$

$$\mathbf{z}_T^P = \mathbf{H}^P \mathbf{x}_T + \mathbf{D}^P \mathbf{u}_{t:t+S-1} + \mathbf{C}^P \mathbf{a}_{t:t+S-1}, \quad (10.21)$$

where:

$$\mathbf{H}^A = \mathbf{J}^A \bar{\mathbf{H}}, \quad \mathbf{D}^A = \mathbf{J}^A \bar{\mathbf{D}}, \quad \mathbf{C}^A = \mathbf{J}^A \bar{\mathbf{C}}, \quad (10.22)$$

$$\mathbf{H}^P = \mathbf{J}^P \bar{\mathbf{H}}, \quad \mathbf{D}^P = \mathbf{J}^P \bar{\mathbf{D}}, \quad \mathbf{C}^P = \mathbf{J}^P \bar{\mathbf{C}}. \quad (10.23)$$

Proof. Pre-multiplying (10.4) by \mathbf{J}^A and \mathbf{J}^P immediately yields (10.20) and (10.21) respectively. ■

As a corollary to proposition 10.2, note that the observer of z_T^A can be alternatively obtained by aggregation of z_T^P . Pre-multiplying (10.21) by \mathbf{J}^* yields:

$$z_T^A = \mathbf{J}^* z_T^P = \mathbf{J}^* H^P x_T + \mathbf{J}^* D^P u_{t:t+S-1} + \mathbf{J}^* C^P a_{t:t+S-1}. \quad (10.24)$$

10.1.4 The effect of aggregation on predictive accuracy

The following proposition states that a model combining annual and quarterly data may predict the annual values better than the corresponding annual model.

Proposition 10.3 *The forecast error variance of the annual model, given by the state equation (10.4) and the observation equation (10.20), is greater than or equal to the error variance of the annual forecasts computed by:*

- Case (1): aggregation of the forecasts for z_t derived from the quarterly model (10.4)–(10.5), and
- Case (2): aggregation of the forecasts for z_T^P derived from the partially aggregated model given by the state equation (10.4) and the observation equation (10.21).

Proof: See Appendix A in Casals, Sotoca, and Jerez [46]¹.

Proposition 10.3 compares the ability of three models to predict the observable annual values: (a) the model for the annual sample, (b) the “true” quarterly data generating process, and (c) the model for the partially aggregated sample. If these models are mutually consistent, given the aggregation constraint, model (b) would provide optimal forecasts, but cannot be empirically specified. Model (a) can be specified by standard methods, but cannot predict better than (b) or (c). Last, model (c) can be inferred from data using, e.g. the method defined in the next Section, and may provide better forecasts than model (a). It is also more flexible, because it has the ability to update the annual forecast as new quarterly information becomes available.

10.2 Observability in the aggregated model

10.2.1 Unobservable modes

As we saw previously, the SS models for quarterly, annual and partially aggregated data have the same state equation (10.4), so aggregation does not affect the states governing a dynamic system. However, it may reduce its observability.

¹This proposition is an extension of previous results. In particular, Wei [220] proved Case (1) for univariate processes, also showing that the loss in forecasting efficiency due to aggregation can be substantial if the nonseasonal component of the model is nonstationary. Conversely, there is no loss in efficiency if the quarterly model is a pure seasonal process. Lutkepohl [155] also discussed Case (1) in a multivariate stationary framework.

Comparing the observation equations (10.5) and (10.20)–(10.21) it is immediate to see that observability loss occurs in two ways: (a) the quarterly observer (10.5) realizes more signals than the aggregated observers (10.20) or (10.21), and (b) the matrices in (10.20)–(10.21) are non-linear functions of the matrices in (10.5), potentially decreasing the observability of some states.

To further discuss this issue, the following definitions particularize the general concept of observability (see Definition 2.4 in Chapter 2) to the notation previously defined.

Definition 10.1 (observability in the quarterly model). All the components in the quarterly model (10.1)–(10.2) are said to be observable if and only if there is no real vector $\mathbf{w} \neq \mathbf{0}$, with $\Phi \mathbf{w} = \lambda \mathbf{w}$ such that $\mathbf{H} \mathbf{w} = \mathbf{0}$, where λ is the eigenvalue associated to the eigenvector \mathbf{w} .

Definition 10.2 (unobservable modes in the annual model). Assuming that the variables are flows, so $\mathbf{J}^A = [\mathbf{I}, \mathbf{I}, \dots, \mathbf{I}]$, the annual model given by (10.4) and (10.20) has unobservable modes if and only if there exists a vector $\mathbf{w} \neq \mathbf{0}$, with $\Phi^S \mathbf{w} = \lambda \mathbf{w}$ such that $\mathbf{H}^A \mathbf{w} = \mathbf{0}$ or, equivalently, $\mathbf{H} \sum_{i=0}^{S-1} \Phi^i \mathbf{w} = \mathbf{0}$; see (10.9) and (10.22).

Under these conditions, the following Theorem characterizes which components of the quarterly model become unobservable after annual aggregation.

Theorem 10.1 (loss of observability in the annual model). Assuming that the quarterly model (10.4)–(10.5) is observable, the annual model given by the state equation (10.4) and the observation equation (10.20) includes unobservable modes in any of the following cases.

- Case (1): when the annual variables are flows, such that $\mathbf{J}^A = [\mathbf{I}, \mathbf{I}, \dots, \mathbf{I}]$, if there exists an eigenvalue λ of Φ such that $1 + \lambda + \lambda^2 + \dots + \lambda^{S-1} = 0$.
- Case (2): if there is a subset of k eigenvalues of Φ , denoted by $\lambda_k(\Phi)$, such that for any pair $\lambda_i, \lambda_j \in \lambda_k(\Phi)$, with $\lambda_i \neq \lambda_j$, it holds that $(\lambda_i)^S = (\lambda_j)^S$.
- Case (3): when Φ has a $k \times k$ Jordan block with null eigenvalues, such that the geometric multiplicity of this block grows with its S -th order power.

Proof: See Appendix B in Casals, Sotoca, and Jerez [46]².

Theorem 10.1 characterizes the situation where some modes of the quarterly model, associated to different transition eigenvalues, collapse to a smaller number of modes in the annual model. Consider, e.g., a quarterly univariate model where all the dynamics of a flow variable are given by a seasonal difference $\nabla_S =$

²Two important works connected with this Theorem are those by Stram and Wei [204] and Wei and Stram [222]. In particular, the loss of observability is related with the “hidden periodicity” notion defined by Stram and Wei [204] for univariate models. In a univariate framework, Case (2) of Theorem 10.1 is equivalent to hidden periodicity because all the indistinguishable modes created by aggregation affect a single time series. However, in the multivariate case hidden periodicity does not imply loss of observability in all cases. For example, observability is not lost if two modes with hidden periodicity affect two different time series. Note also that the mode elimination rules discussed for cases (2) and (3) are coherent in the univariate case with Theorems 4.1 and 3.1, respectively, in Stram and Wei [204].

$(1 - B)(1 + B + B^2 + \dots + B^{S-1})$. Case (1) implies that, after aggregation, this structure is indistinguishable from a nonseasonal difference $\nabla = (1 - B)$ in the annual model. Then, observability of $S-1$ dynamic components of the quarterly model is lost, and the dynamics of the corresponding annual model simplify accordingly.

Under Case (2), aggregation reduces the number of system modes, while maintaining the dimension of the state vector. The number of unobservable modes is $k - \text{rank}(\mathbf{H} \cdot \mathbf{W})$, where k is the number of elements in $\lambda_k(\Phi)$ and \mathbf{W} is a matrix whose columns are the eigenvectors \mathbf{w}_i , $i = 1, \dots, k$, that generate the k -order subspace \mathbf{S}_k . In the univariate case aggregation keeps only one mode, which corresponds with the larger Jordan block. In the multivariate case the reduction in the number of modes depends on how they affect the different observable time series so, in this case, there is no way to compute *a priori* how many modes will become unobservable, but the maximum number of unobservable modes is $k - 1$.

Case (3) occurs when the seasonal structure is a pure moving average. In this situation a systematic rule for mode elimination only exists in the univariate case, where there are $1 + [(k - 1)/S]$ distinguishable modes, with $[\cdot]$ denoting the integer part operator and k the dimension of the corresponding Jordan block.

After discussing the effect of aggregation on the quarterly model dynamics and its observability, it is important to characterize the correspondence between the models for disaggregated and aggregated data. This is done in the following theorem.

Theorem 10.2 (correspondence between quarterly and annual models). *If the quarterly model (10.4)–(10.5) is minimal and all its components are observable from annual data, then the annual model given by (10.4) and (10.20) is unique, allowing for similar transformations (minimal and observable).*

Proof: See Appendix 10.3 in Casals, Sotoca, and Jerez [46]³.

The converse of this result is not true in general, as there are minimal and observable annual models which necessarily correspond to quarterly models with unobservable components. Assume e.g. that the annual model is an AR(1) with a negative parameter and that the seasonal frequency S is even. In this case, there is no high-frequency ARMA(1,1) process that adds up to the annual AR(1) model because the S -th power of the transition matrix will always be positive; see (10.6).

10.2.2 Observability and fixed-interval smoothing

In an SS framework, the optimal method to estimate the system states is the fixed-interval smoother; see Chapter 4. The uncertainty of smoothed estimates depends on the system detectability, see Chapter 2 and, building on this idea, the following result characterizes the effect of non-detectable modes on smoothed estimates.

Proposition 10.4 *The variance of the smoothed estimates of the states in the models given by the state equation (10.4) and the observation equation (10.20) or (10.21) is finite if and only if all the states are detectable.*

³In the case of ARIMA models this result collapses to Lemma 2 in Wei and Stram [222].

Proof: See Appendix 10.3 in Casals, Sotoca, and Jerez [46].

In time series disaggregation, smoothing is typically applied to estimate the unobserved quarterly values. Therefore detectability is a necessary and sufficient condition for these estimates to have a bounded uncertainty, while observability is a sufficient (not necessary) condition.

Infinite smoothed variances arise, for example, when the target variables are flows and their quarterly model includes seasonal roots in the unit circle. In this case the aggregated model has non-detectable components which would have infinite variances. In practice this means that if quarterly indicators have seasonal components, it is advisable to remove them before disaggregation because they can induce spurious dynamics in the disaggregated series.

10.2.3 An algorithm to aggregate a linear model: theory and examples

The VARMAX annual model corresponding to any linear quarterly model can be obtained by combining Propositions 10.1 and 10.2 with Theorem 10.1 and the algorithms described in Chapter 9. The step-by-step procedure to do this is as follows.

Step (1) Quarterly VARMAX Consider any linear and fixed-coefficients quarterly model. Write the model in the SEM form (10.1)–(10.2). If required, first write the MEM form and then compute the parameter matrices of the equivalent SEM; see Subsection 2.2.3. Finally, obtain the VARMAX model equivalent to the SEM using the procedures described in Chapter 9.

Step (2) Annual model Obtain the equivalent quarterly representation (10.4)–(10.5) and the corresponding annual representation, given by (10.4) and (10.20).

Step (3): Annual VARMAX Apply either the algorithm described in Subsection 9.2.1 to obtain the standard VARMAX representation, or the one described in Subsection 9.2.2, to obtain the corresponding canonical VARMAX echelon model.

Tables 10.1 and 10.2 illustrate the application of this algorithm and some previous results by showing the aggregation of several univariate and bivariate models. In particular:

1. Models 1–3 exemplify the observability loss described in Theorem 10.1 cases (1), (2) and (3), respectively. In particular, aggregation of Model 1 shows how a seasonal difference collapses to a nonseasonal unit root. This result is consistent with Granger and Siklos [113] and Stram and Wei [204].
2. Models 1, 4, 5, and 7–11 show that aggregation does not affect the number of unit roots in a time series, so $I(0)$, $I(1)$ or $I(2)$ quarterly flows yield $I(0)$, $I(1)$ or $I(2)$ annual aggregates. A straightforward implication of this is that if the quarterly variables are cointegrated, the corresponding annual aggregates will also be cointegrated, this result being consistent with Pierse and Snell [183] and Marcellino [161].

3. Models 4–5 have regular and seasonal unit roots in the quarterly frequency. In this very frequent case, the corresponding annual model has two unit roots and a MA structure induced by aggregation⁴.
4. Aggregation induces additional MA structure and preserves the order of the stationary AR structure, typically reducing its persistence. Models 2 and 8 are clear examples of this. This result is consistent in the univariate case with Amemiya and Wu [5], Wei [220] and Stram and Wei [204]. In the multivariate stationary case, it is consistent with the findings in Chapters 4 and 6 of Lutkepohl [155] and in Marcellino [161].
5. Models 8–11 show that a feedback in the quarterly frequency is maintained in the annual frequency.
6. Model 6 is the annual model corresponding to a quarterly regression with AR(1) errors; see Chow–Lin [52]. Therefore, this method is empirically justified only if the model relating the target variable with the indicator in the annual frequency is a static regression with ARMA(1,1) errors.
7. Models 7 and 11 show that the aggregation algorithm is not restricted to VAR-MAX or transfer functions, as it can be applied to structural time series models, see Harvey [118], and VARMAX echelon models; see Hannan and Deistler [116]. In general, it supports any model with an equivalent SS representation.
8. Finally, model 10 shows how a monthly model aggregates to a quarterly VARMA, thus exemplifying that the aggregation algorithm can be applied to a general combination of sampling frequencies.

The code required to replicate these examples can be found in the file `Example1033.m`, see Appendix D.

10.3 Specification of the high-frequency model

Assume that a linear model has been fitted to all the available variables in the annual frequency. We will now discuss a systematic method to enforce consistency between this annual model and the unknown quarterly model, given the aggregation constraint and the partially aggregated sample. Without loss of generality, we will refer to the process of specifying a VARMA model, consisting of the successive determination of unit roots, AR and MA dynamics. The basic ideas can be easily applied to other model-building methods such as, e.g., structural time series modeling; see Harvey [118].

10.3.1 Enforcing approximate consistency

It is relatively easy to specify the structure of the quarterly model which realizes approximately a given annual model and, after estimating this high-frequency struc-

⁴This suggests that many annual variables should be I(2) while, in practice, annual models are often specified with a single unit root. This apparent contradiction is easy to explain because aggregation induces MA structures which could compensate an AR unit root if some MA root is close to unity.

Table 10.1 *Aggregation of several univariate models. The columns labelled "states" show the number of dynamic components in the minimal SS representation of the corresponding model. Therefore, the difference between the number of states in the quarterly and annual models is the number of dynamic components that become unobservable after aggregation. The quarterly variables z_t , z_{1t} and z_{2t} are assumed to be flows, so their annual aggregate is the sum of the corresponding quarterly values.*

#	Quarterly model	States	Annual model	States
1	$(1 - B^4)z_t = a_t; \sigma_a^2 = 1$	4	$(1 - B)z_t^A = a_t^A; \sigma_A^2 = 4.00$	1
2	$(1 - 0.8B)(1 + 0.8B)z_t = a_t; \sigma_a^2 = 1$	2	$(1 - 0.410B)z_t^A = (1 + 0.160B)a_t^A; \sigma_A^2 = 7.99$	1
3	$z_t = (1 - 0.6B^4)a_t; \sigma_a^2 = 1$	4	$z_t^A = (1 - 0.600B)a_t^A; \sigma_A^2 = 4.00$	1
4	$(1 - B)(1 - B^4)z_t = a_t; \sigma_a^2 = 1$	5	$(1 - B)^2 z_t^A = (1 + 0.240B)a_t^A; \sigma_A^2 = 41.60$	2
5	$(1 - B)(1 - B^4)z_t = (1 - 0.8B)(1 - 0.6B^4)a_t; \sigma_a^2 = 1$	5	$(1 - B)^2 z_t^A = (1 - 0.997B + 0.238B^2)a_t^A; \sigma_A^2 = 7.05$	2
6	$z_{1t} = 0.5z_{2t} + \frac{1}{1 - 0.8B}a_t; \sigma_a^2 = 1$	1	$z_{1t}^A = 0.500z_{2t}^A + \frac{1 + 0.228B}{1 - 0.410B}a_t^A; \sigma_A^2 = 23.103$	1
7	$(1 - B)T_{t+1} = u_t; \sigma_u^2 = 0.01$	4	$(1 - B)z_t^A = (1 - 0.669B)a_t^A; \sigma_A^2 = 5.844$	1
	$(1 + B + B^2 + B^3)S_{t+1} = v_t; \sigma_v^2 = 0.01$			
	$z_t = T_t + S_t + \varepsilon_t; \sigma_\varepsilon^2 = 1$			

Table 10.2 Aggregation of several bivariate models. All the quarterly variables z_{1t} and z_{2t} are assumed to be flows, so their annual aggregate is the sum of the corresponding quarterly values. If a variable were to be aggregated as an average of the quarterly values, the model would require an appropriate re-scaling.

#	High-frequency model	States	Low-frequency model	States
8	$\begin{bmatrix} 1 - 0.7B & -0.8B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (1 - B^4)z_{1t} \\ (1 - B^4)z_{2t} \end{bmatrix} = \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix};$	10	$\begin{bmatrix} 1 - 0.240B & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (1 - B)z_{1t}^A \\ (1 - B)z_{2t}^A \end{bmatrix} = \begin{bmatrix} 1 - 0.039B & 1.461 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{1t}^A \\ a_{2t}^A \end{bmatrix};$	3
9	$\Sigma_a = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} (1 - B^4)z_{1t} \\ (1 - B^4)z_{2t} \end{bmatrix} = \begin{bmatrix} 1 & -0.8B \\ 1.2B & 1 \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix};$	$\Sigma_A = \begin{bmatrix} 35.36 & 7.62 \\ 7.62 & 4.00 \end{bmatrix} \begin{bmatrix} (1 - B)z_{1t}^A \\ (1 - B)z_{2t}^A \end{bmatrix} = \begin{bmatrix} 1 - 0.080B & -0.053B \\ 0.289B & 1 + 0.015B \end{bmatrix} \begin{bmatrix} a_{1t}^A \\ a_{2t}^A \end{bmatrix};$	2	
10	$\Sigma_a = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} (1 - B^{12})z_{1t} \\ (1 - B^{12})z_{2t} \end{bmatrix} = \begin{bmatrix} 1 & -0.8B \\ 1.2B & 1 \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix};$	24	$\Sigma_A = \begin{bmatrix} 4.09 & 1.41 \\ 1.41 & 13.00 \end{bmatrix} \begin{bmatrix} (1 - B^4)z_{1t}^A \\ (1 - B^4)z_{2t}^A \end{bmatrix} = \begin{bmatrix} 1 - 0.100B & -0.075B \\ 0.365B & 1 + 0.024B \end{bmatrix} \begin{bmatrix} a_{1t}^A \\ a_{2t}^A \end{bmatrix};$	8
11	$\Sigma_a = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} 1 - 0.4B & 0.8B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{1t} \\ a_{2t} \end{bmatrix};$	1	$\Sigma_A = \begin{bmatrix} 1.03 & 9.27 \\ 3.22 & 1.03 \end{bmatrix} \begin{bmatrix} 1 - B & 0 \\ -0.5 & 1 \end{bmatrix} \begin{bmatrix} z_{1t}^A \\ z_{2t}^A \end{bmatrix} = \begin{bmatrix} 1 - 0.745B & 1.808B \\ 0 & 1 \end{bmatrix} \begin{bmatrix} a_{1t}^A \\ a_{2t}^A \end{bmatrix};$	1
	$\Sigma_a = \begin{bmatrix} 1 & 0.5 \\ 0.5 & 1 \end{bmatrix}$		$\Sigma_A = \begin{bmatrix} 51.91 & 29.55 \\ 29.55 & 19.58 \end{bmatrix}$	

ture, to assess whether the quarterly model obtained is statistically adequate or not. To these ends, one can use the following heuristic procedure.

Step (1) Annual modeling Specify and estimate a model relating the target annual variable(s) with the annualized values of the quarterly indicator(s). Any model having an equivalent linear SS representation, such as e.g., a transfer function or VARMAX, is adequate for this purpose.

Step (2) Decomposition of the quarterly indicator Adjust the quarterly indicator(s) to suppress undesired features, such as seasonality and calendar effects.

Step (3) Model specification Specify the VAR and VMA orders by the following steps:

Step (3.1) Set the VAR factor order of the quarterly model to be equal to that of the annual model and, particularly, constrain the number of unit roots to be the same. The foundation of this step results from comparing the quarterly model (10.4)–(10.5) and the annual model given by (10.4) and (10.20). As both models share the same state equation, they have the same (stationary and nonstationary) autoregressive components.

Step (3.2) Add a VMA(q) structure, with $q \leq n$, being n the size of the state vector in the annual model. This upper bound on the MA dynamics results from the fact that, in a minimal SS representation, the size of the state vector is $n = \max\{p, q\}m$, being p and q the orders of the VAR and VMA factors, respectively.

Step (4) Estimation Estimate the model specified in Steps (2) and (3) by maximum likelihood and prune insignificant parameters to obtain a parsimonious parameterization.

Step (5) Diagnostic checking Check the final quarterly model by obtaining the corresponding annual representation, using the algorithm described in the previous Section, and then:

Step (5.1) ... compare this model with the one specified in Step (1), and

Step (5.2) ... check whether it filters the annual data to white noise residuals.

Step (6) Check forecasting accuracy If the sample is long enough, compare out-of-the-sample forecasts for the annual values produced by both the tentative quarterly model and the annual model specified in Step (1).

Our experience in applying this method provides some clues about how to implement it:

- First, a good characterization of unit roots in Step (1) is critical, as misspecification of these components severely impacts the quality of the final results; see Tiao [207]. When in doubt, we think that over-differencing is safer than under-differencing.
- Second, the number of MA parameters specified in Step (3.2) may be excessive, depending on the sample size and number of time series. In this case, it is a good idea to constrain the MA matrices to be diagonal and, later, add off-diagonal parameters in a sequence of overfitting experiments.

- Third, as Nunes [169] points out, there are many situations where missing observations arise in a time series disaggregation framework. Some of these are: different release dates for some variables, changes in the sampling frequency or non-conformable samples, where the series considered have different starting dates. In these cases the ability of SS methods to take care of missing values is an important asset, as it allows for using all the information available.
- Last, the forecasting accuracy check proposed in Step (6) can be implemented by setting some within-sample values to missing, and estimating them afterwards. We have found this alternative useful when the sample is too short to reserve some values for out-of-sample forecasting.

10.3.2 “Bottom-up” determination of the quarterly model

On the other hand, one can adopt a “bottom-up” approach, similar to the one discussed in Subsection 9.3.4, to maximize the consistency between the high and low-frequency models. To this end, one must solve the equations relating the known annual model and the unknown quarterly model, using the algorithm described in Subsection 10.2.3. This can be done by using the following procedure.

Perform steps (1)–(4) of Subsection 10.3.1. The resulting outcomes will be: (a) an annual data model, whose parameters will be denoted by $\hat{\theta}_a$, and (b) a preliminary quarterly model, whose parameter vector will be denoted by $\hat{\Theta}_q^0$. Obtain the corresponding annual model parameters as $\hat{\theta}_a^0 = A(\hat{\Theta}_q^0)$, where the function $A(\cdot)$ denotes the aggregation procedure discussed in Subsection 10.2.3. Then execute:

Step (5) Iteration Estimate the model specified in Steps (2) and (3) above as

$$\hat{\Theta}^* = \arg \min \left[A(\hat{\Theta}_q) - \hat{\theta}_a \right]^T \left[A(\hat{\Theta}_q) - \hat{\theta}_a \right],$$

where the optimization must be performed iteratively because the problem is nonlinear. The estimates obtained in previous steps will be useful to initialize this iteration.

After convergence, a null value for the loss function implies that the quarterly and annual models are perfectly consistent, so the annual aggregation of the former realizes the latter exactly. On the other hand, a nonzero optimal loss means that it is not possible to achieve an exact match between the quarterly and annual models⁵. In this case it would be convenient to perform the diagnostic testing checks described in the previous subsection to formally validate the quarterly model.

10.4 Empirical example

We will now illustrate the application of the methods proposed in previous sections by disaggregating and forecasting the annual series of Value Added by the Industry in Spain (VAI), from 1980 to 2001, using as an indicator the quarterly values

⁵Bear in mind that, as stated in the discussion of Theorem 10.2, a statistically adequate model for the annual data may not have a perfectly consistent quarterly representation.

of a re-balanced Industrial Production Index (IND), from the 1st quarter of 1980 to the 4th quarter of 2001. The latter series is the indicator actually employed by the agency in charge of the Spanish national accounts.

To clarify when a series is expressed in annual or quarterly frequency we will use an uppercase/lowercase notation, so VAI_T^A and IND_T^A are the values of VAI and the annual average of the indicator in year T ($T = 1980, 1981, \dots, 2001$), while vai_t and ind_t denote the values of both variables in quarter t ($t = 1980.1, 1980.2, \dots, 2001.4$).

10.4.1 Annual model

The first step in the analysis consists of modeling the annual values of the target variable and the indicator. A standard analysis, see Jenkins and Alavi [131], resulted in the following model for $T = 1980, \dots, 2001$:

$$\begin{bmatrix} (1-B)^2 & 0 \\ 0 & (1-B)^2 \end{bmatrix} \begin{bmatrix} VAI_T^A \\ IND_T^A \end{bmatrix} = \begin{bmatrix} 1 - 1.85B & 0.74B \\ (0.06) & (0.02) \\ -1.48B & 1 \\ (0.05) & \end{bmatrix} \begin{bmatrix} \hat{a}_T^{VAI} \\ \hat{a}_T^{IND} \end{bmatrix}, \quad (10.25)$$

$$\hat{\Sigma}_a = \begin{bmatrix} 2.64 & 5.53 \\ 5.53 & 13.08 \end{bmatrix}, \quad Q(5) = \begin{bmatrix} 4.84 & 4.31 \\ 6.23 & 4.17 \end{bmatrix},$$

where the values in parentheses are the standard errors of the estimates, $\hat{\Sigma}_a$ represents the estimate of the error covariance matrix, and $Q(5)$ is the matrix of portmanteau statistics.

Model (10.25) provides precise clues about the dynamics of VAI and IND in the quarterly frequency. Following the guidelines in Subsection 10.3.1 a consistent quarterly model must have a nonstationary AR structure with two unit roots for each series and an MA term with a maximum order of 4, because the minimal SS representation of (10.25) requires four state variables.

The code required to replicate these results can be found in the file Example104Step1.m, see Appendix D.

10.4.2 Decomposition of the quarterly indicator

The second step requires modeling the quarterly indicator to extract the components useful for disaggregation. To this end, we will use the following model for $t = 1980.1, 1980.2, \dots, 2001.4$:

$$\begin{aligned} ind_t &= 1.02L_t - 1.85E_t - 5.94S_t^{92.4} + \hat{N}_t, \\ (1-B)(1-B^4)\hat{N}_t &= (1 - 0.83B^4)\hat{a}_t, \\ \hat{\sigma}_a^2 &= 3.21, \quad Q(15) = 10.90, \end{aligned} \quad (10.26)$$

where L_t is the number of non-holidays in quarter t , E_t is a dummy variable to account for Easter effects, and $S_t^{92.4}$ is an intervention variable capturing a persistent change from the 4th quarter of 1992 onwards.

According to the signal extraction ideas explained in Chapter 8, model (10.26) implies that the indicator can be decomposed into: (a) trend, (b) seasonal component, (c) irregular component, (d) calendar effects associated to the number of non-holidays and Easter; and (e) the 1992 4th quarter effect. Figure 10.1 shows these components. In the light of previous results, components (b) and (d) are useless for disaggregation. By adding the remaining components we obtain a seasonally and calendar-adjusted quarterly indicator series.

The code required to replicate these results can be found in the file Example104Step2.m, see Appendix D.

10.4.3 Specification and estimation of the quarterly model

The third and fourth steps consist of specifying and estimating a model for the quarterly data. Building on the results from the previous steps, we estimate a doubly integrated VMA(4) process for the value added and the adjusted indicator in the quarterly frequency. After pruning insignificant parameters we obtained the following model for $t = 1980.1, 1980.2, \dots, 2001.4$:

$$\begin{bmatrix} (1-B)^2 & 0 \\ 0 & (1-B)^2 \end{bmatrix} \begin{bmatrix} vai_t \\ \widetilde{ind}_t \end{bmatrix} = \begin{bmatrix} 1 - 1.48B + 0.36B^2 & 0.044B \\ 0.044B & 1 - 0.78B \end{bmatrix} \begin{bmatrix} \hat{a}_t^{vai} \\ \hat{a}_t^{\widetilde{ind}} \end{bmatrix},$$

$$\hat{\Sigma}_a = \begin{bmatrix} 0.06 & 0.40 \\ 0.40 & 3.64 \end{bmatrix}, \quad (10.27)$$

where vai_t denotes the unobserved value of VAI in quarter t and \widetilde{ind}_t is the corresponding adjusted indicator. Finally, applying to the sample the fixed-interval smoothing described in Chapter 4, we obtain the disaggregates and forecasts shown in Figure 10.2.

The code required to replicate these results can be found in the file Example104Step3and4.m, see Appendix D.

10.4.4 Diagnostics

The fifth step of the analysis consists of assessing whether the quarterly model is empirically consistent with the annual data. To this end, we obtain the annual representation corresponding to model (10.27) by means of the aggregation algorithm described in Section 10.3.1. It is:

$$\begin{bmatrix} (1-B)^2 & 0 \\ 0 & (1-B)^2 \end{bmatrix} \begin{bmatrix} VAI_T^A \\ \widetilde{IND}_T^A \end{bmatrix} = \begin{bmatrix} 1 - 1.52B - .23B^2 & .70B + .10B^2 \\ -1.18B - 0.19B^2 & 1 + .14B - .07B^2 \end{bmatrix} \begin{bmatrix} \hat{a}_T^{VAI} \\ \hat{a}_T^{\widetilde{IND}} \end{bmatrix},$$

$$\hat{\Sigma}_a = \begin{bmatrix} 2.54 & 5.53 \\ 5.53 & 13.79 \end{bmatrix}, \quad Q(5) = \begin{bmatrix} 7.38 & 7.81 \\ 9.49 & 8.48 \end{bmatrix}, \quad (10.28)$$

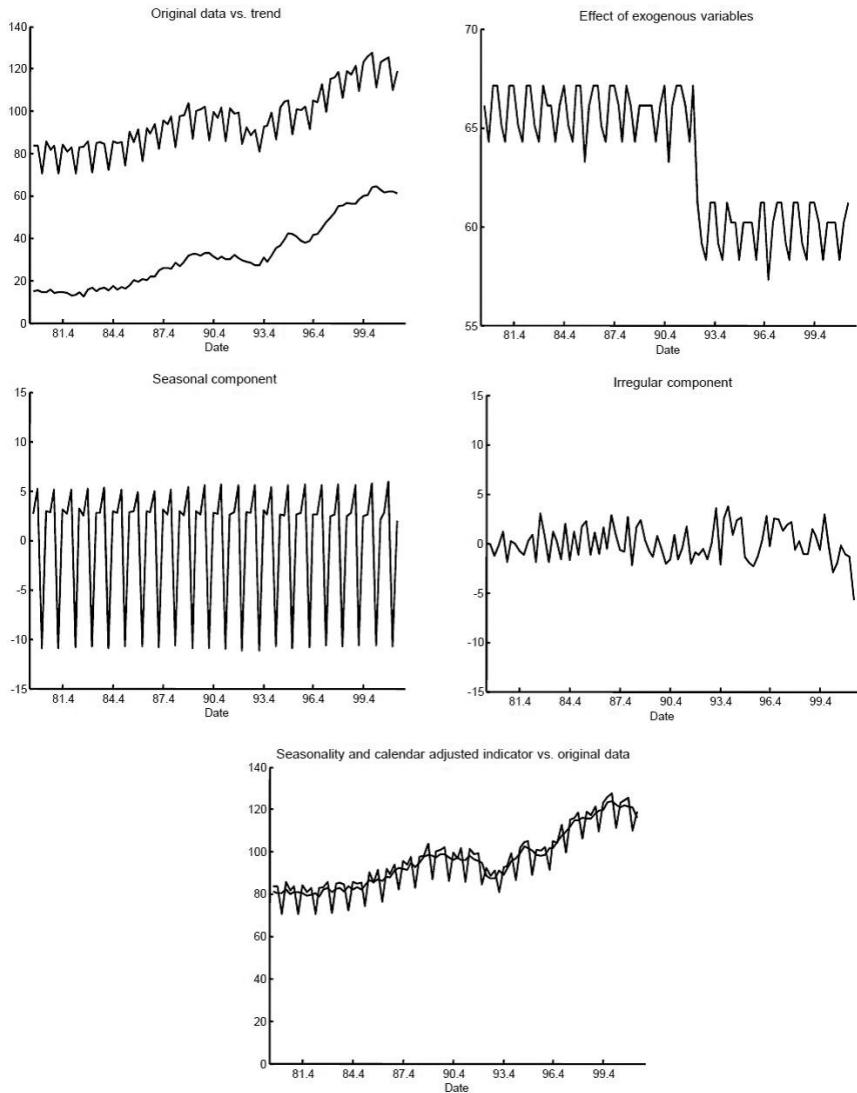


Figure 10.1 *Decomposition of the quarterly indicator. The adjusted indicator includes the trend, irregular, and level change components, and excludes seasonality and calendar affects.*

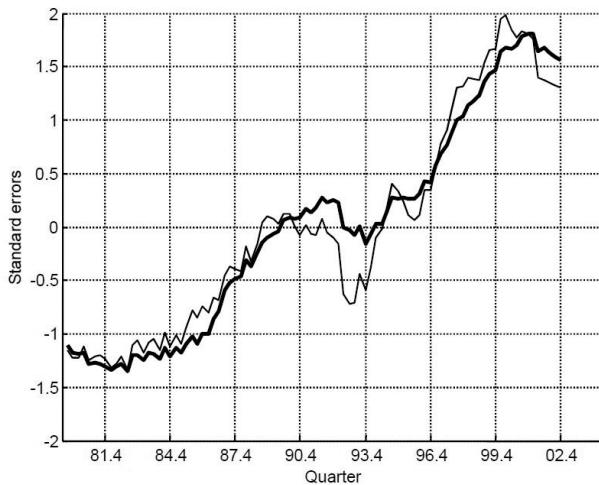


Figure 10.2 *Standardized plot of quarterly estimates of VAI (thick line) and adjusted indicator (thin line). The last values are forecasts computed from 2002.1 to 2002.4 quarters using model (10.27).*

where \widetilde{IND}_T^A denotes the annual average of the adjusted indicator in year T . Models (10.28) and (10.25) differ mainly in the additional second-order MA parameters in (10.28)⁶. On the other hand, the residuals obtained by filtering the annual series using (10.28) are stationary, normal, and do not show important autocorrelations. Therefore, we conclude that model (10.27) is statistically adequate and roughly conformable with (10.25).

Previous results in this example show that our method can be applied to real disaggregation problems. The remaining subsections highlight its advantages when dealing with non-conformable samples and in terms of forecasting power.

The code required to replicate these results can be found in the file `Example104Step5.m`, see Appendix D.

10.4.5 Forecast accuracy and non-conformable samples

Statistical bureaus typically have long records for the target variable, say GDP, and shorter ones for the indicators. In this situation, standard techniques constrain the analysis to the common time window, thus assuming a substantial information loss. However the SS methods employed here allow for missing values, so all the available information can be used. To illustrate this idea, we delete the first four observations of the quarterly indicator and the last annual value of VAI. Re-estimating model (10.27)

⁶We tried to fit an annual model including these parameters and the corresponding estimates resulted insignificant. Perhaps this additional MA structure is due to the disaggregated indicator information included in (10.27) and excluded in (10.25)

with this non-conformable sample yields the following results:

$$\begin{aligned} \begin{bmatrix} (1-B)^2 & 0 \\ 0 & (1-B)^2 \end{bmatrix} \begin{bmatrix} \widehat{vai}_t \\ \widehat{ind}_t \end{bmatrix} &= \begin{bmatrix} 1 - 1.48B + 0.34B^2 & 0.045B \\ (0.06) & (0.05) \\ -0.89B & 1 - 0.78B \\ (0.04) & (0.03) \end{bmatrix} \begin{bmatrix} \widehat{a}_t^{vai} \\ \widehat{a}_t^{ind} \end{bmatrix}, \\ \hat{\Sigma}_a &= \begin{bmatrix} 0.06 & 0.41 \\ 0.41 & 3.74 \end{bmatrix}. \end{aligned} \tag{10.29}$$

Estimates in (10.29) are remarkably close to those in (10.27). Table 10.3 shows the original sample information and the numerical results obtained with models (10.27) and (10.29). Note that: (a) the forecast provided by (10.29) for the 2001 annual value is very accurate, (b) quarterly interpolations and out-of-sample forecasts obtained with both models are very similar, and (c) the indicator extrapolations computed with (10.29) and the non-conformable sample have acceptable errors.

10.4.6 Comparison with alternative methods

Table 10.4 shows the root mean squared errors (RMSEs) obtained by forecasting the last five end-of-year values of VAI with Model (10.27) and some alternative methods. All the forecasts are conditional on the true indicator values⁷. Note that Model (10.27) produced the best forecasts, with an 8–9% advantage over the second-best, and much larger gains in comparison with the remaining methods. Obviously, this comparison is not perfectly fair, as Model (10.27) was carefully fitted to the data while the alternative procedures are mechanical. Therefore, these results seem to reinforce the idea that a model fitted by a trained human is usually able to forecast better than an automatic system.

The code required to replicate these results can be found in the files `Example104Step6.m` and `Example104Step6-Chow-Lin.m`, see Appendix D.

⁷For example, to compute the 1997 values each model was estimated using annual VAI and quarterly indicator values up to 1996, and was then used to predict the end-of-year value of VAI using: (a) the past of VAI up to 1996, and (b) past values of the indicator up to 1997. The same procedure was applied for 1998, 1999, 2000, and 2001 extending the sample in each case with one, two, three, and four additional years of data.

Table 10.3 *Original data and interpolations obtained with full and non-conformable samples. Underlined values correspond to interpolations, retrodictions or forecasts. The column "Indicator" refers to the calendar and seasonally adjusted values computed with model (10.26).*

Obs.	Original sample		Full sample			Non-conformable sample		
	Annual	Indicator	Interpolation	Annual	Indicator	Interpolation	Annual	Indicator
VAI	sum	sum	sum	sum	sum	sum	sum	sum
1980.1	81.421	<u>14.993</u>	81.421	<u>15.009</u>	82.869			
1980.2	80.451	<u>14.810</u>	80.451	<u>14.896</u>	82.237			
1980.3	80.524	<u>14.744</u>	80.524	<u>14.778</u>	81.584			
1980.4	59.341	81.877	<u>14.794</u>	59.341	81.877	<u>14.658</u>	59.341	80.968
1981.1		80.196	<u>14.474</u>		80.196	<u>14.502</u>		80.196
1981.2		80.710	<u>14.531</u>		80.710	<u>14.519</u>		80.710
1981.3		80.823	<u>14.483</u>		80.823	<u>14.474</u>		80.823
1981.4	57.895	80.299	<u>14.407</u>	57.895	80.299	<u>14.400</u>	57.895	80.299
...								
2000.1	123.520	<u>23.326</u>	123.520	<u>23.342</u>	123.520	<u>23.342</u>	123.520	123.520
2000.2	124.005	<u>23.417</u>	124.005	<u>23.431</u>	124.005	<u>23.431</u>	124.005	124.005
2000.3	122.033	<u>23.374</u>	122.033	<u>23.371</u>	122.033	<u>23.371</u>	122.033	122.033
2000.4	93.620	121.147	<u>23.502</u>	93.620	121.147	<u>23.476</u>	93.620	121.147
2001.1		121.823	<u>23.753</u>		121.823	<u>23.690</u>		121.823
2001.2		121.550	<u>23.801</u>		121.550	<u>23.718</u>		121.550
2001.3		121.006	<u>23.830</u>		121.006	<u>23.734</u>		121.006
2001.4	94.711	115.986	<u>23.327</u>	94.711	115.986	<u>23.217</u>	94.359	115.986
2002.1			<u>23.414</u>		115.664	<u>23.311</u>		115.723
2002.2			<u>23.295</u>		115.343	<u>23.199</u>		115.461
2002.3			<u>23.177</u>		115.021	<u>23.088</u>		115.198
2002.4			<u>23.059</u>		114.699	<u>22.976</u>	92.574	114.936

Table 10.4 *Ranking of RMSEs for end-of-year forecasts of annual VAI from 1997 to 2001. RMSEs are computed for forecasts of both, VAI levels and growth rates. The columns RMSE% show the corresponding RMSEs normalized so that the RMSE of Model (10.29) is 100%*

Rig	Model	Levels		Annual growth rates	
		RMSE	RMSE %	RMSE	RMSE %
1	Model (10.29)	0.65	100.0	0.71	100.0
2	Litterman [149] with constant	0.72	109.6	0.77	108.4
3	Fernandez [95] with constant	0.96	147.5	1.03	145.2
4	Chow-Lin [52] AR(1) with constant	1.21	185.2	1.3	182.5
5	Fernandez [95] without constant	1.44	219.8	1.54	216.0
6	Litterman [149] without constant	1.47	224.9	1.57	220.9
7	ADL(1,0) ^a without constant	1.53	234.0	1.73	242.4
8	ADL(1,0) with constant	1.61	247.0	1.75	245.3
9	Chow-Lin [52] AR(1) without constant	1.82	278.9	1.98	277.2
10	Boot, Feibes and Lisman [26]	3.04	466.2	3.39	475.2
11	Denton [75]	18.01	2758.4	20.26	2842.3

^aThe ADL (Autoregressive Distributed Lag model) of orders p and q , denoted by $\text{ADL}(p,q)$, is defined by the equation: $y_t = \beta_0 + \beta_1 y_{t-1} + \beta_2 y_{t-2} + \dots + \beta_p y_{t-p} + \delta_0 x_t + \delta_1 x_{t-1} + \delta_2 x_{t-2} + \dots + \delta_q x_{t-q} + \varepsilon_t$

This page intentionally left blank

Chapter 11

Cross-sectional extension: longitudinal and panel data

11.1 Model formulation	168
11.2 The Kalman Filter	169
11.2.1 Case of uncorrelated state and observational errors	170
11.2.2 Case of correlated state and observational errors	171
11.3 The linear mixed model in SS form	171
11.4 Maximum likelihood estimation	174
11.5 Missing data modifications	176
11.5.1 Missingness in responses only	176
11.5.2 Missingness in both responses and covariates: method 1	178
11.5.3 Missingness in both responses and covariates: method 2	180
11.6 Real data examples	182
11.6.1 A LMM for the mare ovarian follicles data	182
11.6.2 Smoothing and prediction of missing values for the beluga whales data	183

Multiple time series on independent *units* occur routinely in biostatistics and econometrics, among other disciplines. In the classical longitudinal study, multivariate quantitative outcomes may be collected over time on multiple *subjects* (living organisms) and their contemporaneously observed explanatory variables or *covariates*. Cause and effect relationships among the variables are typically of interest here. Similarly, econometrics panel data is comprised of multivariate quantitative and/or qualitative responses on several *entities* (countries, firms, individuals, etc.), frequently collected with the intent of using any available explanatory variables (now called *exogenous variables*) to aid in making future predictions.

This chapter discusses the cross-sectional extension that is required in the classical SS formulation of Chapter 2, in order to accommodate these types of data. The common linear mixed model for longitudinal analyses will be seen to be expressible in SS form. Appropriate Kalman smoothing, filtering, and prediction recursions

in this context will be discussed, as well as ML estimation¹. A particularly attractive and powerful feature of this methodology is its ability to account for missing or irregularly-spaced time series data, together with the ease with which smoothing and forecasting is accomplished. Since this “missingness” is often present in longitudinal analyses, a substantial portion of this chapter is devoted to techniques for accommodating such features. The methods will be illustrated on some real datasets.

Finally, we note that, since the cross-sectional extension is essentially a “replication” of the SS model across independent units, all the methods discussed here apply equally well to the situation of a single unit that has been the subject of all the previous chapters. For this reason (but also in order to follow standard longitudinal presentation of the material), we have strived to provide a somewhat different slant on things. However, some of the sections, and in particular the ML estimation and missing data adjustments, discuss approaches that have not been presented in earlier chapters.

11.1 Model formulation

The modeling of longitudinal or panel type data can be accommodated by a *cross-sectional* extension to the standard SS model. It necessitates the addition of an extra subscript so that $z_{i,t}$ and $x_{i,t}$ now represent the observation and state variables for the i -th unit, $i = 1, \dots, k$, at time t , $t = 1, \dots, n_i$. Using similar notation as in Chapter 2, the state equation for a simplified MEM becomes

$$\mathbf{x}_{i,t} = \Phi \mathbf{x}_{i,t-1} + \Gamma \mathbf{u}_{i,t-1} + \mathbf{w}_{i,t}, \quad (11.1)$$

where $\mathbf{w}_{i,t} \sim NID(\mathbf{0}, \mathbf{Q})$, with initial condition $\mathbf{x}_{i,0} \sim N(\boldsymbol{\mu}, \mathbf{P})$. An alternative formulation that admits a contemporaneous (instead of lagged) relationship between the state and exogenous variables can also be used,

$$\mathbf{x}_{i,t} = \Phi \mathbf{x}_{i,t-1} + \Upsilon \mathbf{u}_{i,t} + \mathbf{w}_{i,t}. \quad (11.2)$$

The corresponding observation equation for either of these is

$$\mathbf{z}_{i,t} = \mathbf{H}_{i,t} \mathbf{x}_{i,t} + \mathbf{D} \mathbf{u}_{i,t} + \mathbf{v}_{i,t}, \quad (11.3)$$

where $\mathbf{v}_{i,t} \sim NID(\mathbf{0}, \mathbf{R})$ represents the observational noise vector. For simplicity in likelihood computations, $\{\mathbf{w}_{i,t}\}$ and $\{\mathbf{v}_{i,t}\}$ are assumed to be uncorrelated, but a more general setup is to allow for

$$\text{cov}(\mathbf{w}_{i,s}, \mathbf{v}_{i,t}) = \begin{cases} \mathbf{S}, & s = t + 1, \\ \mathbf{0}, & s \neq t + 1. \end{cases} \quad (11.4)$$

The notation $\mathbf{H}_{i,t}$ in (11.3) reflects the fact that the coefficient matrix \mathbf{H} , from the formulation in Chapter 6, is now allowed to vary over both unit i and time t . This

¹Quantification of the variability in ML estimates via asymptotic results is discussed in Appendix B.

is an important extension that allows for the accommodation of missing data; to be discussed shortly.

A similar class of models popular for handling longitudinal data in the social and behavioral sciences, are the so-called *structural equation models*, see Bollen [24]. In these fields, the essential problem is the estimation of structural relationships between quantitative observed variables, and such models have become a prominent form of data analysis only in the last twenty years or so (thanks in part to the availability of the software LISREL). Although SS models were originally developed from a time series perspective (long series on one unit), and structural equation models from a panel data perspective (short series on multiple units), there has been a gradual convergence of the two, see Oud and Jansen [172], with Chow *et al.* [53] pointing out equivalences and complementary aspects. The recent monograph by Little [150] is a state-of-the-art authoritative guide to applied longitudinal structural equation modeling.

An early version of the cross-sectional extension to the basic SS model seems to have been first proposed by Goodrich and Caines [112], who also investigated maximum (Gaussian) likelihood estimation and derived the corresponding asymptotic properties in a series of three papers, the results of which appear in Caines [37, ch. 9], Jones [133], Fong [98], Shumway and Stoffer [195, ch. 6], and Hsiao [127], comprise the short list of subsequent attempts at popularizing the use of SS models for longitudinal data analysis. Fahrmeir and Tutz [90] report that the large number of parameters involved in multivariate modeling can make this a difficult endeavor.

In the terminology of Diggle *et al.* [80], the SS approach is categorized as a *transition* (or Markov) model, wherein the response vector $\mathbf{z}_{i,t}$ is affected by earlier values of itself as well as the exogenous $\mathbf{u}_{i,t}$, in the fashion of a classical VAR-MAX model. Two other (currently more popular) generalized linear model based approaches include *marginal* models and *random effects* models. The former models the marginal low-order moments and cross moments of $\mathbf{z}_{i,t}$ directly in a parametrized functional form, whereas correlation in the responses over time is introduced less directly in the latter by adding random effects to a linear model structure for the systematic component that is then related to the mean via the link function.

11.2 The Kalman Filter

The KF enables one to *optimally* (mean-squared error sense) smooth, filter, and forecast the series. Before presenting a version that applies to the situation of multiple subjects, define the following notation. Let $Z_{i,s} = (\mathbf{z}_{i,1}, \dots, \mathbf{z}_{i,s})$ and $U_{i,s} = (\mathbf{u}_{i,1}, \dots, \mathbf{u}_{i,s})$ denote the responses and covariates observed up to time s for subject i . Throughout we will assume the covariates to be *exogenous*, implying conditional independence of all preceding values and of all responses up to the present time. However, the results established here will depend only on the weaker implied condition,

$$\mathbb{E}(U_{i,t} | Z_{i,t}, U_{i,t-1}) = \mathbb{E}(U_{i,t} | U_{i,t-1}).$$

Similar to Shumway and Stoffer [195, ch. 6], we adopt the following notation for the mean of the state vector at time t given the responses and covariates up to time s , and

its error covariance matrix at times t_1 and t_2 ,

$$\mathbf{x}_{i,t}^s = \mathbb{E}(\mathbf{x}_{i,t} | Z_{i,s}, U_{i,s}), \quad (11.5)$$

$$\mathbf{P}_{i,t_1,t_2}^s = \mathbb{E}\left[\left(\mathbf{x}_{i,t_1} - \mathbf{x}_{i,t_1}^s\right) \left(\mathbf{x}_{i,t_2} - \mathbf{x}_{i,t_2}^s\right)^T \middle| Z_{i,s}, U_{i,s}\right]. \quad (11.6)$$

When $t_1 = t_2 = t$, we will write simply $\mathbf{P}_{i,t}^s$ for the conditional variance in (11.6). With this notation defining the information set upon which the conditioning is to be performed, the appropriate Kalman filtering relations for each subject i are then identical to those already presented in Chapter 4. In what follows, we cast these in a slightly different form, and adapted specifically for the SS model defined by (11.2)–(11.3).

Remark 11.1 *The optimality conferred by Kalman relations is in the minimum mean-squared error sense, within the class of all linear predicted/filtered/smoothed values. In this case, the conditional expectation operator in (11.5) can more generally be interpreted as the projection operator of $\mathbf{x}_{i,t}$ on the linear span of $Z_{i,s}$ and $U_{i,s}$, with $\mathbf{P}_{i,t}^s$ as the corresponding mean-squared error.*

11.2.1 Case of uncorrelated state and observational errors

In the special case of $\mathbf{S} = \mathbf{0}$ in (11.4), the Kalman smoothing ($t > s$) and Kalman filtering ($t = s$) relations are given respectively by the recursions

$$\mathbf{x}_{i,t}^{t-1} = \Phi \mathbf{x}_{i,t-1}^{t-1} + \Upsilon \mathbf{u}_{i,t}, \quad (11.7)$$

$$\mathbf{P}_{i,t}^{t-1} = \Phi \mathbf{P}_{i,t-1}^{t-1} \Phi^T + \mathbf{Q}, \quad (11.8)$$

and

$$\mathbf{x}_{i,t}^t = \mathbf{x}_{i,t}^{t-1} + \mathbf{K}_{i,t} \left(\mathbf{z}_{i,t} - \mathbf{H}_{i,t} \mathbf{x}_{i,t}^{t-1} - \mathbf{D} \mathbf{u}_{i,t} \right), \quad (11.9)$$

$$\mathbf{P}_{i,t}^t = (I - \mathbf{K}_{i,t} \mathbf{H}_{i,t}) \mathbf{P}_{i,t}^{t-1}, \quad (11.10)$$

$$\mathbf{K}_{i,t} = \mathbf{P}_{i,t}^{t-1} \mathbf{H}_{i,t}^T \left(\mathbf{H}_{i,t} \mathbf{P}_{i,t}^{t-1} \mathbf{H}_{i,t}^T + \mathbf{R} \right)^{-1}, \quad (11.11)$$

for $i = 1, \dots, k$ and $t = 1, \dots, n_i$, with initial conditions $\mathbf{x}_{i,0}^0 = \boldsymbol{\mu}$ and $\mathbf{P}_{i,0}^0 = \mathbf{P}$.

If desired, h -step ahead prediction, $h \geq 1$, of a particular subject beyond the range of observed values $t = 1, \dots, n_i := n$, can then be accomplished with initial conditions $\mathbf{x}_{i,n}^n$ and $\mathbf{P}_{i,n}^n$, assuming known values for the future exogenous $\{\mathbf{u}_{i,n+h}\}$. Additional equations can also be introduced if prediction of the observed process $\{\mathbf{z}_{i,t}\}$ is of interest. In this case, letting

$$\mathbf{z}_{i,n+h}^n = \mathbb{E}(\mathbf{z}_{i,n+h} | Z_{i,n}, U_{i,n}),$$

and

$$\mathbf{B}_{i,n+h}^n = \mathbb{E}\left[\left(\mathbf{z}_{i,n+h} - \mathbf{z}_{i,n+h}^n\right) \left(\mathbf{z}_{i,n+h} - \mathbf{z}_{i,n+h}^n\right)^T \middle| Z_{i,n}, U_{i,n}\right],$$

denote respectively the prediction and prediction mean squared error of $\mathbf{z}_{i,n+h}$, the appropriate Kalman recursions for h -step ahead prediction are then

$$\mathbf{x}_{i,n+h}^n = \Phi \mathbf{x}_{i,n+h-1}^n + \Upsilon \mathbf{u}_{i,n+h}, \quad (11.12)$$

$$\mathbf{z}_{i,n+h}^n = \mathbf{H}_{i,n+h} \mathbf{x}_{i,n+h}^n + \mathbf{D} \mathbf{u}_{i,n+h}, \quad (11.13)$$

$$\mathbf{P}_{i,n+h}^n = \Phi \mathbf{P}_{i,n+h-1}^n \Phi^T + \mathbf{Q}, \quad (11.14)$$

$$\mathbf{B}_{i,n+h}^n = \mathbf{H}_{i,n+h} \mathbf{P}_{i,n+h}^n \mathbf{H}_{i,n+h}^T + \mathbf{R}. \quad (11.15)$$

11.2.2 Case of correlated state and observational errors

In the most general case of $\mathbf{S} \neq \mathbf{0}$ in (11.4), we express the Kalman smoothing recursions in prediction error form very similar to those of Section 4.2 as

$$\mathbf{x}_{i,t+1}^t = \Phi \mathbf{x}_{i,t}^{t-1} + \Upsilon \mathbf{u}_{i,t+1} + \mathbf{K}_{i,t} \tilde{\mathbf{z}}_{i,t}, \quad (11.16)$$

$$\tilde{\mathbf{z}}_{i,t} = \mathbf{z}_{i,t} - \mathbf{H}_{i,t} \mathbf{x}_{i,t}^{t-1} - \mathbf{D} \mathbf{u}_{i,t}, \quad (11.17)$$

$$\mathbf{P}_{i,t+1}^t = \Phi \mathbf{P}_{i,t}^{t-1} \Phi^T + \mathbf{Q} - \mathbf{K}_{i,t} \mathbf{B}_{i,t} \mathbf{K}_{i,t}^T, \quad (11.18)$$

$$\mathbf{K}_{i,t} = (\Phi \mathbf{P}_{i,t}^{t-1} \mathbf{H}_{i,t}^T + \mathbf{S}) \mathbf{B}_{i,t}^{-1}, \quad (11.19)$$

$$\mathbf{B}_{i,t} = \mathbf{H}_{i,t} \mathbf{P}_{i,t}^{t-1} \mathbf{H}_{i,t}^T + \mathbf{R}, \quad (11.20)$$

and the Kalman filtering recursions

$$\mathbf{x}_{i,t}^t = \mathbf{x}_{i,t}^{t-1} + \mathbf{P}_{i,t}^{t-1} \mathbf{H}_{i,t}^T \mathbf{B}_{i,t}^{-1} \tilde{\mathbf{z}}_{i,t}, \quad (11.21)$$

$$\mathbf{P}_{i,t}^t = \mathbf{P}_{i,t}^{t-1} - \mathbf{P}_{i,t}^{t-1} \mathbf{H}_{i,t}^T \mathbf{B}_{i,t}^{-1} \mathbf{H}_{i,t} \mathbf{P}_{i,t}^{t-1}, \quad (11.22)$$

$$(11.23)$$

for $i = 1, \dots, k$ and $t = 1, \dots, n_i$, with initial conditions $\mathbf{x}_{i,1}^0 = \Phi \boldsymbol{\mu} + \Upsilon \mathbf{u}_{i,1}$ and $\mathbf{P}_{i,1}^0 = \Phi \mathbf{P} \Phi^T + \mathbf{Q}$. Prediction h -steps ahead beyond the range of observed values is identical to (11.12)–(11.15).

Result 11.1 This discussion assumes a time-invariant state coefficient matrix, i.e. $\Phi_t \equiv \Phi$ for all t , but all the results in this Section 11.2 hold with the obvious modification of $\Phi = \Phi_t$ for step t in the case of time-varying state coefficient matrices.

11.3 The linear mixed model in SS form

The *linear mixed model* (LMM) is perhaps the most frequently employed type of random effects model in longitudinal data. This is sometimes also called the Laird-Ware model after Laird and Ware [146] who first formulated it, and it has a SS representation. The classical LMM in its most general compact matrix form is given

by

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{Z}\mathbf{b} + \boldsymbol{\varepsilon}, \quad \text{E}\left(\begin{bmatrix} \mathbf{b} \\ \boldsymbol{\varepsilon} \end{bmatrix}\right) = \begin{bmatrix} \mathbf{0} \\ \mathbf{0} \end{bmatrix}, \quad \text{cov}\left(\begin{bmatrix} \mathbf{b} \\ \boldsymbol{\varepsilon} \end{bmatrix}\right) = \begin{bmatrix} \mathbf{Q} & \mathbf{0} \\ \mathbf{0} & \mathbf{R} \end{bmatrix}, \quad (11.24)$$

which implies the following mean and covariance structure for the response vector, $\mathbf{y} \sim (\mathbf{X}\beta, \mathbf{Z}\mathbf{Q}\mathbf{Z}^T + \mathbf{R})$. The \mathbf{X} and \mathbf{Z} are known design matrices corresponding to the *fixed* and *random* effects, respectively, being entertained. The former are viewed as exogenous deterministic inputs into the system, which therefore makes their accommodation by the corresponding exogenous part of (11.3) the obvious choice in a SS framework. Time does not necessarily enter into (11.24); this basic model can accommodate either a strictly cross-sectional or strictly longitudinal setup, as well as both.

In a longitudinal version of the above LMM, the vector of observations for subject i , $\mathbf{y}_i = [y_{i,1}, \dots, y_{i,n_i}]^T$, $i = 1, \dots, k$, aggregated across times $t = 1, \dots, n_i$, is given as

$$\mathbf{y}_i = \mathbf{X}_i\beta + \mathbf{Z}_i\mathbf{b}_i + \boldsymbol{\varepsilon}_i, \quad \mathbf{b}_i \sim \text{NID}(\mathbf{0}, \mathbf{G}), \quad \boldsymbol{\varepsilon}_i = [\boldsymbol{\varepsilon}_{i,1}, \dots, \boldsymbol{\varepsilon}_{i,n_i}]^T \sim \text{NID}(\mathbf{0}, \Sigma_i), \quad (11.25)$$

so that $\mathbf{y}_i \sim \text{N}(\mathbf{X}_i\beta, \mathbf{Z}_i\mathbf{G}\mathbf{Z}_i^T + \Sigma_i)$. The assumptions of normality on the random effects \mathbf{b}_i (between-subjects) and $\boldsymbol{\varepsilon}_i$ (within-subjects), are for convenience in likelihood based inference. Decomposed further into subject i at time t , (11.25) becomes

$$y_{i,t} = \mathbf{x}_{i,t}^T \beta + \mathbf{z}_{i,t}^T \mathbf{b}_i + \varepsilon_{i,t}, \quad (11.26)$$

where $\mathbf{x}_{i,t}$ and $\mathbf{z}_{i,t}$ are n_i vectors containing the t -th rows of \mathbf{X}_i and \mathbf{Z}_i , respectively.

One way to cast (11.26) in SS form is to place all random components into a large state vector. We can do this in a two-step process. Consider first the case of (11.26) with uncorrelated within-subject errors. Now turn $\mathbf{b}_i \equiv \mathbf{b}_{i,t}$ into a state vector by introducing (vacuous) time dependence with initial condition, $\mathbf{b}_{i,0} \sim \text{N}(\mathbf{0}, \mathbf{G})$. The observation and state equations then become

$$y_{i,t} = \mathbf{z}_{i,t}^T \mathbf{b}_{i,t} + \mathbf{x}_{i,t}^T \beta + \varepsilon_{i,t}, \quad \varepsilon_{i,t} \sim \text{NID}(0, \sigma^2) \quad (11.27)$$

$$\mathbf{b}_{i,t} = \mathbf{b}_{i,t-1}, \quad (11.28)$$

where we note that the $\mathbf{x}_{i,t}^T \beta$ term naturally plays the part of the exogenous $\mathbf{D}\mathbf{u}_{i,t}$ term in (11.3). Note also that the $w_{i,t}$ noise term is absent from (11.28) since it has null variance, this situation being similar to the Recursive Least Squares example in Subsection 4.7.1.

In longitudinal settings however, it will rarely be the case that the $\{\varepsilon_{i,t}\}$ will be serially uncorrelated, and thus appropriate specification of the within-subject error covariance matrix Σ_i is key. Letting $\sigma_i(t, t+h)$ denote the row t and column $t+h$ entry of Σ_i , note that $\text{E}(y_{i,t}) = \mathbf{x}_{i,t}^T \beta$, with covariance

$$\text{cov}(y_{i,t}, y_{j,t+h}) = \begin{cases} \mathbf{z}_{i,t}^T \mathbf{G} \mathbf{z}_{i,t+h} + \sigma_i(t, t+h), & \text{if } i = j, \\ 0, & \text{if } i \neq j. \end{cases}$$

With the hope that any nonstationary behavior has been accounted for by the exogenous term and the random effects $\{z_{i,t}^T b_{i,t}\}$, a stationary within-subject error covariance structure, such as that modeled by an ARMA, is frequently assumed for $\sigma_i(t, t+h) \equiv \sigma_i(h)$. The second step of the two-step process to cast (11.26) in SS form then renders the ARMA model for $\{\varepsilon_{i,t}\}$ in SS form, in the absence of random effects. Finally, these two steps are combined by assembling the separate state vectors into a larger state vector.

The following example illustrates some of the ideas discussed above in converting an LMM into SS form. It focuses on demonstrating the various ways that can potentially be used to effect this conversion, and the trade-offs involved with each. A second example is given in Subsection 11.6.1, and it demonstrates an application of the standard two-step procedure to a real dataset.

Example 11.3.1 A mean with AR(1) errors

Suppose the vector of observations for subject i , $i = 1, \dots, k$, is of length n and follows a constant mean β with AR(1) errors,

$$\mathbf{y}_i = \mathbf{1}\beta + \boldsymbol{\varepsilon}_i, \quad \mathbf{1} = [1, \dots, 1]^T, \quad \boldsymbol{\varepsilon}_i \sim \text{NID}(\mathbf{0}, \Sigma_i),$$

where Σ_i is an $n \times n$ Toeplitz matrix whose (j, l) entry is, $\Sigma_i(j, l) = \sigma^2 \phi^{|j-l|}/(1 - \phi^2)$, and $|\phi| < 1$. The most natural SS formulation is to treat the mean as an exogenous input, leading to the following model for each subject i :

$$y_t = \beta + \varepsilon_t, \quad \varepsilon_t = \phi \varepsilon_{t-1} + w_t, \quad w_t \sim \text{NID}(0, \sigma^2), \quad (11.29)$$

with initial condition $\varepsilon_0 \sim \text{N}(0, \sigma^2/(1 - \phi^2))$. A potential problem with this approach is that β must be estimated, along with ϕ and σ^2 , via nonlinear optimization. To circumvent this difficulty, an alternative route proceeds by including the mean in the (now two-dimensional) state vector by specifying $\beta \equiv \beta_t$ for all t , and defining the system

$$\begin{bmatrix} \varepsilon_t \\ \beta_t \end{bmatrix} = \begin{bmatrix} \phi & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \varepsilon_{t-1} \\ \beta_{t-1} \end{bmatrix} + \begin{bmatrix} w_t \\ 0 \end{bmatrix}, \quad \begin{bmatrix} w_t \\ 0 \end{bmatrix} \sim \text{NID}(\mathbf{0}, \mathbf{W}), \quad (11.30)$$

with initial condition $[\varepsilon_0, \beta_0]^T \sim \text{N}(0, \mathbf{G})$ on the state vector. The structure of the two covariance matrices, \mathbf{G} and \mathbf{W} , is similar,

$$\mathbf{G} = \begin{bmatrix} \frac{\sigma^2}{1-\phi^2} & 0 \\ 0 & 0 \end{bmatrix}, \quad \text{and} \quad \mathbf{W} = \begin{bmatrix} \sigma^2 & 0 \\ 0 & 0 \end{bmatrix}.$$

In comparison with the first, the problem with this second formulation is that a more complex initial condition has to be entertained, along with a singular structure for the two covariance matrices above.

11.4 Maximum likelihood estimation

This section discusses ML estimation for the set of parameters $\Theta = (\boldsymbol{\mu}, \mathbf{P}, \boldsymbol{\Phi}, \boldsymbol{\Upsilon}, \mathbf{Q}, \mathbf{D}, \mathbf{R})$ in SS model (11.2)–(11.3). The asymptotic properties of these estimates are discussed in Appendix B. We extend the notation already introduced in Section 11.2 for the concatenation of observation vectors $Z_{i,s}$ and exogenous vectors $U_{i,s}$ up to time s , to include also the state vectors $X_{i,s} = (\mathbf{x}_{i,0}, \mathbf{x}_{i,1}, \dots, \mathbf{x}_{i,s})$. In addition, let $\mathbf{X} = (X_{1,n_1}, \dots, X_{k,n_k})$, $\mathbf{Z} = (Z_{1,n_1}, \dots, Z_{k,n_k})$, and $\mathbf{U} = (U_{1,n_1}, \dots, U_{k,n_k})$, with $\mathbf{Y} = (\mathbf{X}, \mathbf{Z}, \mathbf{U})$ denoting the *complete* data. Note that (\mathbf{Z}, \mathbf{U}) constitutes the *observed* data. If all the states X_{i,n_i} were observable ($i = 1, \dots, k$), then the complete data joint density could be decomposed into a product of conditional densities for the $(\mathbf{x}_{i,t} | \mathbf{x}_{i,t-1}, \mathbf{u}_{i,t})$, the $(\mathbf{z}_{i,t} | \mathbf{x}_{i,t}, \mathbf{u}_{i,t})$, and the marginals of $\mathbf{x}_{i,0}$, yielding the likelihood function

$$\begin{aligned} f_\Theta(\mathbf{Y}) &= \prod_{i=1}^k \left\{ f_{\boldsymbol{\mu}, \mathbf{P}}(\mathbf{x}_{i,0}) \prod_{t=1}^{n_i} f_{\boldsymbol{\Phi}, \boldsymbol{\Upsilon}, \mathbf{Q}}(\mathbf{x}_{i,t} | \mathbf{x}_{i,t-1}, \mathbf{u}_{i,t}) \prod_{t=1}^{n_i} f_{\mathbf{D}, \mathbf{R}}(\mathbf{z}_{i,t} | \mathbf{x}_{i,t}, \mathbf{u}_{i,t}) \right\} \\ &:= L(\Theta | \mathbf{Y}). \end{aligned}$$

Two times the natural logarithm of this *complete* data likelihood, $2\log L(\Theta | \mathbf{Y})$, ignoring a constant, assuming joint multivariate normality for the error processes and initial state, and letting $\sum_{i=1}^k n_i = N$, can be written as

$$\begin{aligned} \ell(\Theta | \mathbf{Y}) &= k \log |\mathbf{P}^{-1}| \\ &\quad - \sum_{i=1}^k (\mathbf{x}_{i,0} - \boldsymbol{\mu})^T \mathbf{P}^{-1} (\mathbf{x}_{i,0} - \boldsymbol{\mu}) + N (\log |\mathbf{Q}^{-1}| + \log |\mathbf{R}^{-1}|) \\ &\quad - \sum_{i=1}^k \sum_{t=1}^{n_i} (\mathbf{x}_{i,t} - \boldsymbol{\Phi} \mathbf{x}_{i,t-1} - \boldsymbol{\Upsilon} \mathbf{u}_{i,t})^T \mathbf{Q}^{-1} (\mathbf{x}_{i,t} - \boldsymbol{\Phi} \mathbf{x}_{i,t-1} - \boldsymbol{\Upsilon} \mathbf{u}_{i,t}) \\ &\quad - \sum_{i=1}^k \sum_{t=1}^{n_i} (\mathbf{z}_{i,t} - \mathbf{H}_{i,t} \mathbf{x}_{i,t} - \mathbf{D} \mathbf{u}_{i,t})^T \mathbf{R}^{-1} (\mathbf{z}_{i,t} - \mathbf{H}_{i,t} \mathbf{x}_{i,t} - \mathbf{D} \mathbf{u}_{i,t}). \end{aligned} \quad (11.31)$$

The *observed* data counterpart to $\ell(\Theta | \mathbf{Y})$ however, is the function that must be maximized in order to produce the ML estimate of Θ . Ignoring a constant, twice the observed data log-likelihood can conveniently be expressed in innovations form as

$$\ell(\Theta | \mathbf{Z}, \mathbf{U}) := 2\log L(\Theta | \mathbf{Z}, \mathbf{U}) = \sum_{i=1}^k \sum_{t=1}^{n_i} \left(\log |\mathbf{B}_{i,t}^{-1}| - \tilde{\mathbf{z}}_{i,t}^T \mathbf{B}_{i,t}^{-1} \tilde{\mathbf{z}}_{i,t} \right), \quad (11.32)$$

where the *innovations* (or prediction errors) in (11.32) are defined as

$$\tilde{\mathbf{z}}_{i,t} = \mathbf{z}_{i,t} - \mathbb{E}(\mathbf{z}_{i,t} | \mathbf{Z}_{i,t-1}, \mathbf{U}_{i,t-1}) = \mathbf{z}_{i,t} - \mathbf{H}_{i,t} \mathbf{x}_{i,t}^{t-1} - \mathbf{D} \mathbf{u}_{i,t}. \quad (11.33)$$

Clearly, $\mathbb{E}(\tilde{\mathbf{z}}_{i,t}) = \mathbf{0}$, so that the innovations sequence $\{\tilde{\mathbf{z}}_{i,t}\}$, $t = 1, \dots, n_i$, comprise an independent Gaussian process with covariance matrix

$$\mathbf{B}_{i,t} := \text{cov}(\tilde{\mathbf{z}}_{i,t}) = \text{cov} \left[\mathbf{H}_{i,t} \left(\mathbf{x}_{i,t} - \mathbf{x}_{i,t}^{t-1} \right) + \mathbf{v}_{i,t} \right] = \mathbf{H}_{i,t} \mathbf{P}_{i,t}^{t-1} \mathbf{H}_{i,t}^T + \mathbf{R}.$$

Assuming independence among subjects $i, i = 1, \dots, k$, leads to (11.32).

As discussed in Chapter 5, the usual method of maximizing (11.32) is to use an iterative procedure that combines propagation of the Kalman recursions (11.7)–(11.11) for a given parameter value, $\Theta^{(j-1)}$, with a Newton-Raphson algorithm in order to produce an updated parameter value, $\Theta^{(j)}$. This is the usual procedure advocated by say, Jones [132]. However, since the advent of the Expectation-Maximization algorithm, see Dempster, Laird and Rubin [73], an EM based optimization scheme originally devised by Shumway and Stoffer [194] for the standard SS model (without cross-sectional extension) has gained popularity, partly due to the fact that it does not require computation of the gradient vector and Hessian matrix of the log-likelihood (see, also, Shumway and Stoffer [195, ch. 6]). For the cross-sectionally extended model (11.2)–(11.3), Naranjo, Trindade and Casella [165] provide complete closed-form expressions for the E and M-steps as we will now detail.

The EM algorithm consecutively maximizes the expectation of the complete data likelihood given the observations to find the ML estimate of Θ based on the observed data (\mathbf{Z}, \mathbf{U}) , since the complete data $Y = (\mathbf{X}, \mathbf{Z}, \mathbf{U})$ are unknown. It begins with implementation of the E-step. At iteration j , let

$$\varphi(\Theta | \Theta^{(j-1)}) = E[-\ell(\Theta | Y) | \mathbf{Z}, \mathbf{U}, \Theta^{(j-1)}], \quad \text{and} \quad \Psi = [\Phi, \Upsilon],$$

define the relevant E-step function, and the matrix formed by concatenating Φ and Υ , in order to simplify notation. A straightforward computation then yields

$$\begin{aligned} \varphi(\Theta | \Theta^{(j-1)}) &= k \log |\mathbf{P}| + N(\log |\mathbf{Q}| + \log |\mathbf{R}|) \\ &\quad + \sum_{i=1}^k \text{tr} \left\{ \mathbf{P}_{i,0}^{-1} \left[\mathbf{P}_{i,0}^{n_i} + (\mathbf{x}_{i,0}^{n_i} - \boldsymbol{\mu}) (\mathbf{x}_{i,0}^{n_i} - \boldsymbol{\mu})^T \right] \right\} \\ &\quad + \text{tr} \{ \mathbf{Q}^{-1} (\mathbf{S}_{11} - \mathbf{S}_{10}\Psi^T - \Psi\mathbf{S}_{10}^T + \Psi\mathbf{S}_{00}\Psi^T) \} \\ &\quad + \text{tr} \{ \mathbf{R}^{-1} (\mathbf{T}_{11} - \mathbf{T}_{10}\mathbf{D}^T - \mathbf{D}\mathbf{T}_{10}^T + \mathbf{D}\mathbf{T}_{00}\mathbf{D}^T) \}, \end{aligned} \quad (11.34)$$

where we define the following “sums of squares” type terms

$$\begin{aligned} \mathbf{S}_{11} &= \sum_{i=1}^k \sum_{t=1}^{n_i} \left(\mathbf{P}_{i,t}^{n_i} + \mathbf{x}_{i,t}^{n_i} \mathbf{x}_{i,t}^{n_i \top} \right), \\ \mathbf{S}_{10} &= \sum_{i=1}^k \sum_{t=1}^{n_i} \left[\mathbf{P}_{i,t,t-1}^{n_i} + \mathbf{x}_{i,t}^{n_i} \mathbf{x}_{i,t-1}^{n_i \top} \quad \mathbf{x}_{i,t}^{n_i} \mathbf{u}_{i,t}^T \right], \\ \mathbf{S}_{00} &= \sum_{i=1}^k \sum_{t=1}^{n_i} \left[\mathbf{P}_{i,t-1}^{n_i} + \mathbf{x}_{i,t-1}^{n_i} \mathbf{x}_{i,t-1}^{n_i \top} \quad \mathbf{x}_{i,t-1}^{n_i} \mathbf{u}_{i,t}^T \right], \end{aligned}$$

and

$$\begin{aligned}\mathbf{T}_{00} &= \sum_{i=1}^k \sum_{t=1}^{n_i} \mathbf{u}_{i,t} \mathbf{u}_{i,t}^T, \\ \mathbf{T}_{10} &= \sum_{i=1}^k \sum_{t=1}^{n_i} (\mathbf{z}_{i,t} - \mathbf{H}_{i,t} \mathbf{x}_{i,t}^{n_i}) \mathbf{u}_{i,t}^T, \\ \mathbf{T}_{11} &= \sum_{i=1}^k \sum_{t=1}^{n_i} \left[(\mathbf{z}_{i,t} - \mathbf{H}_{i,t} \mathbf{x}_{i,t}^{n_i}) (\mathbf{z}_{i,t} - \mathbf{H}_{i,t} \mathbf{x}_{i,t}^{n_i})^T + \mathbf{H}_{i,t} \mathbf{P}_{i,t}^{n_i} \mathbf{H}_{i,t}^T \right].\end{aligned}$$

Maximization of (11.34) with respect to Θ at iteration j constitutes the M-step. Taking derivatives with respect to each component of Θ , setting to zero, and solving for each parameter, produces the updated estimators

$$\begin{aligned}\boldsymbol{\mu}^{(j)} &= \frac{1}{k} \sum_{i=1}^k \mathbf{x}_{i,0}^{n_i}, \\ \mathbf{P}^{(j)} &= \frac{1}{k} \sum_{i=1}^k \left[\mathbf{P}_{i,0}^{n_i} + (\mathbf{x}_{i,0}^{n_i} - \boldsymbol{\mu}^{(j)}) (\mathbf{x}_{i,0}^{n_i} - \boldsymbol{\mu}^{(j)})^T \right], \\ \boldsymbol{\Psi}^{(j)} &= \mathbf{S}_{10} \mathbf{S}_{00}^{-1}, \\ \mathbf{Q}^{(j)} &= \frac{1}{N} \left(\mathbf{S}_{11} - \mathbf{S}_{10} \mathbf{S}_{00}^{-1} \mathbf{S}_{10}^T \right), \\ \mathbf{D}^{(j)} &= \mathbf{T}_{10} \mathbf{T}_{00}^{-1}, \\ \mathbf{R}^{(j)} &= \frac{1}{N} \left(\mathbf{T}_{11} - \mathbf{T}_{10} \mathbf{T}_{00}^{-1} \mathbf{T}_{10}^T \right).\end{aligned}$$

11.5 Missing data modifications

Missing data is an almost *de facto* consequence of longitudinal studies of any complexity, especially when biological variables are involved. The SS model can, with certain modifications, accommodate missing values in both responses ($\mathbf{z}_{i,t}$) and covariates ($\mathbf{u}_{i,t}$). We first discuss methods for handling missingness in responses only, followed by two methods for handling missingness in both responses and covariates. Strictly speaking, these last two methods apply to different SS model formulations, although they do overlap in a particular instance.

11.5.1 Missingness in responses only

The classical approach to this problem proceeds by imputing the missing values (viewed as irregularly-spaced data) via the Kalman recursions, whence an expression for the complete data likelihood is available; see e.g., Brockwell and Davis [32, ch. 12]. On the other hand, an EM algorithm-based solution offers the additional attraction of providing a procedure for not only imputing, but also maximizing; see Shumway and Stoffer [195, ch. 6]. We focus our attention on the latter.

The necessary modifications to the EM procedure presented in Section 11.4 is most easily discussed by simplifying model (11.2)–(11.3) to exclude covariates,

$$\mathbf{x}_{i,t} = \Phi \mathbf{x}_{i,t-1} + \mathbf{w}_{i,t}, \quad (11.35)$$

$$\begin{bmatrix} \mathbf{z}_{i,t}^{(o)} \\ \mathbf{z}_{i,t}^{(u)} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{i,t}^{(o)} \\ \mathbf{H}_{i,t}^{(u)} \end{bmatrix} \mathbf{x}_{i,t} + \begin{bmatrix} \mathbf{v}_{i,t}^{(o)} \\ \mathbf{v}_{i,t}^{(u)} \end{bmatrix}, \quad (11.36)$$

with

$$\text{cov} \left(\begin{bmatrix} \mathbf{v}_{i,t}^{(o)} \\ \mathbf{v}_{i,t}^{(u)} \end{bmatrix} \right) = \begin{bmatrix} \mathbf{R}_{i,t}^{(o)} & \mathbf{R}_{i,t}^{(ou)} \\ \mathbf{R}_{i,t}^{(ou)T} & \mathbf{R}_{i,t}^{(u)} \end{bmatrix},$$

where we have rearranged the observation vector for unit i at time t in partitioned form, with $\mathbf{z}_{i,t}^{(o)}$ denoting the *observed* part and $\mathbf{z}_{i,t}^{(u)}$ the *unobserved* part. The KF from Section 11.2 can now be applied by “zeroing out” the appropriate portions of the partitioned vectors and matrices in the observation equation, i.e. $\mathbf{z}_{i,t}^{(u)} = \mathbf{0}$, $\mathbf{H}_{i,t}^{(u)} = \mathbf{0}$, and $\mathbf{R}_{i,t}^{(ou)} = \mathbf{0}$; see Stoffer [200] for a justification of the validity of this approach.

In similar notation to Section 11.4 (with \mathbf{U} omitted), if $\mathbf{Z}^{(o)} = (\mathbf{z}_{1,1}^{(o)}, \dots, \mathbf{z}_{k,n_k}^{(o)})$ denotes the concatenation of all observed responses, then $\mathbf{Y} = (\mathbf{X}, \mathbf{Z})$ is still the complete data, but the observed data is now obviously only $\mathbf{Z}^{(o)}$. The expression for twice the complete data log-likelihood (11.31) remains unchanged, but in implementing the E-step we now condition on the portion of \mathbf{Z} that is actually observed, so that we must compute

$$\varphi(\Theta | \Theta^{(j-1)}) = \text{E} \left[-\ell(\Theta | \mathbf{Y}) \mid \mathbf{Z}^{(o)}, \Theta^{(j-1)} \right],$$

which results in the expression

$$\begin{aligned} \varphi(\Theta | \Theta^{(j-1)}) &= k \log |\mathbf{P}| + N(\log |\mathbf{Q}| + \log |\mathbf{R}|) + \text{tr} \{ \mathbf{R}^{-1} \mathbf{T}_{11} \} \\ &\quad + \sum_{i=1}^k \text{tr} \left\{ \mathbf{P}_{i,0}^{-1} \left[\mathbf{P}_{i,0}^{(n_i)} + (\mathbf{x}_{i,0}^{(n_i)} - \boldsymbol{\mu}) (\mathbf{x}_{i,0}^{(n_i)})^T \right] \right\} \\ &\quad + \text{tr} \{ \mathbf{Q}^{-1} (\mathbf{S}_{11} - \mathbf{S}_{10} \Phi^T - \Phi \mathbf{S}_{10}^T + \Phi \mathbf{S}_{00} \Phi^T) \}, \end{aligned}$$

where the sums of squares terms are now

$$\mathbf{S}_{11} = \sum_{i=1}^k \sum_{t=1}^{n_i} \left[\mathbf{P}_{i,t}^{(n_i)} + \mathbf{x}_{i,t}^{(n_i)} \mathbf{x}_{i,t}^{(n_i)T} \right],$$

$$\mathbf{S}_{10} = \sum_{i=1}^k \sum_{t=1}^{n_i} \left[\mathbf{P}_{i,t-1}^{(n_i)} + \mathbf{x}_{i,t}^{(n_i)} \mathbf{x}_{i,t-1}^{(n_i)T} \right],$$

$$\mathbf{S}_{00} = \sum_{i=1}^k \sum_{t=1}^{n_i} \left[\mathbf{P}_{i,t-1}^{(n_i)} + \mathbf{x}_{i,t-1}^{(n_i)} \mathbf{x}_{i,t-1}^{(n_i)T} \right],$$

and

$$\begin{aligned} T_{11} = \sum_{i=1}^k \sum_{t=1}^{n_i} & \left\{ \left(\mathbf{z}_{i,t} - \mathbf{H}_{i,t} \mathbf{x}_{i,t}^{(n_i)} \right) \left(\mathbf{z}_{i,t} - \mathbf{H}_{i,t} \mathbf{x}_{i,t}^{(n_i)} \right)^T \right. \\ & \left. + \mathbf{H}_{i,t} \mathbf{P}_{i,t}^{(n_i)} \mathbf{H}_{i,t}^T + \begin{bmatrix} \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{R}_{i,t}^{(u)} \end{bmatrix} \right\}. \quad (11.37) \end{aligned}$$

The T_{11} matrix now involves the covariance matrix for the noise vector $\mathbf{v}_{i,t}$, due to computation of the terms $E(\mathbf{z}_{i,t}^{(u)} | \mathbf{Z}^{(o)})$ and $E(\mathbf{z}_{i,t}^{(u)} \mathbf{z}_{i,t}^{(u)T} | \mathbf{Z}^{(o)})$. Note also that an adjustment to the filtered state values $\mathbf{x}_{i,t}^s$ and their error covariance matrices $\mathbf{P}_{i,t}^s$ was necessary, so that these have been replaced by

$$\mathbf{x}_{i,t}^{(s)} = E \left(\mathbf{x}_{i,t} \mid \mathbf{z}_{i,1}^{(o)}, \dots, \mathbf{z}_{i,s}^{(o)} \right),$$

and

$$\mathbf{P}_{i,t_1, t_2}^{(s)} = E \left[\left(\mathbf{x}_{i,t_1} - \mathbf{x}_{i,t_1}^{(s)} \right) \left(\mathbf{x}_{i,t_2} - \mathbf{x}_{i,t_2}^{(s)} \right)^T \mid \mathbf{z}_{i,1}^{(o)}, \dots, \mathbf{z}_{i,s}^{(o)} \right],$$

respectively. Implementation of the M-step (maximization with respect to Θ at iteration j) produces the updated estimators

$$\begin{aligned} \boldsymbol{\mu}^{(j)} &= \frac{1}{k} \sum_{i=1}^k \mathbf{x}_{i,0}^{(n_i)}, \\ \mathbf{P}^{(j)} &= \frac{1}{k} \sum_{i=1}^k \left[\mathbf{P}_{i,0}^{(n_i)} + \left(\mathbf{x}_{i,0}^{(n_i)} - \boldsymbol{\mu}^{(j)} \right) \left(\mathbf{x}_{i,0}^{(n_i)} - \boldsymbol{\mu}^{(j)} \right)^T \right], \\ \boldsymbol{\Phi}^{(j)} &= \mathbf{S}_{10} \mathbf{S}_{00}^{-1}, \\ \mathbf{Q}^{(j)} &= \frac{1}{N} \left(\mathbf{S}_{11} - \mathbf{S}_{10} \mathbf{S}_{00}^{-1} \mathbf{S}_{10}^T \right), \\ \mathbf{R}^{(j)} &= \frac{1}{N} \mathbf{M}_{i,t} \mathbf{T}_{11} \mathbf{M}_{i,t}^T, \end{aligned}$$

where $\mathbf{M}_{i,t}$ is a permutation matrix (consisting of zeros and ones) that reorders the response vector for subject i at time t in its original order, i.e., if $\mathbf{z}_{i,t}^* = [\mathbf{z}_{i,t}^{(o)T}, \mathbf{z}_{i,t}^{(u)T}]^T$ denotes the missing data rearranged $\mathbf{z}_{i,t}$, then $\mathbf{M}_{i,t} \mathbf{z}_{i,t}^* = \mathbf{z}_{i,t}$.

11.5.2 Missingness in both responses and covariates: method 1

Accommodation of missing values in both responses ($\mathbf{z}_{i,t}$) and covariates ($\mathbf{u}_{i,t}$), can most easily be achieved for the SS model formulation of (11.1) and (11.3)

$$\mathbf{x}_{i,t}^a = \boldsymbol{\Phi}^a \mathbf{x}_{i,t-1}^a + \boldsymbol{\Gamma} \mathbf{u}_{i,t-1} + \mathbf{w}_{i,t}^a, \quad \{\mathbf{w}_{i,t}^a\} \sim NID(\mathbf{0}, Q^a), \quad (11.38)$$

$$\mathbf{z}_{i,t} = \mathbf{H}_{i,t}^a \mathbf{x}_{i,t}^a + \mathbf{D} \mathbf{u}_{i,t} + \mathbf{v}_{i,t}^a, \quad \{\mathbf{v}_{i,t}^a\} \sim NID(\mathbf{0}, R^a). \quad (11.39)$$

In order to provide a mechanism for imputing the potentially missing (in whole or in part) exogenous $\mathbf{u}_{i,t}$, we now introduce an additional SS model for it

$$\mathbf{x}_{i,t}^b = \Phi^b \mathbf{x}_{i,t-1}^b + \mathbf{w}_{i,t}^b, \quad \{\mathbf{w}_{i,t}^b\} \sim \text{NID}(\mathbf{0}, Q^b), \quad (11.40)$$

$$\mathbf{u}_{i,t} = \mathbf{H}_{i,t}^b \mathbf{x}_{i,t}^b + \mathbf{v}_{i,t}^b, \quad \{\mathbf{v}_{i,t}^b\} \sim \text{NID}(\mathbf{0}, R^b). \quad (11.41)$$

Substituting (11.41) into (11.38) and (11.39), leads to the system

$$\mathbf{x}_{i,t}^a = \Phi^a \mathbf{x}_{i,t-1}^a + \Gamma (\mathbf{H}_{i,t-1}^b \mathbf{x}_{i,t-1}^b + \mathbf{v}_{i,t-1}^b) + \mathbf{w}_{i,t}^a, \quad (11.42)$$

$$\mathbf{z}_{i,t} = \mathbf{H}_{i,t}^a \mathbf{x}_{i,t}^a + \mathbf{D} (\mathbf{H}_{i,t}^b \mathbf{x}_{i,t}^b + \mathbf{v}_{i,t}^b) + \mathbf{v}_{i,t}^a. \quad (11.43)$$

We can now combine (11.42) and (11.40) into a single SS equation,

$$\begin{bmatrix} \mathbf{x}_{i,t}^a \\ \mathbf{x}_{i,t}^b \end{bmatrix} = \begin{bmatrix} \Phi^a & \Gamma \mathbf{H}_{i,t-1}^a \\ \mathbf{0} & \Phi^b \end{bmatrix} \begin{bmatrix} \mathbf{x}_{i,t-1}^a \\ \mathbf{x}_{i,t-1}^b \end{bmatrix} + \begin{bmatrix} \Gamma \mathbf{v}_{i,t-1}^b + \mathbf{w}_{i,t}^a \\ \mathbf{w}_{i,t}^b \end{bmatrix}, \quad (11.44)$$

and (11.41) and (11.43) into the single observation equation,

$$\begin{bmatrix} \mathbf{z}_{i,t} \\ \mathbf{u}_{i,t} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{i,t}^a & \mathbf{D} \mathbf{H}_{i,t}^b \\ \mathbf{0} & \mathbf{H}_{i,t}^b \end{bmatrix} \begin{bmatrix} \mathbf{x}_{i,t}^a \\ \mathbf{x}_{i,t}^b \end{bmatrix} + \begin{bmatrix} \mathbf{D} \mathbf{v}_{i,t}^b + \mathbf{v}_{i,t}^a \\ \mathbf{v}_{i,t}^b \end{bmatrix}. \quad (11.45)$$

Note that this reformulation essentially aggregates (11.38)–(11.41), into an equivalent system that is a SS model without exogenous variables, as in (11.35)–(11.36). In so doing, the exogenous variables become endogenous, i.e., the left-hand side of (11.45) is the new response variable. This “trick” of *endogeneization* (already introduced in Section 3.2) has been known since at least Mehra [163], but was more visibly presented by Terceiro [206] and others.

In order to deal with missing values, we now reorder the (new) response vector in (11.45),

$$\begin{bmatrix} \mathbf{z}_{i,t} \\ \mathbf{u}_{i,t} \end{bmatrix} \longrightarrow \begin{bmatrix} \mathbf{z}_{i,t}^{(o)} \\ \mathbf{u}_{i,t}^{(o)} \\ \mathbf{z}_{i,t}^{(u)} \\ \mathbf{u}_{i,t}^{(u)} \end{bmatrix},$$

where $\mathbf{z}_{i,t}^{(o)}$ and $\mathbf{u}_{i,t}^{(o)}$ denote the observed parts of the response and covariate, and $\mathbf{z}_{i,t}^{(u)}$ and $\mathbf{u}_{i,t}^{(u)}$ the unobserved parts. After corresponding reordering of the state vector, the newly reordered system is now seen to be in the form of (11.35)–(11.36), whence the missing portion of the “response” vector, $[\mathbf{z}_{i,t}^{(u)}, \mathbf{u}_{i,t}^{(u)}]^T$, can be handled through the imputation method outlined in Subsection 11.5.1.

Remark 11.2 A byproduct of this approach for handling missing data is that the resulting state equation (11.44) now has a time-varying state coefficient matrix. As discussed in Result 11.1, this does not pose a problem from the standpoint of Kalman filtering. However, from an empirical perspective, it does raise the issue of how to “estimate” a time-varying state coefficient matrix; see Chapter 6.

11.5.3 Missingness in both responses and covariates: method 2

If instead of (11.1), the alternative formulation (11.2) admitting a contemporaneous relationship between the state and exogenous variables is used,

$$\mathbf{x}_{i,t}^a = \Phi^a \mathbf{x}_{i,t-1}^a + \Upsilon \mathbf{u}_{i,t} + \mathbf{w}_{i,t}^a, \quad \{\mathbf{w}_{i,t}^a\} \sim NID(\mathbf{0}, Q^a), \quad (11.46)$$

$$\mathbf{z}_{i,t} = \mathbf{H}_{i,t}^a \mathbf{x}_{i,t}^a + \mathbf{D} \mathbf{u}_{i,t} + \mathbf{v}_{i,t}^a, \quad \{\mathbf{v}_{i,t}^a\} \sim NID(\mathbf{0}, R^a), \quad (11.47)$$

the trick of endogeneization will not be able to render the resulting system in the form of a new SS model as in (11.44)–(11.45). This is due to the time index match between the state variable $\mathbf{x}_{i,t}$ and covariate $\mathbf{u}_{i,t}$ in (11.46), which prevents a reformulation of the system into an aggregated state equation as in (11.44). Nevertheless, one can follow the analogous idea of introducing an extra SS model in order to provide a mechanism for imputing the missing covariates.

The details of this argument are as follows. We again partition the observation vector into observed and unobserved components, respectively, as $\mathbf{z}_{i,t} = (\mathbf{z}_{i,t}^{(o)T}, \mathbf{z}_{i,t}^{(u)T})^T$, and similarly for the exogenous variables, $\mathbf{u}_{i,t} = (\mathbf{u}_{i,t}^{(o)T}, \mathbf{u}_{i,t}^{(u)T})^T$. The state and observation equations can then be re-expressed as

$$\mathbf{x}_{i,t}^a = \Phi \mathbf{x}_{i,t-1}^a + [\Upsilon^{(o)} \quad \Upsilon^{(u)}] \begin{bmatrix} \mathbf{u}_{i,t}^{(o)} \\ \mathbf{u}_{i,t}^{(u)} \end{bmatrix} + \mathbf{w}_{i,t}^a, \quad (11.48)$$

and

$$\begin{bmatrix} \mathbf{z}_{i,t}^{(o)} \\ \mathbf{z}_{i,t}^{(u)} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{i,t}^{a(o)} \\ \mathbf{H}_{i,t}^{a(u)} \end{bmatrix} \mathbf{x}_{i,t}^a + \begin{bmatrix} D_{11}^{(o)} & D_{12}^{(u)} \\ D_{21}^{(o)} & D_{22}^{(u)} \end{bmatrix} \begin{bmatrix} \mathbf{u}_{i,t}^{(o)} \\ \mathbf{u}_{i,t}^{(u)} \end{bmatrix} + \begin{bmatrix} \mathbf{v}_{i,t}^{a(o)} \\ \mathbf{v}_{i,t}^{a(u)} \end{bmatrix}. \quad (11.49)$$

The covariance between the observed and unobserved portions of the observational error vector is given by

$$\text{cov} \left(\begin{bmatrix} \mathbf{v}_{i,t}^{a(o)} \\ \mathbf{v}_{i,t}^{a(u)} \end{bmatrix} \right) = \begin{bmatrix} \mathbf{R}_{11}^a & \mathbf{R}_{12}^a \\ \mathbf{R}_{21}^a & \mathbf{R}_{22}^a \end{bmatrix}.$$

In (11.48)–(11.49), $\mathbf{x}_{i,t}^a$ and $\mathbf{z}_{i,t}$ depend on the unobserved portion of the exogenous variable vector, $\mathbf{u}_{i,t}^{(u)}$. To model $\mathbf{u}_{i,t}$ with a second SS model, we treat $\mathbf{u}_{i,t}$ as the observation vector for an underlying state process, $\mathbf{x}_{i,t}^b$, $i = 1, \dots, k$ and $t = 1, \dots, n_i$. This second (imputation) model is given by

$$\mathbf{x}_{i,t}^b = \Phi^b \mathbf{x}_{i,t-1}^b + \mathbf{w}_{i,t}^b, \quad (11.50)$$

$$\mathbf{u}_{i,t} = \mathbf{H}_{i,t}^b \mathbf{x}_{i,t}^b + \mathbf{v}_{i,t}^b, \quad (11.51)$$

where Φ^b denotes the state transition matrix, with $\mathbf{w}_{i,t}^b \sim NID(\mathbf{0}, Q^b)$ the corresponding noise vector. $\mathbf{H}_{i,t}^b$ is now the observation matrix, and $\mathbf{v}_{i,t}^b \sim NID(\mathbf{0}, \mathbf{R}^b)$ the observational noise vector, with initial condition $\mathbf{x}_{i,0}^b \sim NID(\boldsymbol{\mu}^b, \mathbf{P}^b)$. For simplicity, $\{\mathbf{w}_{i,t}^a\}$, $\{\mathbf{v}_{i,t}^a\}$, $\{\mathbf{w}_{i,t}^b\}$, and $\{\mathbf{v}_{i,t}^b\}$ are assumed to be independent.

Now, the observation equation in (11.51) can be partitioned into observed and missing parts, yielding

$$\begin{bmatrix} \mathbf{u}_{i,t}^{(o)} \\ \mathbf{u}_{i,t}^{(u)} \end{bmatrix} = \begin{bmatrix} \mathbf{H}_{i,t}^{b(o)} \\ \mathbf{H}_{i,t}^{b(u)} \end{bmatrix} \mathbf{x}_{i,t}^b + \begin{bmatrix} \mathbf{v}_{i,t}^{b(o)} \\ \mathbf{v}_{i,t}^{b(u)} \end{bmatrix}, \quad (11.52)$$

with the covariance between the observed and unobserved portions of the observational error vector $\mathbf{v}_{i,t}^b$ given by

$$\text{cov} \left(\begin{bmatrix} \mathbf{v}_{i,t}^{b(o)} \\ \mathbf{v}_{i,t}^{b(u)} \end{bmatrix} \right) = \begin{bmatrix} \mathbf{R}_{11}^b & \mathbf{R}_{12}^b \\ \mathbf{R}_{21}^b & \mathbf{R}_{22}^b \end{bmatrix}.$$

Finally, the likelihood must be modified to account for the second SS model. Define

$$\begin{aligned} 2\log L(\Theta^b | X^b, U) &= k\log|\mathbf{P}^b|^{-1} + N \left(\log|(\mathbf{Q}^b)^{-1}| + \log|(\mathbf{R}^b)^{-1}| \right) \\ &\quad - \sum_{i=1}^k \left(\mathbf{x}_{i,0}^b - \boldsymbol{\mu}^b \right)^T (\mathbf{P}^b)^{-1} \left(\mathbf{x}_{i,0}^b - \boldsymbol{\mu}^b \right) \\ &\quad - \sum_{i=1}^k \sum_{t=1}^{n_i} \left(\mathbf{x}_{i,t}^b - \boldsymbol{\Phi}^b \mathbf{x}_{i,t-1}^b \right)^T (\mathbf{Q}^b)^{-1} \left(\mathbf{x}_{i,t}^b - \boldsymbol{\Phi}^b \mathbf{x}_{i,t-1}^b \right) \\ &\quad - \sum_{i=1}^k \sum_{t=1}^{n_i} \left(\mathbf{u}_{i,t}^b - \mathbf{H}_{i,t}^b \mathbf{x}_{i,t}^b \right)^T (\mathbf{R}^b)^{-1} \left(\mathbf{u}_{i,t}^b - \mathbf{H}_{i,t}^b \mathbf{x}_{i,t}^b \right), \end{aligned} \quad (11.53)$$

where $\Theta^b = (\boldsymbol{\mu}^b, \mathbf{P}^b, \boldsymbol{\Phi}^b, \mathbf{Q}^b, \mathbf{R}^b)$ represents the parameters from the second SS formulation used to model the exogenous variables, and $\mathbf{X}^b = (X_{1,n_1}^b, \dots, X_{k,n_k}^b)$ denotes the state information for the second SS model, with $X_{i,n_i}^b = (\mathbf{x}_{i,0}^b, \dots, \mathbf{x}_{i,n_i}^b)$. Ignoring constants in (11.31) and (11.53), the natural logarithm of the complete data likelihood is now

$$2\log L(\Theta, \Theta^b | X, Z, X^b, U) = 2\log L(\Theta | X, Z, U) + 2\log L(\Theta^b | X^b, U), \quad (11.54)$$

where Θ are the model parameters to be estimated in (11.46)–(11.47), with $\mathbf{Y} = (X, Z, U)$ the corresponding concatenation of the complete data.

The resulting system is non-standard, requiring the derivation of relevant KF equations. This approach, together with the development of an Expectation/Conditional Maximization (ECM) algorithm for parameter estimation, was derived by Naranjo, Trindade and Casella [165]. The non-standard setup therefore makes this a less attractive method of dealing with missingness than the first method outlined in Subsection 11.5.2. To this end, note from Remark 11.3 below that if we only include the exogenous variables in the observation equation of the system (11.46)–(11.47), then the endogeneization trick will be able to accommodate the missingness in both responses and covariates.

Remark 11.3 Including the exogenous only in the observation equation by setting $\boldsymbol{\Upsilon} = \mathbf{0}$ in the system (11.46)–(11.47), one obtains a simplified version of (11.44)–(11.45) where $\boldsymbol{\Gamma} = \mathbf{0}$. This is tantamount to fitting a so-called regression with ARMA errors (if correlation is allowed between the $\{\mathbf{w}_{i,t}\}$ and $\{\mathbf{v}_{i,t}\}$ noise processes).

11.6 Real data examples

11.6.1 A LMM for the mare ovarian follicles data

The dataset `Ovary` in the `nlme` library of R, see [186], contains the number of ovarian follicles ($> 10\text{mm}$ in diameter) detected in 11 different mares at several times in their estrus cycle. The follicles present in each mare were counted over one complete cycle, starting 3 days before ovulation and ending 3 days after the next ovulation. The times of measurement vary by animal, and are scaled so that ovulation occurs at times 0 and 1.

Pinheiro and Bates [184, ch. 5] consider a LMM with sinusoidal components to explain the oscillatory nature of the data. If $y_{i,t}$ denotes the number of follicles in mare i at time t (really time t_{ij} since the j -th measurement time for the i -th mare varies for each animal), the model is

$$y_{i,t} = (\beta_0 + b_{0,i}) + (\beta_1 + b_{1,i}) \sin(2\pi t) + \beta_2 \cos(2\pi t) + \varepsilon_{i,t},$$

where

$$\boldsymbol{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{bmatrix}, \quad \mathbf{b}_i = \begin{bmatrix} b_{0,i} \\ b_{1,i} \end{bmatrix} \sim \text{NID}(\mathbf{0}, \mathbf{G}), \quad \mathbf{G} = \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix},$$

and the vector of within-mare errors, ε_i , follows a zero-mean ARMA(1,1) with covariance matrix $\Sigma_i := \Sigma(\phi, \theta, \sigma_a^2)$ for each $i = 1, \dots, 11$, specified by the recursions $\phi(B)\varepsilon_{i,t} = \theta(B)a_{i,t}$ with $a_{i,t} \sim \text{NID}(0, \sigma_a^2)$. This is easily seen to be a LMM in the form of (11.26), with

$$\mathbf{x}_{i,t}^T = [1, \sin(2\pi t), \cos(2\pi t)], \quad \text{and} \quad \mathbf{z}_{i,t}^T = [1, \sin(2\pi t)].$$

In the first step of a SS formulation we write $\mathbf{b}_i := \mathbf{b}_{i,t}$ for the state vector, and cast the model in the special case of $\varepsilon_{i,t} = 0$ as

$$y_{i,t} = \mathbf{z}_{i,t}^T \mathbf{b}_{i,t} + \mathbf{x}_{i,t}^T \boldsymbol{\beta}, \quad \text{and} \quad \mathbf{b}_{i,t} = \mathbf{I}_2 \mathbf{b}_{i,t-1},$$

with initial condition, $\mathbf{b}_{i,0} \sim \text{N}(\mathbf{0}, \mathbf{G})$. In the second step, using \mathbf{H} and Φ to denote the observation and state transition matrices (defined below), we appeal to a SS representation for ARIMA models in the special case of $\mathbf{b}_{i,t} = 0$, and write

$$y_{i,t} = \underbrace{[\theta, 1]}_{\mathbf{H}} \begin{bmatrix} \eta_{i,t-1} \\ \eta_{i,t} \end{bmatrix} + \mathbf{x}_{i,t}^T \boldsymbol{\beta}, \quad \text{and} \quad \begin{bmatrix} \eta_{i,t-1} \\ \eta_{i,t} \end{bmatrix} = \underbrace{\begin{bmatrix} 0 & 1 \\ 0 & \phi \end{bmatrix}}_{\Phi} \begin{bmatrix} \eta_{i,t-2} \\ \eta_{i,t-1} \end{bmatrix} + \begin{bmatrix} 0 \\ a_{i,t} \end{bmatrix},$$

with initial condition,

$$\begin{bmatrix} \eta_{i,-1} \\ \eta_{i,0} \end{bmatrix} \sim N(\mathbf{0}, \mathbf{V}), \quad \mathbf{V} = \frac{\sigma_a^2}{1-\phi^2} \begin{bmatrix} 1 & \phi \\ \phi & 1 \end{bmatrix}.$$

The SS representation we chose here is outlined in Brockwell and Davis [32, ch. 12], and ensures there is no observational noise process; but in principle any such representation could be used (e.g., the ones discussed in Chapter 2). Putting these two steps together, we create the combined 4-dimensional state vector $\mathbf{s}_{i,t} = [\mathbf{b}_{i,t}^T, \eta_{i,t-1}, \eta_{i,t}]^T$, leading to the observation and state equations

$$\mathbf{y}_{i,t} = [\mathbf{z}_{i,t}^T, \mathbf{H}] \mathbf{s}_{i,t} + \mathbf{x}_{i,t}^T \boldsymbol{\beta}, \quad \text{and} \quad \mathbf{s}_{i,t} = \begin{bmatrix} \mathbf{I}_2 & \mathbf{0} \\ \mathbf{0} & \Phi \end{bmatrix} \mathbf{s}_{i,t-1} + \underbrace{\begin{bmatrix} \mathbf{0} \\ a_{i,t} \end{bmatrix}}_{\mathbf{w}_{i,t}},$$

where the state noise vector and state initial conditions are given by,

$$\mathbf{w}_{i,t} \sim NID\left(\mathbf{0}, \begin{bmatrix} \mathbf{0} & 0 \\ 0 & \sigma_a^2 \end{bmatrix}\right), \quad \text{and} \quad \mathbf{s}_{i,0} \sim N\left(\mathbf{0}, \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{V} \end{bmatrix}\right).$$

Fitting the LMM via the `lme` function of R [186], yields the maximum likelihood estimates

$$\begin{aligned} \hat{\beta}_0 &= 12.13, & \hat{\sigma}_0^2 &= 6.034, & \hat{\phi} &= 0.7855, \\ \hat{\beta}_1 &= -2.921, & \hat{\sigma}_1^2 &= 0.7263, & \hat{\theta} &= -0.2835, \\ \hat{\beta}_2 &= -0.8486, & \hat{\sigma}_\varepsilon^2 &= 13.72, \end{aligned}$$

where $\hat{\sigma}_\varepsilon^2$ is the estimated variance of the ARMA(1,1) model for the (unobserved) $\{\varepsilon_{i,t}\}$ process, from which the estimate of the corresponding white noise variance is obtained from the ARMA(1,1) model variance equation:

$$\hat{\sigma}_a^2 = \hat{\sigma}_\varepsilon^2 \left[1 + \frac{(\hat{\theta} + \hat{\phi})^2}{1 - \hat{\phi}^2} \right]^{-1} = 8.7263.$$

Taking these estimates to be the (true) model, and employing the KF and recursions for 12-step ahead forecasting (11.7)–(11.15), results in the predictions and 95% prediction bands for the 11th mare ($i = 11$) displayed in Figure 11.1.

11.6.2 Smoothing and prediction of missing values for the beluga whales data

The data file `whale.dat` in Chatterjee, Handcock and Simonoff [51] gives the values of several time series variables pertaining to the nursing behavior of two newborn captive beluga whale calves, Hudson and Casey. The variables recorded for each animal are as follows.

- **period:** The (consecutive) time period since birth, each period corresponding to a 6 hour time interval. This is the *time* variable, and will be designated by t .

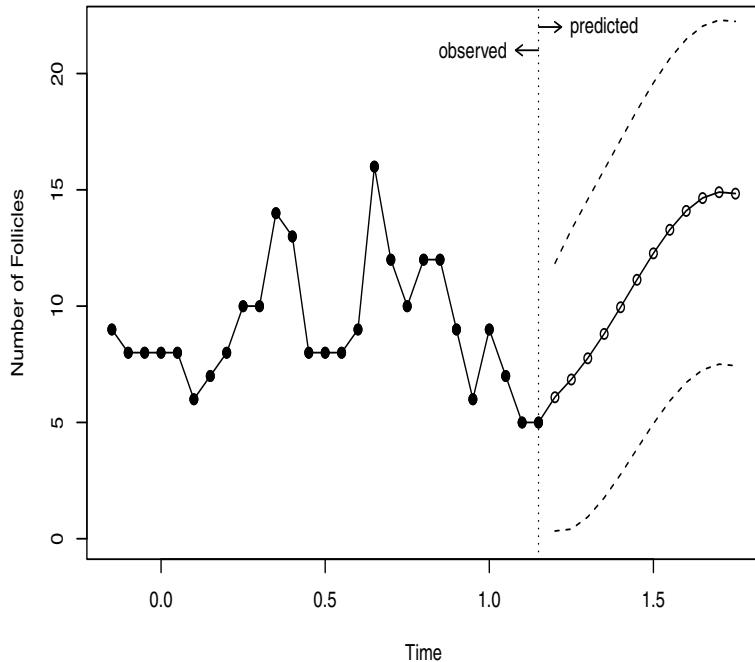


Figure 11.1 *Observed and predicted time series of ovarian follicles for the 11th mare in dataset Ovary of Section 11.6.1. The 12-step ahead predictions (open circles) and corresponding 95% prediction bands (dashed lines), are obtained by running the Kalman recursions for the SS formulation of the LMM described for the series.*

- **nursing:** Total time (in seconds) spent nursing in the period. This is the *response* variable, and will be designated by y_t .
- **bouts:** Number of nursing bouts that occurred in the period. A nursing bout is defined as a successful nursing episode where milk was obtained from the mother.
- **lockons:** Number of lockons in period. A lockon occurs when the calf attaches itself to the mother's mammary gland while suckling milk.
- **daytime:** Indicator variable indicating whether the period was day (1), or night (0). Day is defined as 8AM-8PM, and night 8PM-8AM.

The calves were observed over $n = 228$ consecutive periods, each period 6 hours long, totaling approximately 57 days. However, Casey was only observed up to period 223. The interesting question then arises of how to appropriately “fill in” the remaining 5 (missing) values for Casey, while “borrowing strength” from a com-

bined analysis of both calves. Since the animals were born to different mothers at different times, it makes sense to treat them as independent subjects.

Chatterjee, Handcock and Simonoff [51] did not attempt a longitudinal analysis, but merely explored an ARMAX model for Hudson in order to predict *nursing* as a function of the remaining variables, eventually arriving at

$$y_t = [\beta_1, \dots, \beta_4] \mathbf{u}_t + \phi y_{t-1} + v_t, \quad \{v_t\} \sim \text{NID}(0, \sigma^2), \quad (11.55)$$

where the exogenous covariate vector $\mathbf{u}_t^T = [1, t, \text{lockons}_t, \text{lockons}_{t-1}]$ allows for a linear time trend and the instantaneous and lagged effects of the *lockons* variable. The remaining covariates, instantaneous and at lag one, were not found to be statistically significant. Since our goal is to handle the missingness in *nursing*, *bouts*, and *lockons* for Casey, we entertain the following SS representation of the above ARMAX-model-based longitudinal framework for both Hudson ($i = 1$) and Casey ($i = 2$),

$$x_{i,t}^a = \Phi^a x_{i,t-1}^a + \Gamma \mathbf{u}_{i,t-1} + \mathbf{w}_{i,t}^a, \quad \{\mathbf{w}_{i,t}^a\} \sim \text{NID}(0, Q^a), \quad (11.56)$$

$$y_{i,t} = H_{i,t}^a x_{i,t}^a + \mathbf{v}_{i,t}^a, \quad \{\mathbf{v}_{i,t}^a\} \sim \text{NID}(0, R^a), \quad (11.57)$$

with $\mathbf{u}_{i,t}^T = [1, t, \text{bouts}_{i,t}, \text{lockons}_{i,t}]$ as the chosen exogenous covariate vector, and initial state distribution $x_{i,0} \sim \text{N}(\mu^a, P^a)$. Note that both observation and state variables are one-dimensional, and that we have omitted the covariates in the observation equation in order to render the resulting system in the ARMAX form of model (11.55) for each i . We must also allow for correlation between the state and observation error processes, so that we have additionally,

$$\mathbf{w}_{i,t}^a = \Phi^a \mathbf{v}_{i,t-1}^a, \quad \Rightarrow \quad S^a := \text{cov}(\mathbf{v}_{i,t}^a, \mathbf{w}_{i,t+1}^a) = \text{cov}(\mathbf{v}_{i,t}^a, \Phi^a \mathbf{v}_{i,t}^a) = \Phi^a R^a. \quad (11.58)$$

In order to provide a mechanism for imputing the missing values, we introduce the additional SS model defined by (11.40)–(11.41) as described in Subsection 11.5.2, and note that both observation and state variables in the latter are of dimension 4. Combining the two models, leads to the overall (5-dimensional) state equation,

$$\mathbf{x}_{i,t} \equiv \begin{bmatrix} x_{i,t}^a \\ \mathbf{x}_{i,t}^b \end{bmatrix} = \underbrace{\begin{bmatrix} \Phi^a & \Gamma H_{i,t}^a \\ \mathbf{0} & \Phi^b \end{bmatrix}}_{\Phi} \begin{bmatrix} x_{i,t-1}^a \\ \mathbf{x}_{i,t-1}^b \end{bmatrix} + \mathbf{w}_{i,t}, \quad \{\mathbf{w}_{i,t}\} \sim \text{NID}(0, Q), \quad (11.59)$$

and the overall (5-dimensional) observation equation,

$$\mathbf{z}_{i,t} \equiv \begin{bmatrix} y_{i,t} \\ \mathbf{u}_{i,t} \end{bmatrix} = \begin{bmatrix} H_{i,t}^a & \mathbf{0} \\ \mathbf{0} & \mathbf{H}_{i,t}^b \end{bmatrix} \begin{bmatrix} x_{i,t}^a \\ \mathbf{x}_{i,t}^b \end{bmatrix} + \mathbf{v}_{i,t}, \quad \{\mathbf{v}_{i,t}\} \sim \text{NID}(0, R), \quad (11.60)$$

with the following values for the transition matrices,

$$H_{i,t}^a = \begin{cases} 1, & \text{if } y_{i,t} \text{ is observed,} \\ 0, & \text{if } y_{i,t} \text{ is missing,} \end{cases} \quad \text{and} \quad \mathbf{H}_{i,t}^b = \begin{cases} I_4, & \text{if } \mathbf{u}_{i,t} \text{ is observed,} \\ \mathbf{0}, & \text{if } \mathbf{u}_{i,t} \text{ is missing,} \end{cases}$$

and initial state distribution, $x_{i,0} \sim N(\mu, P)$.

The EM algorithm of Subsection 11.5.1 can now be used to estimate the entire set of parameters, $\Theta = \{\mu, P, \Phi, Q, R\}$, whence the estimates for ARMAX model (11.56)–(11.57) can be reconstructed, e.g., $\Gamma = \Phi(1,2 : 5)$ is obtained by selecting columns 2 through 5 in the first row of Φ , etc. The high-dimensionality of Θ , 75 parameters in total, makes the determination of initial values for this exercise a daunting task! A structured approach was attempted by fitting separate ARMAX models for each whale, estimating parameters via least squares in a regression setting, and then averaging out parameter values for the two models. This did not however, seem to work as well as the simple strategy of starting with values close to zero for all parameters. Table 11.1 displays the final parameter estimates, as well as the values used to initialized the EM algorithm (square brackets). The procedure converged in 45 iterations to within a tolerance of 10^{-4} (as measured by the absolute relative difference in the log-likelihood between successive iterations) with a final log-likelihood value of -4908.3 .

Table 11.1 Final parameter estimates for the SS model fitted to the beluga whale data of Subsection 11.6.2. Estimates were obtained via the EM algorithm initialized with the values in square brackets. The final estimate of the initial state mean was $\hat{\mu}^T = [3.296, 1.009, 1.390, -8.848 \times 10^{-3}, 6.653 \times 10^{-1}]$, starting from the zero vector.

i	j	$\hat{\Phi}(i,j)$	$\hat{P}(i,j)$	$\hat{Q}(i,j)$	$\hat{R}(i,j)$
1	1	1.100 [1.0]	8.780×10^{-2} [0.1]	5.881×10^3 [0.1]	4.156×10^3 [0.1]
1	2	1.098×10^2 [0.0]	-7.742×10^{-6} [0.0]	-4.242×10^{-1} [0.0]	0.000 [0.0]
1	3	5.684×10^{-2} [0.0]	1.462×10^{-4} [0.0]	1.301 [0.0]	0.000 [0.0]
1	4	-1.011×10^1 [0.0]	8.191×10^{-4} [0.0]	1.850×10^2 [0.0]	0.000 [0.0]
1	5	-1.877 [0.0]	1.599×10^{-3} [0.0]	5.910×10^2 [0.0]	0.000 [0.0]
2	1	-9.467×10^{-6} [0.0]	=	=	=
2	2	9.912×10^{-1} [1.0]	2.068×10^{-5} [0.1]	1.172×10^{-4} [0.1]	1.149×10^{-2} [0.1]
2	3	-5.914×10^{-6} [0.0]	-2.740×10^{-4} [0.0]	-1.260×10^{-4} [0.0]	0.000 [0.0]
2	4	8.238×10^{-4} [0.0]	-1.810×10^{-5} [0.0]	-1.471×10^{-2} [0.0]	0.000 [0.0]
2	5	1.656×10^{-4} [0.0]	5.752×10^{-6} [0.0]	-4.572×10^{-2} [0.0]	0.000 [0.0]
3	1	-1.149×10^{-4} [0.0]	=	=	=
3	2	8.772×10^{-1} [0.0]	=	=	=
3	3	9.999×10^{-1} [1.0]	2.420×10^{-2} [0.1]	1.985×10^{-2} [0.1]	5.622×10^2 [0.1]
3	4	1.983×10^{-2} [0.0]	-3.229×10^{-3} [0.0]	5.817×10^{-2} [0.0]	0.000 [0.0]
3	5	-1.986×10^{-4} [0.0]	-8.867×10^{-4} [0.0]	1.739×10^{-1} [0.0]	0.000 [0.0]
4	1	5.695×10^{-3} [0.0]	=	=	=
4	2	4.141 [0.0]	=	=	=
4	3	3.697×10^{-3} [0.0]	=	=	=
4	4	5.820×10^{-1} [1.0]	1.919×10^{-2} [0.1]	6.892 [0.1]	2.539 [0.1]
4	5	-8.297×10^{-2} [0.0]	1.505×10^{-2} [0.0]	2.123×10^1 [0.0]	0.000 [0.0]
5	1	1.871×10^{-2} [0.0]	=	=	=
5	2	1.172×10^1 [0.0]	=	=	=
5	3	1.175×10^{-2} [0.0]	=	=	=
5	4	-1.060 [0.0]	=	=	=
5	5	6.726×10^{-1} [1.0]	6.739×10^{-2} [0.1]	6.791×10^1 [0.1]	1.158×10^1 [0.1]

Figure 11.2 displays the results of applying Kalman smoothing (period 1 to 223) and prediction (period 224 to 228) for Casey, based on the fitted model. Smoothed

values (solid lines) and predictions (open circles) are shown for nursing time, number of bouts, and number of lockons, along with appropriate 95% confidence bands. In order not to overly compress the display, the plots only span the time frame from period 100 onward. The actual predictions and 95% confidence bounds for Casey's 5 missing values are tabulated in Table 11.2.

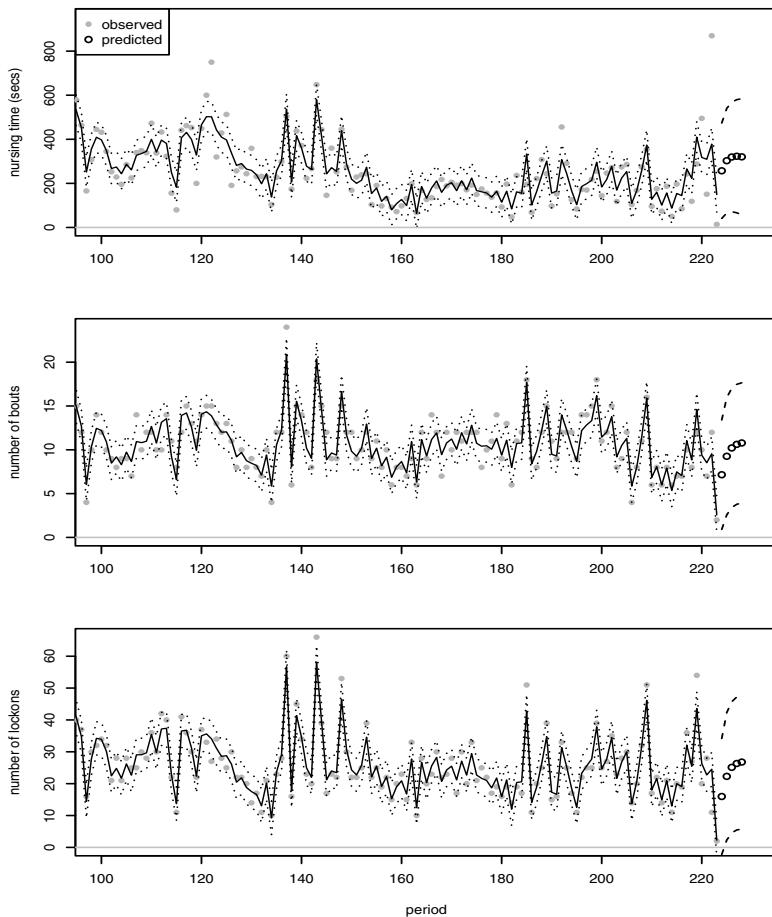


Figure 11.2 *Kalman smoothing and prediction for beluga whale Casey in the whale.dat dataset of Subsection 11.6.2. Shown are smoothed values (solid lines) and 5-step ahead predictions (open circles) for nursing time (top panel), number of bouts (middle panel), and number of lockons (bottom panel), from period 100 onward. Corresponding 95% confidence bands for the smoothed and predicted values are shown as dotted and dashed lines, respectively.*

In conclusion, a few remarks are in order concerning the complexity of longitudinal modeling via SS models, that this particular example highlights.

Table 11.2 Five-step ahead predictions and 95% confidence bounds for Casey in the beluga whale data of Subsection 11.6.2.

Variable	$t = 224$	$t = 225$	$t = 226$	$t = 227$	$t = 228$
nursing	257.3 (40.9, 473.8)	303.1 (65.9, 540.2)	319.8 (71.4, 568.2)	323.3 (66.7, 579.9)	320.6 (57.3, 584.0)
bouts	7.2 (0.9, 13.4)	9.3 (2.6, 16.0)	10.3 (3.4, 17.0)	10.6 (3.8, 17.5)	10.8 (3.9, 17.7)
lockons	16.0 (-2.2, 34.2)	22.3 (2.3, 42.3)	25.1 (4.6, 45.7)	26.3 (5.5, 47.2)	26.8 (5.7, 47.8)

- (i) Properly speaking, the ML estimates in Table 11.1 are not genuine ML estimates, since we did not enforce the constraint implied by (11.59) whereby the first column of Φ below the first entry should be set to zero. To do so would require the development of a constrained version of EM, or application of one of the constrained estimation techniques discussed in Chapter 7, Section 7.3. However, a glance at the estimates for Φ in Table 11.1 reveals that the relevant entries are all less than 1.5×10^{-2} in absolute value, and thus the difference from running a constrained version of ML estimation is likely to be negligible.
- (ii) A second issue we have side-stepped in the estimation of Φ is the fact that 4 entries in the first row are in essence time-varying, but since this involves only the 5 missing values for Casey, a proportion of only 5/456 of the total sample size used, the effect is likewise negligibly small. This time-varying nature was however, accounted for in all the Kalman filtering steps.
- (iii) Finally, we can infer the (obvious) parameter estimates for the original ARMAX model (11.56)–(11.58),

$$\hat{\Phi}^a = 1.100, \quad \hat{\mu}^a = 3.296, \quad \hat{P}^a = 8.780 \times 10^{-2},$$

and

$$\hat{\Gamma} = [109.8 \quad 5.684 \times 10^{-2} \quad -10.11 \quad -1.877].$$

The error variance/covariances are not so straightforward. Note first that from (11.44)–(11.45) one has

$$\mathbf{w}_{i,t} = \begin{bmatrix} \mathbf{\Gamma} \mathbf{v}_{i,t-1}^b + w_{i,t}^a \\ w_{i,t}^b \end{bmatrix}, \quad \text{and} \quad \mathbf{v}_{i,t} = \begin{bmatrix} v_{i,t-1}^a \\ v_{i,t}^b \end{bmatrix},$$

which implies the following covariance structure in the entries of $\mathbf{Q} = \text{var}(\mathbf{w}_{i,t})$,

$$\begin{aligned} \mathbf{Q}(1,1) &= \text{var}(\mathbf{\Gamma} \mathbf{v}_{i,t-1}^b + w_{i,t}^a) = \mathbf{\Gamma} \mathbf{R}^b \mathbf{\Gamma}^T + Q^a, \\ \mathbf{Q}(2:5, 2:5) &= \text{var}(\mathbf{w}_{i,t}^b) = \mathbf{Q}^b, \end{aligned}$$

and

$$\begin{aligned} \mathbf{Q}(1,2:5) &= \text{cov}(\Gamma \mathbf{v}_{i,t-1}^b + \mathbf{w}_{i,t}^a, \mathbf{w}_{i,t}^b) \\ &= \Gamma \text{cov}(\mathbf{v}_{i,t-1}^b, \mathbf{w}_{i,t}^b) + \underbrace{\Gamma \text{cov}(\mathbf{w}_{i,t}^a, \mathbf{w}_{i,t}^b)}_{\mathbf{0}} \\ &= \Gamma (\mathbf{S}^b)^T, \end{aligned}$$

where $\mathbf{S}^b := \text{cov}(\mathbf{w}_{i,t}^b, \mathbf{v}_{i,t-1}^b) = \text{cov}(\mathbf{v}_{i,t-1}^b, \mathbf{w}_{i,t}^b)^T$ denotes the correlation between the state and observation noise processes in the imputation model. The corresponding result for $\mathbf{R} = \text{var}(\mathbf{v}_{i,t})$ is simpler,

$$\mathbf{R} = \begin{bmatrix} \text{var}(\mathbf{v}_{i,t}^a) & \text{cov}(\mathbf{v}_{i,t}^a, \mathbf{v}_{i,t}^b) \\ \mathbf{0} & \text{var}(\mathbf{v}_{i,t}^b) \end{bmatrix} = \begin{bmatrix} \mathbf{R}^a & \mathbf{0} \\ \mathbf{0} & \mathbf{R}^b \end{bmatrix},$$

whence we obtain $\widehat{\mathbf{R}}^a = 4.156 \times 10^3$ and $\widehat{\mathbf{R}}^b = \widehat{\mathbf{R}}(2:5, 2:5)$. Armed with these results we can finally retrieve the estimates for \mathbf{Q}^a ,

$$\widehat{\mathbf{Q}}^a = \widehat{\mathbf{Q}}(1,1) - \widehat{\Gamma} \widehat{\mathbf{R}}^b \widehat{\Gamma}^T = 5440.3,$$

and for \mathbf{S}^a ,

$$\widehat{\mathbf{S}}^a = \widehat{\Phi}^a \widehat{\mathbf{R}}^a = 4572.8.$$

The fact that $\widehat{\mathbf{Q}}(1,2:5) \neq \mathbf{0}$ automatically implies that $\widehat{\mathbf{S}}^b \neq \mathbf{0}$, which suggests the presence of correlation between the state and observation noise processes in the fitted imputation model. However, throughout this analysis we have ignored the fact that the noise and observational error processes may be correlated in the overall model (11.59)–(11.60), and assumed $\mathbf{S} = \mathbf{0}$ in the relevant estimation and Kalman filtering steps. A more sophisticated analysis should not make this simplification, since the above estimates suggest, in particular, that the relevant correlation between $\mathbf{v}_{i,t}^a$ and $\mathbf{v}_{i,t+1}^a$ is of the order of 0.9.

This page intentionally left blank

Appendices

This page intentionally left blank

Appendix A

Some results in numerical algebra and linear systems

A.1	QR Decomposition	193
A.2	Schur decomposition	195
A.3	The Hessenberg Form	196
A.4	SVD Decomposition	197
A.5	Canonical Correlations	198
A.6	Algebraic Lyapunov and Sylvester equations	198
A.7	Numerical solution of a Sylvester equation	200
A.8	Block-diagonalization of a matrix	201
A.9	Reduced rank least squares	202
A.10	Riccati equations	203
A.10.1	Definition	204
A.10.2	Solving the ARE in the general case	205
A.10.3	Solving the ARE for GARCH models	207
A.11	Kalman filter	208

This Appendix provides a summary of several important results in numerical algebra and linear systems theory which are used liberally throughout the book. While this may be helpful in making the book more “self-contained,” it is not meant to provide a complete treatment of this material, so the interested reader may need to consult additional resources.

A.1 QR Decomposition

Any real matrix $\mathbf{A}(m \times n)$, with $m \geq n$, can be factored as:

$$\mathbf{A} = \mathbf{Q} \mathbf{R}, \tag{A.1}$$

where the matrix $\mathbf{Q}(m \times n)$ is orthonormal, and $\mathbf{R}(n \times n)$ is an upper triangular matrix. This common factorization is known as the *QR decomposition*.

The numerical determination of \mathbf{Q} and \mathbf{R} is complicated, and requires additional discussion concerning *triangularization*. Any matrix \mathbf{A} , with dimension $m \times n$ and $m \leq n$, can be reduced to an upper triangular matrix by applying a sequence of orthogonal transformations, such as the *Givens rotations* or the *Householder reflections* (see [60]). A Householder reflection is an orthonormal and symmetric matrix of the form:

$$\mathbf{U} = \mathbf{I} - 2 \frac{\mathbf{u}\mathbf{u}^T}{\mathbf{u}^T\mathbf{u}}. \quad (\text{A.2})$$

Given a vector \mathbf{x} , by defining $\mathbf{u} = \mathbf{x} + h\mathbf{e}$, with $h = \text{sign}(x_1)\sqrt{\mathbf{x}^T\mathbf{x}}$ and $\mathbf{e} = [1 \ 0 \ \dots \ 0]^T$, then $\mathbf{U}\mathbf{x} = -h\mathbf{e}$. The algorithm for the triangularization of the matrix \mathbf{A} by elemental reflection (known as *Householder reduction*), can be described by the following sequence of steps:

1. Build the reflection matrix \mathbf{U}_1 , such that $\mathbf{U}_1\mathbf{a}_1 = [r_{11} \ 0 \ \dots \ 0]^T$, where \mathbf{a}_1 denotes the first column of \mathbf{A} . Then, define \mathbf{A}_2 with the following structure:

$$\mathbf{A}_2 = \begin{bmatrix} r_{11} & r_{12} & \dots & r_{1n} \\ 0 & & & \\ \vdots & & \bar{\mathbf{A}}_2 & \\ 0 & & & \end{bmatrix}. \quad (\text{A.3})$$

2. For all $i = 2, \dots, n$:

- (a) Build the reflection matrix $\bar{\mathbf{U}}_i$ corresponding to the first column of the matrix $\bar{\mathbf{A}}_i(\bar{\mathbf{a}}_{i1})$, such that: $\bar{\mathbf{U}}_i\bar{\mathbf{a}}_1 = [r_{ii} \ 0 \ \dots \ 0]^T$.
- (b) Calculate $\mathbf{A}_{i+1} = \mathbf{U}_i\mathbf{A}_i$, where $\mathbf{U}_i = \text{diag}(\mathbf{I}_{i-1}, \bar{\mathbf{U}}_i)$, and define $\bar{\mathbf{A}}_{i+1}$ as the block made up of the last $m-i$ rows and $n-i$ columns of \mathbf{A}_{i+1} .

After n iterations, we obtain the matrix $\mathbf{A}_{n+1} = \mathbf{U}_n\mathbf{U}_{n-1}\dots\mathbf{U}_1\mathbf{A}$, which exhibits the structure:

$$\mathbf{A}_{n+1} = \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}, \quad (\text{A.4})$$

with \mathbf{R} an upper triangular matrix. By defining the matrix $\mathbf{Q} = (\mathbf{U}_n\mathbf{U}_{n-1}\dots\mathbf{U}_1)^T$, it is straightforward to see that $\mathbf{Q}^T\mathbf{Q} = \mathbf{Q}\mathbf{Q}^T = \mathbf{I}$, and premultiplying (A.4) by \mathbf{Q} yields:

$$\mathbf{A} = \mathbf{Q}\mathbf{A}_{n+1} = \mathbf{Q} \begin{bmatrix} \mathbf{R} \\ \mathbf{0} \end{bmatrix}. \quad (\text{A.5})$$

Moreover, if \mathbf{Q} is partitioned as $\mathbf{Q} = [\mathbf{Q}_1 \ \mathbf{Q}_2]$, \mathbf{A} can be written as $\mathbf{A} = \mathbf{Q}_1\mathbf{R}$, where $\mathbf{Q}_1^T\mathbf{Q}_1 = \mathbf{I}_m$.

The QR decomposition is stable numerically no matter the type of orthogonal transformation employed (*Givens* or *Householder*), although the Householder reflection is more efficient computationally; see [60]. In particular, the algorithm described above has a computational load of $n^2(m - \frac{1}{3}n)$ flops, in contrast with the $n^2(m - \frac{1}{3}n)$ flops needed when using the Givens transformations. The subspace algorithms described in Chapter 7 often use this decomposition.

A.2 Schur decomposition

Let \mathbb{C}^n denote the complex-valued n -space, and $\mathbb{C}^{n \times n} \equiv \mathbb{C}^n \times \mathbb{C}^n$ the crossproduct space of \mathbb{C}^n with itself. Similarly, denote by $\mathbb{R}^{n \times n}$ the corresponding crossproduct space of real numbers. Any matrix $\mathbf{A} \in \mathbb{C}^{n \times n}$ can be factored as:

$$\mathbf{A} = \mathbf{U} \mathbf{D} \mathbf{U}^H, \quad (\text{A.6})$$

where $\mathbf{U} \in \mathbb{C}^{n \times n}$ is a unitary matrix, \mathbf{U}^H denotes the conjugate transpose of \mathbf{U} , and \mathbf{D} is an upper triangular matrix known as the *complex Schur form* of \mathbf{A} , whose diagonal elements correspond to the eigenvalues of \mathbf{A} .

The importance of this decomposition is due to the fact that any matrix can be reduced to its Schur form by means of a succession of similar unitary transformations, so this reduction provides a numerically stable method to calculate eigenvalues; see Golub and Van Loan [109] and Datta [60].

If \mathbf{A} has complex eigenvalues, matrices \mathbf{U} and \mathbf{D} will be complex. In this case, an alternative procedure to compute (A.6) is the so-called *real Schur decomposition*, which is noteworthy due to its avoidance of operations with complex numbers. The real Schur decomposition of $\mathbf{A} \in \mathbb{R}^{n \times n}$ is given by:

$$\mathbf{A} = \mathbf{W} \mathbf{R} \mathbf{W}^T, \quad (\text{A.7})$$

where $\mathbf{W} \in \mathbb{R}^{n \times n}$ is a orthonormal matrix and $\mathbf{R} \in \mathbb{R}^{n \times n}$ is an upper Hessenberg matrix (called the *real Schur form* of \mathbf{A}) with the following structure:

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \dots & \mathbf{R}_{1k} \\ \mathbf{0} & \mathbf{R}_{22} & \dots & \mathbf{R}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_{kk} \end{bmatrix}, \quad (\text{A.8})$$

where $\mathbf{R}_{11}, \mathbf{R}_{22}, \dots, \mathbf{R}_{kk}$ are either scalars or (2×2) matrices with complex conjugate eigenvalues. The eigenvalues of \mathbf{A} coincide with those of the matrices $\mathbf{R}_{11}, \mathbf{R}_{22}, \dots, \mathbf{R}_{kk}$, so that k equals the number of real eigenvalues of \mathbf{A} plus the number of pairs of complex conjugate eigenvalues of \mathbf{A} . For a proof of this result, see Golub and Van Loan [109].

From a computational point of view, the reduction of a real Schur form should be performed before obtaining the complex Schur form. The transformation of the real Schur form to the complex representation only requires applying the complex Schur decomposition to each of the \mathbf{R}_{ii} blocks in (A.8).

Computing the Schur form of a matrix requires iterations. A detailed description of the available algorithms for real and complex Schur forms can be found in Golub and Van Loan [109] and Datta [60]. In general, these procedures include two steps:

Step 1) Reduce the matrix \mathbf{A} to an upper Hessenberg form. This consists in computing the factorization $\mathbf{A} = \mathbf{U} \mathbf{H} \mathbf{U}^T$, where \mathbf{H} is the upper Hessenberg matrix and \mathbf{U} is an orthonormal matrix.

Step 2) QR iterations. With the initial condition $\mathbf{A}_0 = \mathbf{A}$, the iterations:

$$\mathbf{A}_i = \mathbf{Q}_i \mathbf{R}_i, \quad (\text{A.9})$$

and

$$\mathbf{A}_{i+1} = \mathbf{R}_i \mathbf{Q}_i = \mathbf{Q}_i^T \mathbf{A}_i \mathbf{Q}_i, \quad (\text{A.10})$$

with $i = 0, 1, 2, \dots$, generate a sequence \mathbf{A}_i of similar matrices which converges to a real Schur form under non restrictive conditions. If $\mathbf{A}_0 = \mathbf{H}$ and \mathbf{H} is an upper unreduced Hessenberg form, then all the matrices \mathbf{A}_i are Hessenberg forms. This has a significant computational advantage, because the computational burden in flops of determining the QR decomposition of a Hessenberg matrix is proportional to n^2 , which is smaller than the cost of decomposing an arbitrary matrix (proportional to n^3).

In some cases it may be desirable for the Schur form to be a lower Hessenberg matrix. This requires applying the Schur decomposition only to the transpose of \mathbf{A} . Then, $\mathbf{A} = \mathbf{W} \mathbf{R}^T \mathbf{W}^T$, where \mathbf{R}^T is a lower Hessenberg matrix.

A.3 The Hessenberg Form

Any matrix $\mathbf{A}(n \times n)$, can be reduced to an upper Schur form by applying a finite sequence of orthogonal transformations such that \mathbf{A} is factored as:

$$\mathbf{A} = \mathbf{U}^T \mathbf{H} \mathbf{U}, \quad (\text{A.11})$$

where $\mathbf{U}(n \times n)$ is an orthonormal matrix, and \mathbf{H} is an upper Hessenberg matrix.

The procedure to reduce a Hessenberg matrix can be divided into the following steps:

Step 1) Build a Householder reflection matrix $\bar{\mathbf{U}}_1$ such that:

$$\bar{\mathbf{U}}_1 \begin{bmatrix} a_{21} \\ a_{31} \\ \vdots \\ a_{n1} \end{bmatrix} = \begin{bmatrix} h_{21} \\ 0 \\ \vdots \\ 0 \end{bmatrix}, \quad (\text{A.12})$$

and define $\mathbf{U}_1 = \text{diag}(1, \bar{\mathbf{U}}_1)$, where \mathbf{U}_1 is such that $\mathbf{A}_2 = \mathbf{U}_1 \mathbf{A} \mathbf{U}_1^T$, being \mathbf{A}_2 a matrix with the following structure:

$$\mathbf{A}_2 = \begin{bmatrix} \times & \times & \dots & \times \\ \times & \times & \dots & \times \\ 0 & \times & \dots & \times \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \times & \dots & \times \end{bmatrix}, \quad (\text{A.13})$$

where “ \times ” represents potential nonzero elements. Now define $\bar{\mathbf{a}}_2 = \mathbf{A}_2(3 : n, 2)$ to be the vector formed from rows 3 to n of the second column of \mathbf{A}_2 .

Step 2) For all $i = 2, \dots, n-2$:

Build the reflection matrix \bar{U}_i such that $\bar{U}_i \bar{a}_i = [\mathbf{h}_{i+1,i} \ 0 \ \dots \ 0]^T$, and then compute $\mathbf{A}_{i+1} = \mathbf{U}_i \mathbf{A}_i \mathbf{U}_i^T$, with $\mathbf{U}_i = \text{diag}(\mathbf{I}_i, \bar{U}_i)$. Now define $\bar{a}_{i+1} = \mathbf{A}_{i+1}(i+2 : n, i+1)$, similar to what was done at the end of Step 1.

The matrix \mathbf{A}_{n-1} has an upper Hessenberg form which is obtained from \mathbf{A} as:

$$\mathbf{A}_{n-1} = \mathbf{U}_{n-2} \mathbf{U}_{n-3} \dots \mathbf{U}_1 \mathbf{A} \mathbf{U}_1^T \dots \mathbf{U}_{n-3}^T \mathbf{U}_{n-2}^T = \mathbf{U} \mathbf{A} \mathbf{U}^T, \quad (\text{A.14})$$

where $\mathbf{U} = \mathbf{U}_{n-2} \mathbf{U}_{n-3} \dots \mathbf{U}_1$. Note that \mathbf{U} is orthonormal, and therefore $\mathbf{A} = \mathbf{U}^T \mathbf{A}_{n-1} \mathbf{U}$.

A.4 SVD Decomposition

The *Singular Value Decomposition* (SVD) is based on the following theorem.

Theorem A.1 For any real matrix $\mathbf{A} \in \mathbb{R}^{m \times n}$, there exist two orthogonal matrices, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_m] \in \mathbb{R}^{m \times m}$ and $\mathbf{V} = [\mathbf{v}_1, \dots, \mathbf{v}_n] \in \mathbb{R}^{n \times n}$, such that $\mathbf{U}^T \mathbf{A} \mathbf{V} = \mathbf{S} = \text{diag}(\sigma_1, \dots, \sigma_p) \in \mathbb{R}^{m \times n}$, with $p = \min\{m, n\}$ and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_p \geq 0$.

Proof. See [109], Theorem 2.5.2.

We denote by σ_i , $i = 1, \dots, p$, the singular values of \mathbf{A} , while the vectors \mathbf{u}_i and \mathbf{v}_i are the i left and right singular vectors of \mathbf{A} , respectively. We may also define the Euclidean norm of \mathbf{A} as $\|\mathbf{A}\|_2 = \sigma_1$, and the Frobenius norm of \mathbf{A} as $\|\mathbf{A}\|_F = \sqrt{\sum_{i=1}^p \sigma_i^2}$. Some of the properties of the SVD used in Chapters 7 and 9 are:

1. Rank and singular values. If $\text{rank}(\mathbf{A}) = k \leq p$ then $\text{rank}(\mathbf{A}) = \text{rank}(\mathbf{USV}^T) = \text{rank}(\mathbf{S}_k) = k$ holds, where $\mathbf{S}_k = \text{diag}(\sigma_1, \dots, \sigma_k)$ and $\sigma_1 \geq \sigma_2 \geq \dots \geq \sigma_k > 0$. Therefore, the rank of \mathbf{A} coincides with the number of nonzero singular values.
2. Moore–Penrose pseudo-inverse. The matrix $\mathbf{A}^+ = \mathbf{V}_k \mathbf{S}_k^{-1} \mathbf{U}_k^T$ is the so-called Moore–Penrose pseudo-inverse of matrix \mathbf{A} , where $\text{rank}(\mathbf{A}) = k \leq p$. The matrix \mathbf{A}^+ satisfies the following properties:
 - (a) $(\mathbf{A}^+)^+ = \mathbf{A}$
 - (b) $\mathbf{A} \mathbf{A}^+ \mathbf{A} = \mathbf{A}$
 - (c) $\mathbf{A}^+ \mathbf{A} \mathbf{A}^+ = \mathbf{A}^+$
 - (d) $(\mathbf{A}^+)^T = (\mathbf{A}^T)^+$
 - (e) $\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^+ \mathbf{A}^T$
 - (f) When $\text{rank}(\mathbf{A}) = n \leq m$, then $\mathbf{A}^+ = (\mathbf{A}^T \mathbf{A})^{-1}$ and $\mathbf{A}^+ \mathbf{A} = \mathbf{I}_n$
3. Low-rank approximation. Assume that $\text{rank}(\mathbf{A}) = r$. For any matrix \mathbf{B} such that $\text{rank}(\mathbf{B}) = k$, the matrix $\mathbf{A}_k = \mathbf{U}_k \mathbf{S}_k \mathbf{V}_k^T$ with $k < r$ and $\text{rank}(\mathbf{A}_k) = k$, satisfies the condition:

$$\|\mathbf{A} - \mathbf{B}\|_2 \geq \|\mathbf{A} - \mathbf{A}_k\|_2, \quad \text{and} \quad \|\mathbf{A} - \mathbf{B}\|_F \geq \|\mathbf{A} - \mathbf{A}_k\|_F. \quad (\text{A.15})$$

This property assures that the SVD is the adequate method to determine the rank of a matrix as the number of singular values that are significantly different from zero.

A.5 Canonical Correlations

Let \mathbf{x} and \mathbf{y} be two vectors of random variables with dimensions m and n , respectively, and satisfying: $m \geq n$, $\text{var}(\mathbf{x}) = \Sigma_{xx} > \mathbf{0}$, $\text{var}(\mathbf{y}) = \Sigma_{yy} > \mathbf{0}$, and $\text{cov}(\mathbf{x}, \mathbf{y}) = \Sigma_{xy}$. The n canonical correlations between \mathbf{x} and \mathbf{y} , denoted by ρ_k , $k = 1, \dots, n$, are defined as:

$$\rho_k(\mathbf{x}, \mathbf{y}) = \max_{\substack{\alpha_k \in \mathbb{R}^m \\ \beta_k \in \mathbb{R}^n}} \text{cov}(\alpha_k^T \mathbf{x}, \beta_k^T \mathbf{y}), \quad (\text{A.16})$$

subject to the conditions that for all $i = 1, \dots, k-1$, we have

$$\text{var}(\alpha_i^T \mathbf{x}) = 1, \quad \text{var}(\beta_i^T \mathbf{y}) = 1, \quad \text{cov}(\alpha_i^T \mathbf{x}, \alpha_i^T \mathbf{x}) = 0, \quad \text{cov}(\beta_i^T \mathbf{y}, \beta_i^T \mathbf{y}) = 0.$$

Obviously, we must have $\rho_1 \geq \rho_2 \geq \dots \geq \rho_n \geq 0$. The first canonical correlation, ρ_1 , is the maximum correlation between any linear combination of \mathbf{x} and \mathbf{y} with unit variance. The vectors $\alpha_1^T \mathbf{x}$ and $\beta_1^T \mathbf{y}$ are the so-called *canonical variables* corresponding to ρ_1 . The second canonical correlation, ρ_2 , represents the maximum correlation between linear combinations of \mathbf{x} and \mathbf{y} that are uncorrelated with the previous canonical variables, and so on.

Proposition A.1 *The canonical correlations between \mathbf{x} and \mathbf{y} correspond to the singular values of the matrix:*

$$\Sigma_{xx}^{-\frac{1}{2}} \Sigma_{xy} \Sigma_{yy}^{-\frac{1}{2}} \stackrel{\text{svd}}{=} \mathbf{U}_n \mathbf{S}_n \mathbf{V}_n^T, \quad (\text{A.17})$$

where $[\alpha_1 \ \alpha_2 \ \dots \ \alpha_n] = \Sigma_{xx}^{-\frac{1}{2}} \mathbf{U}_n$, and $[\beta_1 \ \beta_2 \ \dots \ \beta_n] = \Sigma_{yy}^{-\frac{1}{2}} \mathbf{V}_n$.

Proof. See Casals [38], Proposition A.9.2.

In practice, the information in $\mathbf{X} \in \mathbb{R}^{N \times m}$ and $\mathbf{Y} \in \mathbb{R}^{N \times n}$ is used to estimate the following sample covariance matrices:

$$\hat{\Sigma}_{xx} = \frac{\mathbf{X}^T \mathbf{X}}{N} > 0, \quad \hat{\Sigma}_{yy} = \frac{\mathbf{Y}^T \mathbf{Y}}{N} > 0, \quad \hat{\Sigma}_{xy} = \frac{\mathbf{X}^T \mathbf{Y}}{N}.$$

Therefore, the canonical correlations between \mathbf{X} and \mathbf{Y} coincide with the singular values of $(\mathbf{X}^T \mathbf{X})^{-\frac{1}{2}} \mathbf{X}^T \mathbf{Y} (\mathbf{Y}^T \mathbf{Y})^{-\frac{1}{2}} \stackrel{\text{svd}}{=} \mathbf{U}_n \mathbf{S}_n \mathbf{V}_n^T$.

A.6 Algebraic Lyapunov and Sylvester equations

The so-called algebraic Lyapunov and Sylvester equations often appear in the theory of linear systems and optimal control. The discrete-time Lyapunov equation is given by:

$$\mathbf{X} = \mathbf{A} \mathbf{X} \mathbf{A}^T + \mathbf{C}, \quad (\text{A.18})$$

where \mathbf{A} is a $(n \times n)$ matrix, \mathbf{C} is a $(n \times n)$ symmetric matrix and \mathbf{X} is an unknown $(n \times n)$ symmetric matrix.

On the other hand, the Sylvester equation is given by:

$$\mathbf{A}\mathbf{X} + \mathbf{X}\mathbf{B} = \mathbf{C}, \quad (\text{A.19})$$

where the matrices \mathbf{A} , \mathbf{B} , \mathbf{X} , and \mathbf{C} are $(n \times n)$, $(m \times m)$, $(n \times m)$, and $(n \times m)$, respectively, with \mathbf{A} , \mathbf{B} , and \mathbf{C} unknown matrices. Both equations, (A.18) and (A.19), can be represented as linear systems equations. Applying the $\text{vec}(\cdot)$ operator to (A.18) and (A.19), yields:

$$(\mathbf{I} - \mathbf{A}^T \otimes \mathbf{A})\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{C}), \quad (\text{A.20})$$

and

$$(\mathbf{I}_m \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_n)\text{vec}(\mathbf{X}) = \text{vec}(\mathbf{C}). \quad (\text{A.21})$$

The matrix $(\mathbf{I} - \mathbf{A}^T \otimes \mathbf{A})$ has eigenvalues $1 - \lambda_i \lambda_j$, for all $i = 1, \dots, n$ and $j = 1, \dots, n$, with λ_i denoting the i -th eigenvalue of \mathbf{A} . Thus, system (A.20) has a unique solution if matrix \mathbf{A} has no reciprocal eigenvalues, that is, if $\lambda_i \lambda_j \neq 1$, for all $i = 1, \dots, n$ and $j = 1, \dots, n$. For example, if \mathbf{A} is a transition matrix in a stable dynamic system, then $\lambda_i < 1$, for all $i = 1, \dots, n$, and the condition of reciprocity between eigenvalues is satisfied. Moreover, in this case, if \mathbf{C} is positive semidefinite, then so is the solution to equation (A.18). Writing (A.18) as:

$$\mathbf{X} = \mathbf{A}^N \mathbf{X} (\mathbf{A}^T)^N + \sum_{i=0}^{N-1} \mathbf{A}^i \mathbf{C} (\mathbf{A}^T)^i, \quad (\text{A.22})$$

demonstrates that if the eigenvalues of \mathbf{A} are all strictly within the unit circle, it holds that (see Anderson and Moore [6, ch. 4]):

$$\lim_{N \rightarrow \infty} \mathbf{A}^N \mathbf{X} (\mathbf{A}^T)^N = \mathbf{0}, \quad (\text{A.23})$$

$$\lim_{N \rightarrow \infty} \sum_{i=0}^N \mathbf{A}^i \mathbf{C} (\mathbf{A}^T)^i < \infty, \quad (\text{A.24})$$

and therefore:

$$\mathbf{X} = \sum_{i=0}^{\infty} \mathbf{A}^i \mathbf{C} (\mathbf{A}^T)^i. \quad (\text{A.25})$$

That is, the matrix \mathbf{X} is defined as a convergent infinite sum of positive semidefinite matrices. If $\text{rank}([\mathbf{A}^{n-1} \mathbf{C}^{\frac{1}{2}} \mathbf{A}^{n-2} \mathbf{C}^{\frac{1}{2}} \dots \mathbf{A} \mathbf{C}^{\frac{1}{2}} \mathbf{C}^{\frac{1}{2}}]) = n$, it is easy to see that \mathbf{X} is full rank.

In the Sylvester equation, matrix $(\mathbf{I}_m \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_n)$ has eigenvalues $\lambda_i + \mu_j$, for all $i = 1, \dots, n$ and $j = 1, \dots, m$, where μ_j denotes the j -th eigenvalue of \mathbf{B} . Then, the system equations given by (A.21) have a unique solution if matrices \mathbf{A} and $-\mathbf{B}$ have no common eigenvalues.

Calculating the solutions to (A.18) and (A.19) by solving the systems equations (A.20) and (A.21) is inefficient for large values of n and m . First, because of the computational load involved $(1/6n^6 + n^4)$ flops when performing the Cholesky decomposition of $(\mathbf{I} - \mathbf{A}^T \otimes \mathbf{A})$, and second because of the memory storage required. Note that the matrix $(\mathbf{I} - \mathbf{A}^T \otimes \mathbf{A})$ is $(n^2 \times n^2)$, while $(\mathbf{I}_m \otimes \mathbf{A} + \mathbf{B}^T \otimes \mathbf{I}_n)$ is $(nm \times nm)$. There are, however, computationally more efficient alternative methods based on the reduction of the matrices \mathbf{A} and \mathbf{B} to simpler similar forms.

A.7 Numerical solution of a Sylvester equation

Solving the Sylvester equation (A.19) is far from trivial. It can be done by an algorithm due to Bartels and Stewart [15], or using the *Hessenberg–Schur* procedure ([181])¹. The Bartels and Stewart algorithm proceeds as follows:

Step 1) Reduce \mathbf{A} and \mathbf{B} to lower and upper real Schur forms, respectively, so that $\mathbf{U}^T \mathbf{A} \mathbf{U} = \mathbf{A}^*$ and $\mathbf{V}^T \mathbf{B} \mathbf{V} = \mathbf{B}^*$.

Step 2) Premultiplying (A.19) by \mathbf{U}^T and postmultiplying by \mathbf{V} , yields:

$$\mathbf{U}^T \mathbf{A} \mathbf{U} \mathbf{U}^T \mathbf{X} \mathbf{V} + \mathbf{U}^T \mathbf{X} \mathbf{V} \mathbf{V}^T \mathbf{B} \mathbf{V} = \mathbf{U}^T \mathbf{C} \mathbf{V}. \quad (\text{A.26})$$

Now, defining $\mathbf{X}^* = \mathbf{U}^T \mathbf{X} \mathbf{V}$ and $\mathbf{C}^* = \mathbf{U}^T \mathbf{C} \mathbf{V}$, (A.26) can be written as:

$$\mathbf{A}^* \mathbf{X}^* + \mathbf{X}^* \mathbf{B}^* = \mathbf{C}^*. \quad (\text{A.27})$$

Consider now the following partition of matrices \mathbf{X}^* and \mathbf{C}^* :

$$\mathbf{X}^* = \begin{bmatrix} \mathbf{X}_{11}^* & \mathbf{X}_{1p}^* \\ \vdots & \vdots \\ \mathbf{X}_{k1}^* & \mathbf{X}_{kp}^* \end{bmatrix}, \quad (\text{A.28})$$

and

$$\mathbf{C}^* = \begin{bmatrix} \mathbf{C}_{11}^* & \mathbf{C}_{1p}^* \\ \vdots & \vdots \\ \mathbf{C}_{k1}^* & \mathbf{C}_{kp}^* \end{bmatrix}. \quad (\text{A.29})$$

For all $h = 1, \dots, k$ and $l = 1, \dots, k$, the following equations are then immediately derived from the quasi-triangular form of \mathbf{A}^* and \mathbf{B}^* :

$$\sum_{i=1}^l \mathbf{A}_{li}^* \mathbf{X}_{ih}^* + \sum_{i=1}^h \mathbf{X}_{li}^* \mathbf{B}_{ih}^* = \mathbf{C}_{lh}^*, \quad (\text{A.30})$$

or

$$\mathbf{A}_{ll}^* \mathbf{X}_{lh}^* + \mathbf{X}_{lh}^* \mathbf{B}_{hh}^* = \mathbf{C}_{lh}^* - \sum_{i=1}^{l-1} \mathbf{A}_{li}^* \mathbf{X}_{ih}^* - \sum_{i=1}^{h-1} \mathbf{X}_{li}^* \mathbf{B}_{ih}^*. \quad (\text{A.31})$$

Equations (A.31) are Sylvester equations with dimension (1×1) or (2×2) . If these equations are solved in the order $\mathbf{X}_{11}^*, \mathbf{X}_{21}^*, \dots, \mathbf{X}_{12}^*, \mathbf{X}_{22}^*, \dots, \mathbf{X}_{kp}^*$, then the right-hand side of expression (A.31) does not contain any element of \mathbf{X}^* that has not been previously calculated. The solution of equations (A.31) constitutes Step 2).

Step 3) Calculate the matrix $\mathbf{X} = \mathbf{U} \mathbf{X}^* \mathbf{V}^T$.

¹The main difference between the two methods lies in the fact that Bartels and Stewart [15] reduce the matrix \mathbf{A} to a Schur form, while the Hessenberg–Schur algorithm uses the more efficient Hessenberg form.

The computational load of this algorithm is roughly $13(n^3 + m^3) + 5/2(nm^2 + mn^2)$ flops and it is numerically stable, considering the stability properties of the Schur decomposition. A variant of this algorithm reduces matrices \mathbf{A} and \mathbf{B} to Schur complex forms, such that \mathbf{A}^* and \mathbf{B}^* are lower and upper triangular matrices, respectively. In this case, Step 2 is simplified as all of the equations in (A.31) are (1×1) , but it is necessary to perform complex arithmetic operations.

The procedure to solve the Lyapunov equation (A.18) is very similar to the solution method that we have just described, see Petkov, Christov and Konstantinov [181].

A.8 Block-diagonalization of a matrix

For any real square matrix \mathbf{A} there is a transformation matrix \mathbf{T} , such that $\mathbf{T}^{-1}\mathbf{AT} = \mathbf{R}$, where \mathbf{R} is a real block diagonal matrix $\mathbf{R} = \text{diag}(\mathbf{R}_{11}, \mathbf{R}_{22}, \dots, \mathbf{R}_{kk})$, and blocks $\mathbf{R}_{11}, \mathbf{R}_{22}, \dots, \mathbf{R}_{kk}$ have no common eigenvalues.

The block-diagonalization algorithm is described in Petkov, Christov and Konstantinov [181], and is structured in three steps:

Step 1) Obtain the real Schur transformation of matrix \mathbf{A} :

$$\mathbf{U}^T \mathbf{AU} = \mathbf{R}, \quad (\text{A.32})$$

where

$$\mathbf{R} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} & \dots & \mathbf{R}_{1k} \\ \mathbf{0} & \mathbf{R}_{22} & \dots & \mathbf{R}_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \dots & \mathbf{R}_{kk} \end{bmatrix}, \quad (\text{A.33})$$

with $\mathbf{R}_{11}, \mathbf{R}_{22}, \dots, \mathbf{R}_{kk}$ either scalars or (2×2) matrices with complex conjugate eigenvalues. Transformation matrices \mathbf{T} and \mathbf{T}^{-1} are initialized to: $\mathbf{T} = \mathbf{U}$ and $\mathbf{T}^{-1} = \mathbf{U}^T$.

Step 2) Block diagonalize of the Schur form. We can define a transformation matrix with the following structure:

$$\mathbf{X} = \begin{bmatrix} \mathbf{I} & \mathbf{Z} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad (\text{A.34})$$

and given that:

$$\mathbf{X}^{-1} = \begin{bmatrix} \mathbf{I} & -\mathbf{Z} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}, \quad (\text{A.35})$$

we have:

$$\mathbf{X}^{-1} \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix} \mathbf{X} = \begin{bmatrix} \mathbf{R}_{11} & \mathbf{R}_{11}\mathbf{Z} - \mathbf{Z}\mathbf{R}_{22} + \mathbf{R}_{12} \\ \mathbf{0} & \mathbf{R}_{22} \end{bmatrix}. \quad (\text{A.36})$$

The problem then consists of finding the value of \mathbf{Z} that solves the following Sylvester equation:

$$\mathbf{R}_{11}\mathbf{Z} - \mathbf{Z}\mathbf{R}_{22} + \mathbf{R}_{12} = 0. \quad (\text{A.37})$$

This equation has a unique solution if \mathbf{R}_{11} and \mathbf{R}_{22} do not have common eigenvalues. Once we select a block \mathbf{R}_{ij} to make it zero, we should take into account that the proposed transformation affects blocks $\mathbf{R}_{1:i-1,j}$ and $\mathbf{R}_{i,j+1:k}$, as follows:

$$\mathbf{R}_{1:i-1,j} = \mathbf{R}_{1:i-1,j} + \mathbf{R}_{1:i-1,i} \mathbf{Z}, \quad (\text{A.38})$$

and

$$\mathbf{R}_{i,j+1:k} = \mathbf{R}_{i,j+1:k} - \mathbf{Z} \mathbf{R}_{j,j+1:k}. \quad (\text{A.39})$$

Additionally, the transformation matrix T has to be updated as:

$$\mathbf{T}_{1:k,j} = \mathbf{T}_{1:k,j} + \mathbf{T}_{1:k,i} \mathbf{Z}, \quad (\text{A.40})$$

and

$$(\mathbf{T}^{-1})_{i,1:k} = (\mathbf{T}^{-1})_{i,1:k} - \mathbf{Z} (\mathbf{T}^{-1})_{j,1:k}. \quad (\text{A.41})$$

Then the algorithm consists of starting bottom-to-top and left-to-right: $\mathbf{R}_{k-1,k}, \mathbf{R}_{k-2,k-1}, \mathbf{R}_{k-2,k}, \dots, \mathbf{R}_{1,2}, \dots, \mathbf{R}_{1,k}$, and solving the Sylvester equations for pairs of blocks with no common eigenvalues.

Step 3) Reorder the matrix \mathbf{R} obtained in Step 2. Given that the Schur transformation usually does not provide the eigenvalues ordered in a specific way, we could have common eigenvalues in non-consecutive blocks. For example, if \mathbf{R}_{11} and \mathbf{R}_{33} share the same eigenvalue, the block \mathbf{R}_{13} will not be null. Then, to provide a block diagonal matrix, we have to reorder the rows and columns of \mathbf{R} , while ensuring that common eigenvalues are placed in consecutive blocks. We apply the same reordering operation to both the columns of T and the rows of T^{-1} .

A.9 Reduced rank least squares

Consider the regression model:

$$\mathbf{y}_t = \mathbf{ABx}_t + \boldsymbol{\varepsilon}_t, \quad t = 1, 2, \dots, N, \quad (\text{A.42})$$

where \mathbf{A} is $(j \times n)$ and \mathbf{B} is $(n \times i)$ such that $n < \min\{i, j\}$, and $\boldsymbol{\varepsilon}_t$ is a zero-mean white noise process with $E[\boldsymbol{\varepsilon}_t \boldsymbol{\varepsilon}_t^T] = \boldsymbol{\Omega}$. Therefore, $\mathbf{C} = \mathbf{AB}$ is a reduced-rank matrix. A general representation of (A.42) used in Chapter 7 is:

$$\mathbf{y}_t = \mathbf{ABx}_t + \mathbf{Hu}_t + \boldsymbol{\varepsilon}_t, \quad (\text{A.43})$$

where the rank restriction only affects one part of the model parameters. In the following propositions, we will derive least-squares estimators for \mathbf{A} and \mathbf{B} in model (A.42), and for \mathbf{A} , \mathbf{B} , and \mathbf{H} in model (A.43).

Proposition A.2 *Let $\boldsymbol{\Omega}$, \mathbf{Y} , \mathbf{X} , \mathbf{A} , and \mathbf{B} be matrices with dimensions $(j \times j)$, $(j \times N)$, $(i \times N)$, $(j \times n)$, and $(n \times i)$, respectively, such that $\text{rank}(\boldsymbol{\Omega}) = j$, $\text{rank}(\mathbf{Y}) = j$, and $\text{rank}(\mathbf{X}) = i$. Then, the problem given by:*

$$\min_{\mathbf{A}, \mathbf{B}} \|\boldsymbol{\Omega}^{-1/2}(\mathbf{Y} - \mathbf{ABX})\|_F^2, \quad (\text{A.44})$$

1. Has an infinite number of solutions.
2. May be solved using the SVD decomposition of the matrix:

$$\Omega^{-1/2} \mathbf{Y} \mathbf{X}^T (\mathbf{X} \mathbf{X}^T)^{-1/2} \stackrel{\text{svd}}{=} \mathbf{U} \mathbf{S} \mathbf{V}^T. \quad (\text{A.45})$$

3. $\hat{\mathbf{A}} = \Omega^{1/2} \mathbf{U}_n$ and $\hat{\mathbf{B}} = \mathbf{S}_n \mathbf{V}_n^T (\mathbf{X} \mathbf{X}^T)^{-1/2}$ are optimal solutions, where \mathbf{U}_n and \mathbf{V}_n correspond to the matrices made up of the first n columns from \mathbf{U} and \mathbf{V} , respectively, and \mathbf{S}_n is the diagonal matrix that contains the first n singular values of \mathbf{S} .

Proof. See Casals [38], Proposition A.13.1.

Proposition A.3 Let us assume now that the rank restriction only affects one part of the model parameters, so that the problem (A.44) can be expressed as:

$$\min_{\mathbf{A}, \mathbf{B}, \mathbf{H}} \|\Omega^{-1/2} (\mathbf{Y} - \mathbf{ABX} - \mathbf{HU})\|_F^2, \quad (\text{A.46})$$

where \mathbf{H} is $(j \times r)$, \mathbf{U} is $(r \times N)$, and $\text{rank}(\mathbf{U}) = r$. Then:

1. Matrices $\hat{\mathbf{A}}$ and $\hat{\mathbf{B}}$ arise from applying Proposition A.2 to:

$$\min_{\mathbf{A}, \mathbf{B}} \|\Omega^{-1/2} (\mathbf{Y} \Pi_U^\perp - \mathbf{ABX} \Pi_U^\perp)\|_F^2, \quad (\text{A.47})$$

where $\Pi_U^\perp = \mathbf{I} - \mathbf{U}^T (\mathbf{UU}^T)^{-1} \mathbf{U}$ is the projection onto the orthogonal complement of \mathbf{U} .

2. The optimal estimate for \mathbf{H} is obtained by solving the unrestricted problem:

$$\min_{\mathbf{H}} \|\Omega^{-1/2} (\mathbf{Y} - \hat{\mathbf{A}} \hat{\mathbf{B}} \mathbf{X} - \mathbf{HU})\|_F^2. \quad (\text{A.48})$$

Proof. See Casals [38], Proposition A.13.2.

Under the assumption of Gaussian innovations, setting the weighting matrix Ω to be equal to the variance of the innovations of models (A.42) and (A.43), leads to ML estimates when solving (A.44) and (A.46). However, as Ω is generally unknown, it should be replaced by its estimate $\hat{\Omega} = \frac{1}{N} \mathbf{Y} \Pi_P^\perp \mathbf{Y}^T$, where $\mathbf{P} = \mathbf{X}$ in (A.44), or $\mathbf{P} = [\mathbf{X} \quad \mathbf{U}]$ in (A.46), and $\Pi_P^\perp = \mathbf{I} - \mathbf{P}^T (\mathbf{PP}^T)^{-1} \mathbf{P}$ is the projection onto the orthogonal complement of \mathbf{P} . That is, $\hat{\Omega}$ is obtained as the residual variance ignoring the rank restriction.

The literature also considers $\Omega = \mathbf{Y} \mathbf{Y}^T$ in model (A.44), and $\Omega = \mathbf{Y} \Pi_U^\perp \mathbf{Y}^T$ in model (A.46), as weighting matrices. In such cases, the decomposition (A.45) is equivalent to calculating the canonical correlations; see Section A.5 in this Appendix.

It is noteworthy that there is no asymptotic difference between the two previous approaches, as the optimal solutions obtained with $\Omega = \mathbf{Y} \Pi_P^\perp \mathbf{Y}^T$ equal those obtained with $\Omega = \mathbf{Y} \Pi_U^\perp \mathbf{Y}^T$.

A.10 Riccati equations

A.10.1 Definition

The so-called Riccati difference equation in discrete time is given by the expression:

$$\mathbf{X}_{t+1} = \mathbf{A}\mathbf{X}_t\mathbf{A}^T + \mathbf{M} - \mathbf{A}\mathbf{X}_t\mathbf{H}^T(\mathbf{H}\mathbf{X}_t\mathbf{H}^T + \mathbf{N})^{-1}\mathbf{H}\mathbf{X}_t\mathbf{A}^T, \quad (\text{A.49})$$

where $\{\mathbf{X}_t\}$ is a sequence of symmetric matrices of dimension $n \times n$. If $\{\mathbf{X}_t\}$ converges as $t \rightarrow \infty$ to an equilibrium solution, $\bar{\mathbf{X}}$ (steady-state), then this solution satisfies the so-called *Algebraic Riccati Equation* (ARE):

$$\bar{\mathbf{X}} = \mathbf{A}\bar{\mathbf{X}}\mathbf{A}^T + \mathbf{M} - \mathbf{A}\bar{\mathbf{X}}\mathbf{H}^T(\mathbf{H}\bar{\mathbf{X}}\mathbf{H}^T + \mathbf{N})^{-1}\mathbf{H}\bar{\mathbf{X}}\mathbf{A}^T. \quad (\text{A.50})$$

Equations (A.49)–(A.50) are common in optimal control and signal extraction problems. There is an extensive literature about the convergence of the equation (A.49), and the existence, uniqueness, and computation of the solutions to the ARE (A.50); see Hewer [123], Pappas, Laub, and Sandell [176], Van Dooren [214], Chan, Goodwin, and Sin [49], De Souza, Gevers, and Goodwin [71], or Ionescu, Oara, and Weiss [128]. In particular, the KF equation:

$$\begin{aligned} \mathbf{P}_{t+1} &= \Phi\mathbf{P}_t\Phi^T + \mathbf{E}\mathbf{Q}\mathbf{E}^T \\ &\quad - (\Phi\mathbf{P}_t\mathbf{H}^T + \mathbf{E}\mathbf{S}\mathbf{C}^T)(\mathbf{H}\mathbf{P}_t\mathbf{H}^T + \mathbf{C}\mathbf{R}\mathbf{C}^T)^{-1}(\Phi\mathbf{P}_t\mathbf{H}^T + \mathbf{E}\mathbf{S}\mathbf{C}^T)^T, \end{aligned} \quad (\text{A.51})$$

is a Riccati difference equation, despite the fact that it looks different from (A.49).

Assuming matrix $\mathbf{C}\mathbf{R}\mathbf{C}^T$ to be nonsingular, the term $(\Phi\mathbf{P}_t\mathbf{H}^T + \mathbf{E}\mathbf{S}\mathbf{C}^T)$ can be written as:

$$\Phi\mathbf{P}_t\mathbf{H}^T + \mathbf{E}\mathbf{S}\mathbf{C}^T = \bar{\Phi}\mathbf{P}_t\mathbf{H}^T + \mathbf{E}\mathbf{S}\mathbf{C}^T(\mathbf{C}\mathbf{R}\mathbf{C}^T)^{-1}(\mathbf{H}\mathbf{P}_t\mathbf{H}^T + \mathbf{C}\mathbf{R}\mathbf{C}^T), \quad (\text{A.52})$$

where $\bar{\Phi} = \Phi - \mathbf{E}\mathbf{S}\mathbf{C}^T(\mathbf{C}\mathbf{R}\mathbf{C}^T)^{-1}\mathbf{H}$. Substituting (A.52) in (A.51), we obtain:

$$\mathbf{P}_{t+1} = \bar{\Phi}\mathbf{P}_t\bar{\Phi}^T + \bar{\mathbf{Q}} - \bar{\Phi}\mathbf{P}_t\mathbf{H}^T(\mathbf{H}\mathbf{P}_t\mathbf{H}^T + \mathbf{C}\mathbf{R}\mathbf{C}^T)^{-1}\mathbf{H}\mathbf{P}_t\bar{\Phi}^T, \quad (\text{A.53})$$

with $\bar{\mathbf{Q}} = \mathbf{E}(\mathbf{Q} - \mathbf{S}\mathbf{C}^T(\mathbf{C}\mathbf{R}\mathbf{C}^T)^{-1}\mathbf{C}\mathbf{S}^T)\mathbf{E}^T$. This last equation corresponds to the definition given by (A.49).

De Souza, Gevers, and Goodwin [71] study the convergence conditions of the difference Riccati equation (A.49), as well as the existence of the solution to the ARE (A.50). Some of their main results are as follows.

Definition A.1. The symmetric, real, and positive semidefinite matrix $\bar{\mathbf{X}}$ is a strong solution to the ARE (A.50) if $|\lambda(\mathbf{A} - \bar{\mathbf{K}}\mathbf{H})| \leq 1$, for all $i = 1, \dots, n$, where $\bar{\mathbf{K}} = \mathbf{A}\bar{\mathbf{X}}\mathbf{H}^T(\mathbf{H}\bar{\mathbf{X}}\mathbf{H}^T + \mathbf{N})^{-1}$ corresponds to the KF steady state gain. If $|\lambda(\mathbf{A} - \bar{\mathbf{K}}\mathbf{H})| < 1$, for all $i = 1, \dots, n$, then the strong solution is called “stabilizable.”

Theorem A.2 *The strong solution to the ARE exists and is unique if the pair (\mathbf{H}, \mathbf{A}) is detectable. Moreover, the solution coincides with the stabilizing solution if the pair (\mathbf{H}, \mathbf{A}) is detectable, and the pair $(\mathbf{A}, \bar{\mathbf{M}})$ has no unreachable modes on the unit circle, where $\bar{\mathbf{M}}$ a matrix such that $\mathbf{M} = \bar{\mathbf{M}}\bar{\mathbf{M}}^T$ and $\text{rank}(\mathbf{M}) = \text{rank}(\bar{\mathbf{M}})$.*

Proof. See De Souza, Gevers, and Goodwin [71], Theorem 3.2.

Corollary A.1 A minimal SEM satisfies that: (a) the ARE has a strong and unique solution; (b) if the eigenvalues of $\Phi - EH$ are not outside the unit circle, then the strong solution is $P = 0$; and (c) if these eigenvalues are strictly within the unit circle (invertibility), then $P = 0$ is a stabilizable solution.

Proof. See Casals [38], Appendix A.12.

Theorem A.3 When $X_0 - \bar{X} \geq 0$ then $\lim_{t \rightarrow \infty} X_t = \bar{X}$, where \bar{X} is the strong solution for the ARE (A.50), if and only if the pair (H, A) is detectable.

Proof. See De Souza, Gevers, and Goodwin [71], Theorem 4.2.

Corollary A.2 If the eigenvalues of $\Phi - EH$ are not outside the unit circle² in a minimal SEM, then the difference Riccati equation (A.51) converges to $P = 0$. When the eigenvalues of $\Phi - EH$ are strictly within the unit circle, this convergence becomes exponentially fast.

Proof. See Casals [38], Appendix A.12.

A.10.2 Solving the ARE in the general case

The procedures to solve the ARE (A.50) can be divided into iterative and non-iterative methods. The first are based on the convergence properties of the difference Riccati equation to achieve an equilibrium solution by propagating this equation. Among them, it is important to mention Hewer's method [123], which is based on the following result.

Suppose that $\{X_k\}$ is the sequence of Lyapunov equations:

$$X_k = A_k X_k A_k^T + M + K_k N K_k^T, \quad k = 0, 1, 2, \dots, \quad (\text{A.54})$$

with $A_k = A - K_k H$, and:

$$K_k = A_k X_{k-1} H^T (H X_{k-1} H^T + N)^{-1}, \quad k = 1, 2, \dots, \quad (\text{A.55})$$

where K_0 is an arbitrary matrix such that $A_0 = A - K_0 H$ is stable. Then:

$$X_0 \geq X_1 \geq \dots \geq \bar{X}, \quad (\text{A.56})$$

and

$$\lim_{k \rightarrow \infty} X_k = \bar{X}. \quad (\text{A.57})$$

The Hewer [123] algorithm can be summarized by the following iterative scheme:

²This condition is equivalent to requiring that the MA roots of the model lie on or within the unit circle, thus excluding strict noninvertibility.

Step 1) Select \mathbf{K}_0 so that \mathbf{A}_0 is stable. Note that if \mathbf{A} is stable, the procedure can be initialized with the condition $\mathbf{K}_0 = \mathbf{0}$.

Step 2) Calculate \mathbf{X}_k , for all $k = 0, 1, 2, \dots$, by solving Lyapunov equation (A.54).

Step 3) Calculate \mathbf{K}_{k+1} , for all $k = 0, 1, 2, \dots$, from (A.55).

Step 4) If $\|\mathbf{X}_k - \mathbf{X}_{k-1}\|/\|\mathbf{X}_{k-1}\| > \tau$, where τ is a scalar that sets a convergence criterion, go back to Step 2). Otherwise, stop the iteration.

The value for τ will be set according to the ε -machine and the condition number of the Riccati equation; see Petkov, Christov, and Konstantinov [181]. Obviously, the computational load of this algorithm depends on the number of iterations until convergence.

Non-iterative methods are based on solving a problem of generalized eigenvalues; see Pappas, Laub, and Sandell [176], Van Dooren [214], Chan, Goodwin, and Sin [49], and Ionescu, Oara, and Weiss [128]. Suppose that the pair (\mathbf{C}, \mathbf{A}) is detectable, meaning that there is a strong and unique solution to the difference Riccati equation. The equation (A.50) can then be solved by computing the generalized eigenvalue matrix in:

$$\mathbf{T}\mathbf{v} = \lambda \mathbf{Y}\mathbf{v}, \quad (\text{A.58})$$

where:

$$\mathbf{T} = \begin{bmatrix} \mathbf{A}^T & \mathbf{0} \\ -\mathbf{M} & \mathbf{I}_n \end{bmatrix}, \quad (\text{A.59})$$

and

$$\mathbf{Y} = \begin{bmatrix} \mathbf{I}_n & (\mathbf{H}^T \mathbf{N}^{-1} \mathbf{H}) \\ \mathbf{0} & \mathbf{A} \end{bmatrix}. \quad (\text{A.60})$$

If matrix \mathbf{A} is nonsingular, then \mathbf{Y} is also nonsingular, and solving (A.58) then reduces to solving the following eigenvalue problem³:

$$(\mathbf{Y}^{-1}\mathbf{T})\mathbf{v} = \lambda \mathbf{v}. \quad (\text{A.61})$$

The problem defined in (A.61) has the following properties (see De Souza, Gevers, and Goodwin [71]):

1. If λ is an eigenvalue of the pencil $\mathbf{T} - \lambda \mathbf{Y}$, then $1/\lambda$ is an eigenvalue of the same pencil, with the same multiplicity.
2. If \mathbf{A} is nonsingular, then all its generalized eigenvalues are nonzero.
3. If \mathbf{A} is singular, then the pencil $\mathbf{T} - \lambda \mathbf{Y}$ has at least one null generalized eigenvalue.
4. If $\lambda = 0$ is a generalized eigenvalue of \mathbf{A} with multiplicity r , then the pencil $\mathbf{T} - \lambda \mathbf{Y}$ has only $2n - r$ finite generalized eigenvalues.

³Petkov, Christov, and Konstantinov [181] provide an alternative form to this generalized eigenvalue problem that avoids the inversion of the matrix \mathbf{N} in (A.60). This is the solution implemented in E^4 ; see Appendix C

The eigenvalues of the problem (A.58) can be sorted as:

$$\overbrace{0, \dots, 0}^r, \lambda_{r+1}, \dots, \lambda_n, \frac{1}{\lambda_n}, \dots, \frac{1}{\lambda_{r+1}}, \overbrace{\infty, \dots, \infty}^r \quad (\text{A.62})$$

where $0 < |\lambda_i| \leq 1$, for all $i = r+1, \dots, n$. If \mathbf{V} is the matrix of ordered generalized eigenvalues according to (A.62), partitioning \mathbf{V} into four blocks of dimension $(n \times n)$ yields:

$$\mathbf{V} = \begin{bmatrix} \mathbf{V}_{11} & \mathbf{V}_{12} \\ \mathbf{V}_{21} & \mathbf{V}_{22} \end{bmatrix}. \quad (\text{A.63})$$

Then:

1. \mathbf{V}_{11} and \mathbf{V}_{12} are nonsingular.
2. $\tilde{\mathbf{X}} = \mathbf{V}_{21}\mathbf{V}_{11}^{-1}$ and $\tilde{\mathbf{X}} = \mathbf{V}_{22}\mathbf{V}_{12}^{-1}$ are solutions to the ARE (A.50).
3. The eigenvalues of the matrix $\mathbf{A} - \bar{\mathbf{K}}\mathbf{H}$, where $\bar{\mathbf{K}} = \mathbf{A}\tilde{\mathbf{X}}\mathbf{H}^T(\mathbf{H}\tilde{\mathbf{X}}\mathbf{H}^T + \mathbf{N})^{-1}$, coincide with the n eigenvalues of the problem given by (A.58) with modulus less than or equal to unity.
4. $\tilde{\mathbf{X}} = \mathbf{V}_{21}\mathbf{V}_{11}^{-1}$ is the strong solution to (A.50).

Petkov, Christov, and Konstantinov [181] recommend first using the non-iterative procedure explained above, and then refining the solution $\tilde{\mathbf{X}}$ by means of Hewer's algorithm with $\mathbf{X}_0 = \tilde{\mathbf{X}}$.

A.10.3 Solving the ARE for GARCH models

In Section 6.3 we stated that in a representation such as (6.56)–(6.57), with GARCH errors, $\bar{\mathbf{P}} = \mathbf{0}$ is the solution of the algebraic Riccati equation:

$$\begin{aligned} \bar{\mathbf{P}} = & \Phi \bar{\mathbf{P}} \Phi^T + \mathbf{E} \Sigma_t \mathbf{E}^T \\ & - (\Phi \bar{\mathbf{P}} \mathbf{H}^T + \mathbf{E} \Sigma_t)(\mathbf{H} \bar{\mathbf{P}} \mathbf{H}^T + \Sigma_t)^{-1} (\Phi \bar{\mathbf{P}} \mathbf{H}^T + \mathbf{E} \Sigma_t)^T. \end{aligned} \quad (\text{A.64})$$

This statement can be proved as follows. In models with GARCH errors the conditional covariances Σ_t are a source of parametric variation. However, it holds that:

1. $\bar{\mathbf{P}} = \mathbf{0}$ is the solution to the algebraic Riccati equation (A.64), associated with the Riccati difference equation (6.62) in Chapter 6, for any value of Σ_t .
2. Following De Souza, Gevers, and Goodwin [71] and denoting $\bar{\Sigma} = \sup\{\Sigma_t\}$, it can be shown that:
 - (a) $\mathbf{P}_{t+1|t} = \Phi \mathbf{P}_{t|t-1} \Phi^T + \mathbf{E} \bar{\Sigma} \mathbf{E}^T - \mathbf{K}_t \mathbf{B}_t \mathbf{K}_t^T$, where $\mathbf{K}_t = (\Phi \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{E} \bar{\Sigma}) \mathbf{B}_t^{-1}$ and $\mathbf{B}_t = \mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^T + \bar{\Sigma}$, converges to its strong solution if and only if the model (6.56)–(6.57) is detectable.
 - (b) The strong solution is $\bar{\mathbf{P}} = \mathbf{0}$ if the system (6.56)–(6.57) does not have moving average roots strictly inside the unit circle.
3. Following Bitmead *et al.* [23] it can be shown that, if $\mathbf{P}_{t|t-q}^* \leq \mathbf{P}_{t|t-1}$ then $\mathbf{P}_{t+1|t}^* \leq \mathbf{P}_{t+1|t}$, where $\mathbf{P}_{t+1|t}^*$ is propagated with Σ_t instead of $\bar{\Sigma}$. Because of

this, the convergence of $P_{t+1|t}$ to zero implies that $P_{t+1|t}^*$ will also converge to zero. ■

A.11 Kalman filter

The Kalman filter (KF) is a minimum-variance recursive algorithm which predicts the state vector \mathbf{x}_{t+1} of a linear dynamic system from the information available up to time t , as well as the covariance matrix of the prediction error. The Kalman filter equations are easily derived from the following results:

Result 1) Given the random vectors \mathbf{x} and \mathbf{y} , with $E(\mathbf{x}) = \bar{\mathbf{x}}$, $E(\mathbf{y}) = \bar{\mathbf{y}}$, $\tilde{\mathbf{x}} = \mathbf{x} - \bar{\mathbf{x}}$, $\tilde{\mathbf{y}} = \mathbf{y} - \bar{\mathbf{y}}$, $E(\tilde{\mathbf{x}}\tilde{\mathbf{x}}^T) = \Sigma_{xx}$, $E(\tilde{\mathbf{y}}\tilde{\mathbf{y}}^T) = \Sigma_{yy}$, and $E(\tilde{\mathbf{y}}\tilde{\mathbf{x}}^T) = \Sigma_{yx}$, consider linear estimators of \mathbf{x} of the form $F(\mathbf{y}) = \mathbf{A}\mathbf{y} + \mathbf{b}$. Then, the minimum-variance (linear) estimator of \mathbf{x} conditional on \mathbf{y} can be shown to be⁴:

$$E(\mathbf{x}|\mathbf{y}) = \bar{\mathbf{x}} + \Sigma_{xy}\Sigma_{yy}^{-1}\tilde{\mathbf{y}}, \quad (\text{A.65})$$

where the prediction error variance is given by:

$$E[(\mathbf{x} - E(\mathbf{x}|\mathbf{y}))(\mathbf{x} - E(\mathbf{x}|\mathbf{y}))^T] = \Sigma_{xx} - \Sigma_{xy}\Sigma_{yy}^{-1}\Sigma_{xy}^T. \quad (\text{A.66})$$

This estimator satisfies,

$$E[\|\mathbf{x} - E(\mathbf{x}|\mathbf{y})\|_F^2] \leq E[\|\mathbf{x} - F(\mathbf{y})\|_F^2],$$

among all linear estimators of the form $F(\mathbf{y}) = \mathbf{A}\mathbf{y} + \mathbf{b}$. If \mathbf{x} and \mathbf{y} are jointly normally distributed, then the linear minimum-variance estimator (A.65) coincides with the minimum-variance estimator of \mathbf{x} conditional on \mathbf{y} , where the form of the estimator is any function $F(\mathbf{y})$, not necessarily linear in \mathbf{y} .

Result 2) If \mathbf{x} , \mathbf{y} , and \mathbf{z} are random vectors, then:

$$E(\mathbf{x}|\mathbf{y}, \mathbf{z}) = E(\mathbf{x}|\mathbf{z}) - \bar{\Sigma}_{xx}\bar{\Sigma}_{yy}^{-1}[\mathbf{y} - E(\mathbf{y}|\mathbf{z})], \quad (\text{A.67})$$

and

$$E[(\mathbf{x} - E(\mathbf{x}|\mathbf{y}, \mathbf{z}))(\mathbf{x} - E(\mathbf{x}|\mathbf{y}, \mathbf{z}))^T] = \bar{\Sigma}_{xx} - \bar{\Sigma}_{xy}\bar{\Sigma}_{yy}^{-1}\bar{\Sigma}_{xy}^T, \quad (\text{A.68})$$

with

$$\bar{\Sigma}_{xx} = E[(\mathbf{x} - E[\mathbf{x}|\mathbf{z}])(\mathbf{x} - E[\mathbf{x}|\mathbf{z}])^T] \quad (\text{A.69})$$

$$\bar{\Sigma}_{xy} = E[(\mathbf{x} - E[\mathbf{x}|\mathbf{z}])(\mathbf{y} - E[\mathbf{y}|\mathbf{z}])^T] \quad (\text{A.70})$$

$$\bar{\Sigma}_{yy} = E[(\mathbf{y} - E[\mathbf{y}|\mathbf{z}])(\mathbf{y} - E[\mathbf{y}|\mathbf{z}])^T]. \quad (\text{A.71})$$

⁴Although we follow the frequently used convention of denoting this estimator by the conditional expectation operator, $E(\mathbf{x}|\mathbf{y})$, it does not coincide with the actual conditional expectation of \mathbf{x} given \mathbf{y} , unless \mathbf{x} and \mathbf{y} are jointly normally distributed. Thus, in general $E(\mathbf{x}|\mathbf{y})$ is used here in the sense of optimal (minimum error variance) projection operator of \mathbf{x} onto the space of all linear combinations of $(\mathbf{1}, \mathbf{y})$.

Result 3) Assume that the vector \mathbf{x} can be defined as $\mathbf{x} = \mathbf{A}\mathbf{y} + \mathbf{b} + \mathbf{v}$, where \mathbf{v} is a zero-mean random vector with $\text{E}(\mathbf{v}\mathbf{v}^T) = \mathbf{Q}$ which is uncorrelated with random vectors \mathbf{y} and \mathbf{z} . Then:

$$\text{E}(\mathbf{x}|\mathbf{y}) = \mathbf{A}\mathbf{y} + \mathbf{b} \quad (\text{A.72})$$

$$\text{E}[(\mathbf{x} - \text{E}(\mathbf{x}|\mathbf{y}))(\mathbf{x} - \text{E}(\mathbf{x}|\mathbf{y}))^T] = \mathbf{Q} \quad (\text{A.73})$$

$$\text{E}(\mathbf{x}|\mathbf{z}) = \mathbf{A}\text{E}(\mathbf{y}|\mathbf{z}) + \mathbf{b} \quad (\text{A.74})$$

$$\bar{\Sigma}_{xx} = \mathbf{A}\bar{\Sigma}_{yy}\mathbf{A}^T + \mathbf{Q}. \quad (\text{A.75})$$

Anderson and Moore [6, ch. 5] provide a detailed treatment of the linear minimum-variance estimator, from which it is straightforward to derive Results (1)–(3).

In deriving the KF recursions below, we will use the following notation:

$$\hat{\mathbf{x}}_{t|t-j} = \text{E}(\mathbf{x}_t | \mathbf{z}_0, \dots, \mathbf{z}_{t-j}, \mathbf{u}_0, \dots, \mathbf{u}_N) \quad (\text{A.76})$$

$$\tilde{\mathbf{x}}_{t|t-j} = \mathbf{x}_t - \hat{\mathbf{x}}_{t|t-j} \quad (\text{A.77})$$

$$\hat{\mathbf{z}}_{t|t-j} = \text{E}(\mathbf{z}_t | \mathbf{z}_0, \dots, \mathbf{z}_{t-j}, \mathbf{u}_0, \dots, \mathbf{u}_N) \quad (\text{A.78})$$

$$\tilde{\mathbf{z}}_{t|t-j} = \mathbf{z}_t - \hat{\mathbf{z}}_{t|t-j} \quad (\text{A.79})$$

$$\mathbf{P}_{t|t-j} = \text{E}(\tilde{\mathbf{x}}_{t|t-j}\tilde{\mathbf{x}}_{t|t-j}^T) \quad (\text{A.80})$$

$$\mathbf{B}_t = \text{E}(\tilde{\mathbf{z}}_t\tilde{\mathbf{z}}_t^T). \quad (\text{A.81})$$

Now, for a MEM in state space and according to Result 3), we have:

$$\begin{bmatrix} \hat{\mathbf{x}}_{t+1|t-1} \\ \hat{\mathbf{z}}_{t|t-1} \end{bmatrix} = \begin{bmatrix} \Phi \\ \mathbf{H} \end{bmatrix} \hat{\mathbf{x}}_{t|t-1} + \begin{bmatrix} \Gamma \\ \mathbf{D} \end{bmatrix} \mathbf{u}_t, \quad (\text{A.82})$$

$$\begin{bmatrix} \mathbf{P}_{t+1|t-1} \\ \tilde{\mathbf{B}}_t \end{bmatrix} = \begin{bmatrix} \Phi \\ \mathbf{H} \end{bmatrix} \mathbf{P}_{t|t-1} \begin{bmatrix} \Phi^T & \mathbf{H}^T \end{bmatrix} + \begin{bmatrix} \mathbf{E} & \mathbf{0} \\ \mathbf{0} & \mathbf{C} \end{bmatrix} \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix} \begin{bmatrix} \mathbf{E}^T & \mathbf{0} \\ \mathbf{0} & \mathbf{C}^T \end{bmatrix}, \quad (\text{A.83})$$

and applying Result 2):

$$\hat{\mathbf{x}}_{t+1|t} = \hat{\mathbf{x}}_{t+1|t-1} + (\Phi \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}^T) \mathbf{B}_t^{-1} (\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}), \quad (\text{A.84})$$

$$\mathbf{P}_{t+1|t} = \mathbf{P}_{t+1|t-1} - (\Phi \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}^T) \mathbf{B}_t^{-1} (\Phi \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}^T)^T. \quad (\text{A.85})$$

Then, the Kalman filter is given by the following expressions:

$$\hat{\mathbf{z}}_{t|t-1} = \mathbf{H}\hat{\mathbf{x}}_{t|t-1} + \mathbf{D}\mathbf{u}_t \quad (\text{A.86})$$

$$\hat{\mathbf{x}}_{t+1|t} = \Phi\hat{\mathbf{x}}_{t|t-1} + \Gamma\mathbf{u}_t + \mathbf{K}_t (\mathbf{z}_t - \hat{\mathbf{z}}_{t|t-1}) \quad (\text{A.87})$$

$$\mathbf{K}_t = (\Phi \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{E} \mathbf{S} \mathbf{C}^T) \mathbf{B}_t^{-1} \quad (\text{A.88})$$

$$\mathbf{B}_t = \mathbf{H} \mathbf{P}_{t|t-1} \mathbf{H}^T + \mathbf{C} \mathbf{R} \mathbf{C}^T \quad (\text{A.89})$$

$$\mathbf{P}_{t+1|t} = \Phi \mathbf{P}_{t|t-1} \Phi^T + \mathbf{E} \mathbf{Q} \mathbf{E}^T - \mathbf{K}_t \mathbf{B}_t \mathbf{K}_t^T, \quad (\text{A.90})$$

where the initial conditions for the filter are:

$$\hat{\mathbf{x}}_{0|-1} = \mathbb{E}(\mathbf{x}_0 | \mathbf{u}_0, \dots, \mathbf{u}_N), \quad (\text{A.91})$$

and

$$\mathbf{P}_{0|-1} = \mathbb{E}[(\mathbf{x}_0 - \hat{\mathbf{x}}_{0|-1})(\mathbf{x}_0 - \hat{\mathbf{x}}_{0|-1})^T]. \quad (\text{A.92})$$

Matrix \mathbf{K}_t is the Kalman filter gain, and \mathbf{B}_t corresponds to the prediction error variance of \mathbf{z}_t obtained from past values of the input and output information.

Equations (A.86)–(A.92) correspond to the general case, where the input of the system is stochastic. The analytical expression for the moments (A.91) and (A.92) was derived in Chapter 5.

Appendix B

Asymptotic properties of maximum likelihood estimates

B.1	Preliminaries	211
B.2	Basic likelihood results for the State-Space model	214
B.2.1	The information matrix	214
B.2.2	Regularity conditions	216
B.2.3	Choice of estimation method	217
B.3	The State-Space model with cross-sectional extension	218

The aim of this Appendix is to provide a brief and accessible discussion of the asymptotic theory underlying the ML estimates for simplified versions of the SS model with time-invariant parameters. These include the MEM of Chapter 5, and its cross-sectionally extended counterpart in Chapter 11 without inputs. The hope is that this will provide the reader with enough background to understand something of what goes on “behind the scenes” as software packages produce statistical inference output in the form of standard errors and p-values.

B.1 Preliminaries

Assume that the observation equation sample is $\mathbf{z} = \text{vec}(\mathbf{z}_1, \mathbf{z}_2, \dots, \mathbf{z}_N)$, with N being the sample size and $\mathbf{z} \in \mathbb{R}^{mN}$ a vectorization of all the observed values. Denote by $\theta_0 \in \mathbb{D}$ the *true* value of the d -dimensional parameter vector θ in the model under consideration, which is simply a chosen vectorization of the entire set of model coefficients, Θ , and where $\mathbb{D} \subset \mathbb{R}^d$ is the parameter space. In addition, let $F(\mathbf{z}; \theta)$ denote the cumulative distribution function of \mathbf{z} . In this context, θ_0 is said to be *identifiable* in the sense already defined in Chapter 9, i.e., if and only if

$$\theta_1 \neq \theta_0 \quad \Rightarrow \quad F(\mathbf{z}; \theta_1) \neq F(\mathbf{z}; \theta_0), \quad \text{for at least one } \mathbf{z}. \quad (\text{B.1})$$

One can then define two types of identifiability for θ_0 (or the model): *global identifiability* if (B.1) holds for all $\theta_1 \in \mathbb{D}$, and *local identifiability* if (B.1) holds for

all θ_1 only in a neighborhood of θ_0 . For the asymptotic properties discussed here, it suffices to have local identifiability.

Throughout the discussion, we will let $\ell(\boldsymbol{\theta})$ denote the *negative of the log-likelihood* function. The ML estimate, which we denote generically by $\hat{\boldsymbol{\theta}}_N$, is then the minimizer of $\ell(\boldsymbol{\theta})$. We will assume that the reader of this more technical material is familiar with basic notions of large sample theory for estimators in the context of parametric models, namely *consistency* and *asymptotic normality*, a very accessible introduction to which can be found in, e.g., Davidson and MacKinnon [61].

The ML approach was first proposed by Fisher [96]. Cramer [58] proved its consistency in the case of IID observations, under the assumption that the likelihood function is differentiable at least three times. Many relevant estimation problems, including those treated in this book, do not correspond to this classical formulation because the observations are neither identically nor independently distributed. Extensions concerning the consistency of the ML estimator when the observations are dependent, were first proposed by Cramer [58] and Wald [218], among others.

Nowadays, a standard approach for establishing asymptotic normality (and hence consistency) uses a Taylor series expansion of the *score* equations,

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \mathbf{0},$$

in conjunction with a central limit theorem on the (uncorrelated) elements of the score vector, and involves the associated *Hessian* matrix,

$$\mathcal{H}_N(\boldsymbol{\theta}) = -\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}.$$

From this, three versions of the (Fisher) *information* matrix can be defined:

1. the *expected* information matrix,

$$\mathbf{J}_N(\boldsymbol{\theta}) = -\mathbb{E}[\mathcal{H}_N(\boldsymbol{\theta})],$$

2. the *observed* (or *empirical*) information matrix,

$$\hat{\mathbf{J}}_N(\hat{\boldsymbol{\theta}}) = -\mathcal{H}_N(\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}},$$

3. and the *asymptotic* information matrix,

$$\mathcal{I}(\boldsymbol{\theta}) = \text{plim}_{N \rightarrow \infty} \frac{1}{N} \mathbf{J}_N(\boldsymbol{\theta}),$$

where the “plim” in the last expression denotes “limit in probability.”

For time series data, and under appropriate regularity conditions, the same (familiar) results are obtained as in the IID case.

Consistency:

$$\text{plim}_{N \rightarrow \infty} \hat{\boldsymbol{\theta}}_N = \boldsymbol{\theta}_0. \tag{B.2}$$

Asymptotic unbiasedness: as a consequence of consistency,

$$\lim_{N \rightarrow \infty} E(\hat{\theta}_N) = \theta_0. \quad (\text{B.3})$$

Asymptotic normality. As $N \rightarrow \infty$, a scaled and centered version of $\hat{\theta}_N$ converges in distribution to a Gaussian random vector with mean $\mathbf{0}$ and covariance matrix given by the inverse of the asymptotic information matrix,

$$\sqrt{N}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(\mathbf{0}, \mathcal{I}(\theta_0)^{-1}). \quad (\text{B.4})$$

Asymptotic efficiency: as a consequence of the above asymptotic normality result,

$$\lim_{N \rightarrow \infty} \text{var}\left(\sqrt{N}(\hat{\theta}_N - \theta_0)\right) = \mathcal{I}(\theta_0)^{-1}, \quad (\text{B.5})$$

so that the variance of the ML estimator attains the *Cramer-Rao lower bound*.

These results rely on invoking the *information matrix equality*,

$$\mathcal{I}(\theta) = -\underset{N \rightarrow \infty}{\text{plim}} \frac{1}{N} \mathcal{H}_N(\theta),$$

which holds under correct model specification; with correspondingly more complex results otherwise; see Davidson and MacKinnon [62]. The primary usefulness of (B.4) lies in the fact that it can be exploited for the construction of a $(1 - \alpha)100\%$ confidence region for θ_0 defined by the set:

$$CR(\theta) = \left\{ \theta \in \mathbb{R}^d : (\theta - \hat{\theta}_N)^T \mathcal{I}(\theta_0) (\theta - \hat{\theta}_N) \leq \frac{1}{N} \chi_{1-\alpha}^2(d) \right\}, \quad (\text{B.6})$$

where $\chi_{1-\alpha}^2(d)$ is the $1 - \alpha$ quantile from a chi-square distribution with d degrees of freedom. From a practical point of view, one replaces the (unknown) asymptotic information matrix in (B.6) with either the expected or observed information matrices normalized by N , that is,

$$\mathcal{I}(\theta_0) \approx \frac{\mathcal{J}_N(\hat{\theta}_N)}{N}, \quad \text{or} \quad \mathcal{I}(\theta_0) \approx \frac{\hat{\mathcal{J}}_N(\hat{\theta}_N)}{N}.$$

The latter is often the obvious choice in a software package implementation seeking automation in the calculation of standard errors, since it requires only numerical evaluation of the Hessian. However, the results can be very inaccurate and highly unstable if (as is often the case) nothing is assumed about the form of the derivatives; see Jamshidian and Jennrich ([129]). Improvements can therefore be obtained by understanding something about the structure of the Hessian.

Estimates of the Hessian are useful not only in inference, but also in diagnosing model identifiability. It can be shown that (except for possibly a set of Lebesgue measure zero) local identifiability and asymptotic nonsingularity of the Hessian are equivalent conditions; see Caines [37].

B.2 Basic likelihood results for the State-Space model

Consider the MEM with exogenous inputs and time-invariant transition matrices:

$$\mathbf{x}_{t+1} = \Phi \mathbf{x}_t + \Gamma \mathbf{u}_t + E \mathbf{w}_t, \quad (\text{B.7})$$

$$\mathbf{z}_t = \mathbf{H} \mathbf{x}_t + \mathbf{D} \mathbf{u}_t + \mathbf{C} \mathbf{v}_t, \quad (\text{B.8})$$

observed for $t = 1, \dots, N$, with corresponding error covariances

$$\mathbb{E} \left(\begin{bmatrix} \mathbf{w}_t \\ \mathbf{v}_t \end{bmatrix} \begin{bmatrix} \mathbf{w}_t^T & \mathbf{v}_t^T \end{bmatrix} \right) = \begin{bmatrix} \mathbf{Q} & \mathbf{S} \\ \mathbf{S}^T & \mathbf{R} \end{bmatrix}, \quad (\text{B.9})$$

and initial condition $\mathbf{x}_1 \sim \mathcal{N}(\boldsymbol{\mu}, \mathbf{P})$. As seen in Chapter 5, the prediction error decomposition (also called innovations) provides a very concise expression for the log-likelihood. Dropping constants and making explicit the dependence on $\boldsymbol{\theta}$, we have the negative log-likelihood expression

$$\begin{aligned} \ell(\boldsymbol{\theta}) := -\log L(\boldsymbol{\theta}) &= \frac{1}{2} \sum_{t=1}^N \{ \log |\mathbf{B}_t(\boldsymbol{\theta})| + \tilde{\mathbf{z}}_t(\boldsymbol{\theta})^T \mathbf{B}_t(\boldsymbol{\theta})^{-1} \tilde{\mathbf{z}}_t(\boldsymbol{\theta}) \} \\ &:= \frac{1}{2} \sum_{t=1}^N v_t(\boldsymbol{\theta}), \end{aligned} \quad (\text{B.10})$$

where the innovations, $\tilde{\mathbf{z}}_t(\boldsymbol{\theta}) = \mathbf{z}_t - \mathbf{H}_t \mathbf{x}_{t|t-1} - \mathbf{D} \mathbf{u}_t$, are as defined in Chapter 5, and we have $\tilde{\mathbf{z}}_t(\boldsymbol{\theta}) \sim \text{NID}(\mathbf{0}, \mathbf{B}_t(\boldsymbol{\theta}))$ under the Gaussian assumption. The ML estimate $\hat{\boldsymbol{\theta}}_N$ is therefore the minimizer of (B.10).

In the context of linear SS models, Pagan [173] obtained consistency and asymptotic normality for $\hat{\boldsymbol{\theta}}_N$ by applying the conditions developed by Crowder [59] to a univariate model without inputs; a result which generalizes immediately to the multivariate case. The conditions required to assure properties (B.2)–(B.5) are: (a) model identifiability, (b) some regularity conditions on $F(\mathbf{z}; \boldsymbol{\theta})$, and (c) some conditions on the convergence properties of the Hessian evaluated in a neighborhood of $\boldsymbol{\theta}_0$.

B.2.1 The information matrix

Calculation of the Hessian for $\ell(\boldsymbol{\theta})$ relies on evaluating the matrix of second partial derivatives. Caines [37] provides this in element-wise form, while Lutkepohl [158] opts for a matrix representation, but both omit some details which we present in the following lemma.

Lemma 1 *For the expression $v_t(\boldsymbol{\theta})$ as defined in (B.10), and omitting explicit dependence on t and $\boldsymbol{\theta}$ for notational expedience, the $(d \times d)$ Hessian matrix with respect to the vector $\boldsymbol{\theta}$ is:*

(i) given in matrix form as

$$\frac{\partial^2 v(\theta)}{\partial \theta \partial \theta^T} = 2J_{\tilde{z}}^T B^{-1} J_{\tilde{z}} - J_B^T (B^{-1} \otimes B^{-1}) J_B \quad (\text{B.11})$$

$$+ J_B^T (B^{-1} \tilde{z} \tilde{z}^T B^{-1} \otimes B^{-1} + B^{-1} \otimes B^{-1} \tilde{z} \tilde{z}^T B^{-1}) J_B \quad (\text{B.12})$$

$$- 2J_B^T (B^{-1} J_{\tilde{z}} \otimes B^{-1} \tilde{z}) + 2H_{\tilde{z}}^T (I_d \otimes B^{-1} \tilde{z}) \quad (\text{B.13})$$

$$- J_{\tilde{z}}^T (I_m \otimes \tilde{z}^T + \tilde{z}^T \otimes I_m) (B^{-1} \otimes B^{-1}) J_B \quad (\text{B.14})$$

$$+ H_B^T \{ I_d \otimes \text{vec}[B^{-1} (I_d - \tilde{z} \tilde{z}^T B^{-1})] \} \quad (\text{B.15})$$

(ii) whose (p, q) element is given by

$$\begin{aligned} \frac{\partial^2 v(\theta)}{\partial \theta_p \partial \theta_q} &= 2\tilde{z}^T B^{-1} \frac{\partial^2 \tilde{z}}{\partial \theta_p \partial \theta_q} - 2\tilde{z}^T B^{-1} \frac{\partial B}{\partial \theta_q} B^{-1} \frac{\partial \tilde{z}}{\partial \theta_p} + 2 \frac{\partial \tilde{z}^T}{\partial \theta_p} B^{-1} \frac{\partial \tilde{z}}{\partial \theta_q} \\ &\quad - 2\tilde{z}^T B^{-1} \frac{\partial B}{\partial \theta_p} B^{-1} \frac{\partial \tilde{z}}{\partial \theta_q} + 2\tilde{z}^T B^{-1} \frac{\partial B}{\partial \theta_q} B^{-1} \frac{\partial B}{\partial \theta_p} B^{-1} \tilde{z} \\ &\quad - \tilde{z}^T B^{-1} \frac{\partial^2 B}{\partial \theta_p \partial \theta_q} B^{-1} \tilde{z} + \text{tr} \left(B^{-1} \frac{\partial^2 B}{\partial \theta_p \partial \theta_q} - B^{-1} \frac{\partial B}{\partial \theta_p} B^{-1} \frac{\partial B}{\partial \theta_q} \right). \end{aligned}$$

Both expressions depend on Jacobians and Hessians of \tilde{z} and B . While this is explicit in (ii), the notation employed in (i) is as follows:

- $m \times d$ Jacobian matrix of \tilde{z} , and $m^2 \times d$ Jacobian matrix of B ,

$$J_{\tilde{z}} = \frac{\partial \tilde{z}}{\partial \theta^T}, \quad \text{and} \quad J_B = \frac{\partial \text{vec}(B)}{\partial \theta^T},$$

- $md \times d$ Hessian matrix of \tilde{z} , and $m^2 d \times d$ Hessian matrix of B ,

$$H_{\tilde{z}} = \frac{\partial \text{vec}(J_{\tilde{z}})}{\partial \theta^T}, \quad \text{and} \quad H_B = \frac{\partial \text{vec}(J_B)}{\partial \theta^T}.$$

Proof for Lemma 1 These results are tedious to derive, but follow by straightforward application of elementary matrix calculus rules, which can be found in, e.g., Lutkepohl [157] and Harville [121]. ■

Now, using the fact that $E(\tilde{z}_t) = \mathbf{0}$ and $E(\tilde{z}_t \tilde{z}_t^T) = B_t$, we obtain the expected information matrix,

$$J_N(\theta) = \frac{1}{2} \sum_{t=1}^N E \left[\frac{\partial^2 v_t(\theta)}{\partial \theta \partial \theta^T} \right] \quad (\text{B.16})$$

$$= \frac{1}{2} \sum_{t=1}^N \left[\frac{\partial \text{vec}(B_t)^T}{\partial \theta} (B_t^{-1} \otimes B_t^{-1}) \frac{\partial \text{vec}(B_t)}{\partial \theta^T} + 2E \left(\frac{\partial \tilde{z}_t^T}{\partial \theta} B_t^{-1} \frac{\partial \tilde{z}_t}{\partial \theta^T} \right) \right] \quad (\text{B.17})$$

$$:= \frac{1}{2} \Sigma_N(\theta). \quad (\text{B.18})$$

This follows straightforwardly by taking the expectation of terms (B.11) and (B.12), and noting that the expectation of (B.15) is zero. Each of the remaining terms contains both \tilde{z}_t and its Jacobian $J_{\tilde{z}}$; consequently, the fact that these two terms are statistically independent implies that the expectation of terms (B.13) and (B.14) is also null. Using similar arguments, the (p, q) element of the expected information matrix is:

$$\mathbf{J}_N(\boldsymbol{\theta})_{p,q} = \frac{1}{2} \sum_{t=1}^N \left[\text{tr} \left(\mathbf{B}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_p} \mathbf{B}_t^{-1} \frac{\partial \mathbf{B}_t}{\partial \theta_q} \right) + 2\mathbb{E} \left(\frac{\partial \tilde{z}_t^T}{\partial \theta_p} \mathbf{B}_t^{-1} \frac{\partial \tilde{z}_t}{\partial \theta_q} \right) \right]. \quad (\text{B.19})$$

B.2.2 Regularity conditions

Under appropriate regularity conditions, (B.2)–(B.5) will hold, and in particular,

$$\sqrt{N} (\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \left[\frac{\Sigma_N(\boldsymbol{\theta}_0)}{2N} \right]^{-1} \right), \quad (\text{B.20})$$

with $\Sigma_N(\boldsymbol{\theta})$ defined by (B.18). To give some idea of the nature of the assumptions involved we adapt the conditions given by Lutkepohl [158] to the situation of SS model (B.7)–(B.9).

- (i) The model is *stable* in the sense of the Definition 2.1, given in Chapter 2.
- (ii) The state and observational white noise errors are Gaussian, $\mathbf{w}_t \sim N(\mathbf{0}, \mathbf{Q})$, $\mathbf{v}_t \sim N(\mathbf{0}, \mathbf{R})$, with $\text{cov}(\mathbf{w}_t, \mathbf{v}_t) = \mathbf{S} \equiv \mathbf{0}$.
- (iii) The initial state vector is Gaussian, $\mathbf{x}_1 \sim N(\boldsymbol{\mu}, \mathbf{P})$, and is independent of $\{\mathbf{w}_t, \mathbf{v}_{t+1}\}$, for $t \geq 1$.
- (iv) \mathbf{H} , Γ , and \mathbf{D} are known.
- (v) The parameter set Θ is therefore comprised of $\{\Phi, \mathbf{E}, \mathbf{C}, \mathbf{Q}, \mathbf{R}, \boldsymbol{\mu}, \mathbf{P}\}$, with $\boldsymbol{\theta}_0$ belonging to the interior of Θ , assumed to be compact.
- (vi) The inputs $\{\mathbf{u}_t\}$ are non-stochastic and uniformly bounded, meaning that there exist real numbers c_1 and c_2 such that $c_1 \leq \mathbf{u}_t^T \mathbf{u}_t \leq c_2$, for all $t \geq 0$.
- (vii) There exists a constant c_0 such that $N^{-1} \mathbf{J}_N(\boldsymbol{\theta}_0) - c_0 \mathbf{I}_N$ is positive definite in the limit as $N \rightarrow \infty$, where \mathbf{I}_N denotes the identity matrix of dimension N .

This is obviously a fairly restricted situation, particularly (iv), but it is important to realize that these are a set of *sufficient* (not *necessary*) conditions; they do not exclude other less restrictive sets of conditions. Condition (ii), independence of the errors in state and observation equations, seems quite restrictive, but it really is not when one realizes that any model formulation with $\mathbf{S} \neq \mathbf{0}$ can be transformed into an observationally equivalent formulation with $\mathbf{S} = \mathbf{0}$; see Anderson and Moore [6, ch. 5]. Condition (vi) excludes the possibility of having time trends; while (vii) insists that the sequence of normalized information matrices be bounded from below by a positive definite matrix.

Generalizing the above results to a model where the input matrices Γ and \mathbf{D} are unknown, is analytically complex and requires further assumptions on the properties

of the inputs. In particular, Caines [37] discussed the asymptotic properties of $\hat{\theta}_N$ under the assumptions of deterministic and stochastic inputs. Under general conditions, he proves consistency and asymptotic normality for these estimates. An analogous analysis was done by Hannan and Deistler [116, ch. 4], in a slightly more general framework, as they avoided the hypothesis of compactness.

A bootstrap procedure can be implemented in situations where: (a) the asymptotics may not hold, (b) specific (non-Gaussian) distributions for the noise terms are entertained, or (c) small sample sizes deliver poor asymptotic performance. An instance of the latter is detailed by Stoffer and Wall [201].

B.2.3 Choice of estimation method

A natural generalization of ML is to use the ML criterion as the objective function to maximize, but not assume a distributional structure for the model noise terms. This is what Caines [37, ch. 8] calls the “minimum prediction error” identification method, where the term is broadly applied in reference to a particular *objective function* (computing terminology) or *loss function* (statistical terminology), $\ell(\theta)$, that one chooses to optimize (minimize) in order to produce parameter estimates. In this regard, the innovations (prediction errors) $\tilde{z}_t(\theta) \sim (0, B_t(\theta))$ are a natural starting point, since they measure the discrepancies between the realized values of the process and its optimal (mean squared error sense) one-step ahead predictions.

“Good” loss functions might then seek to minimize the sum of these discrepancies, leading to the ubiquitous *least squares* criterion,

$$\ell_1(\theta) = \sum_{t=1}^N \tilde{z}_t(\theta)^T \tilde{z}_t(\theta).$$

Since the innovations have different variances in general, it may be more appealing to minimize the *generalized least squares*, by standardizing the innovations to have unit variance,

$$\ell_2(\theta) = \sum_{t=1}^N \tilde{z}_t(\theta)^T B_t(\theta)^{-1} \tilde{z}_t(\theta).$$

Employing the ML principle with \tilde{z}_t Gaussian, adds additionally a log determinant term in ℓ_2 , so that one is lead to the objective function,

$$\ell_3(\theta) = \sum_{t=1}^N \{ \log |B_t(\theta)| + \tilde{z}_t(\theta)^T B_t(\theta)^{-1} \tilde{z}_t(\theta) \},$$

given in (B.10). However, note that any one of $\{\ell_1, \ell_2, \ell_3\}$ is a sensible loss function, leading to different estimators with (possibly) distinct asymptotic properties. In a quest for “robustness,” one then tries to establish the relevant asymptotics under more lenient assumptions that do **not** include specifying a distributional structure for \tilde{z}_t (e.g., Gaussian), thereby freeing one from making unrealistic assumptions¹.

¹The resulting estimates under this paradigm where one does not fully specify the likelihood, are sometimes known as *pseudo-ML* or *quasi-ML* estimates.

Several such results for ARMA models are well-known in the literature; e.g., Davis and Dunsmuir [64] established consistency and asymptotic normality under minimization of a least absolute deviations criterion, Brockwell and Davis [32, ch. 8] do so for the ML criterion ℓ_3 above, while Barnard, Trindade, and Wickramasinghe [14] shed light on the robustness of ℓ_3 with regard to the exponential power family of distributions.

For SS models, under this paradigm (and appropriate regularity conditions) it is often still possible to recover asymptotic properties (B.2)–(B.4). The situation when ℓ_1 and ℓ_2 are used as loss functions is discussed in Caines [37, ch. 6], while that for ℓ_3 is detailed in Caines [37, ch. 7] (an overview of which we have strived to give in this section); see also Ljung [152].

In essence, the quest for robustness involves a trade-off in relinquishing a particular distributional assumption (e.g., normality) at the expense of asymptotic efficiency, as the variance of the resulting estimator is no longer guaranteed to attain the Cramer-Rao lower bound. For example, the asymptotic variance of the estimator under ℓ_3 loss and Gaussian errors is known to be the smallest it can possibly be (in the class of all asymptotically normal estimators), if the errors are *truly* Gaussian. However, if the errors do not follow a Gaussian distribution, then the asymptotic variance of the estimator under ℓ_3 loss may well be larger.

B.3 The State-Space model with cross-sectional extension

Unlike Chapter 5 where there is only one asymptotic regime as the longitudinal sample size $N = n_1 \rightarrow \infty$, here we may also take into account the limiting behavior of the ML estimators as the cross-sectional sample size $k \rightarrow \infty$. (The most general situation where both the longitudinal and cross-sectional sample sizes grow will not be discussed here, as it does not appear to have yet been fully investigated.) We will also simplify the panel structure in order to have a “balanced” layout: $n_i = n$ for each $i = 1, \dots, k$. Under this paradigm we then have that the effective sample size, denoted generically by N , will be $N = n$ under the increasing longitudinal sample size situation, and $N = k$ under the increasing cross-sectional sample size regime. As we shall see shortly, and for the simplified model (B.21)–(B.22) below, the same asymptotic behavior is obtained in either regime.

To avoid overly technical regularity conditions we will restrict attention to the SS model without exogenous inputs and constant transition matrices:

$$\boldsymbol{x}_{i,t+1} = \boldsymbol{\Phi}\boldsymbol{x}_{i,t} + \boldsymbol{w}_{i,t+1}, \quad (\text{B.21})$$

$$\boldsymbol{z}_{i,t} = \boldsymbol{H}\boldsymbol{x}_{i,t} + \boldsymbol{v}_{i,t}, \quad (\text{B.22})$$

for $i = 1, \dots, k$ and $t = 1, \dots, n$, with corresponding error covariances:

$$\mathbb{E} \left(\begin{bmatrix} \boldsymbol{w}_{i,t+1} \\ \boldsymbol{v}_{i,t} \end{bmatrix} \begin{bmatrix} \boldsymbol{w}_{i,t+1}^T & \boldsymbol{v}_{i,t}^T \end{bmatrix} \right) = \begin{bmatrix} \boldsymbol{Q} & \boldsymbol{S} \\ \boldsymbol{S}^T & \boldsymbol{R} \end{bmatrix},$$

and initial condition $\boldsymbol{x}_{i,0} \sim N(\boldsymbol{\mu}, \boldsymbol{P})$. The estimator for the set of parameters $\Theta = \{\boldsymbol{\mu}, \boldsymbol{P}, \boldsymbol{\Phi}, \boldsymbol{Q}, \boldsymbol{R}, \boldsymbol{S}\}$ we will be concerned with is the ML estimator $\hat{\theta}_N$ discussed in

Chapter 11, obtained under the usual Gaussian optimality criterion. If $\boldsymbol{\theta} \in \Theta$ is a vectorization of the parameters (d -dimensional), the negative log-likelihood expression from (11.32) is given by

$$\begin{aligned}\ell(\boldsymbol{\theta}) := -\log L(\boldsymbol{\theta}) &= -\frac{1}{2} \sum_{i=1}^k \sum_{t=1}^n \left\{ \log |\mathbf{B}_{i,t}(\boldsymbol{\theta})^{-1}| - \tilde{\mathbf{z}}_{i,t}(\boldsymbol{\theta})^T \mathbf{B}_{i,t}(\boldsymbol{\theta})^{-1} \tilde{\mathbf{z}}_{i,t}(\boldsymbol{\theta}) \right\} \\ &= \frac{1}{2} \sum_{i=1}^k \sum_{t=1}^n \left\{ \log |\mathbf{B}_t(\boldsymbol{\theta})| + \tilde{\mathbf{z}}_{i,t}(\boldsymbol{\theta})^T \mathbf{B}_t(\boldsymbol{\theta})^{-1} \tilde{\mathbf{z}}_{i,t}(\boldsymbol{\theta}) \right\} \quad (\text{B.23})\end{aligned}$$

$$:= \frac{1}{2} \sum_{i=1}^k \sum_{t=1}^n v_{i,t}(\boldsymbol{\theta}), \quad (\text{B.24})$$

where the innovations are as defined in (11.33). Note that the structure of the model (B.21)–(B.22) implies the simpler structure for the distribution of the innovations, $\tilde{\mathbf{z}}_{i,t}(\boldsymbol{\theta}) \sim \text{NID}(\mathbf{0}, \mathbf{B}_t(\boldsymbol{\theta}))$, so that we have made the substitution $\mathbf{B}_{i,t}(\boldsymbol{\theta}) \equiv \mathbf{B}_t(\boldsymbol{\theta})$ in (B.23).

Calculation of the expected Hessian for $\ell(\boldsymbol{\theta})$ now follows, similar to the arguments given in Section B.2:

$$\mathbf{J}_N(\boldsymbol{\theta}) = \frac{1}{2} \mathbb{E} \left[\sum_{i=1}^k \sum_{t=1}^n \frac{\partial^2 v_{i,t}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right] = \frac{k}{2} \sum_{t=1}^n \mathbb{E} \left[\frac{\partial^2 v_t(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T} \right],$$

and by using the fact that the moments of the innovations vectors are the same over the index i . Now note that the summand in the above is identical to that of (B.16), and thus we obtain a correspondingly similar form,

$$\begin{aligned}\mathbf{J}_N(\boldsymbol{\theta}) &= \frac{k}{2} \sum_{t=1}^n \left[\frac{\partial \text{vec}(\mathbf{B}_t)^T}{\partial \boldsymbol{\theta}} (\mathbf{B}_t^{-1} \otimes \mathbf{B}_t^{-1}) \frac{\partial \text{vec}(\mathbf{B}_t)}{\partial \boldsymbol{\theta}^T} + 2\mathbb{E} \left(\frac{\partial \tilde{\mathbf{z}}_t^T}{\partial \boldsymbol{\theta}} \mathbf{B}_t^{-1} \frac{\partial \tilde{\mathbf{z}}_t}{\partial \boldsymbol{\theta}^T} \right) \right] \\ &= \frac{k}{2} \boldsymbol{\Sigma}_n(\boldsymbol{\theta}). \quad (\text{B.25})\end{aligned}$$

Similar regularity conditions to those given in Section B.2 can now be imposed to guarantee that the asymptotic results (B.2)–(B.5) will hold. Among these, Caines [37, ch. 9] cites observability, controllability (and thus minimality), and identifiability as key. Gaussian innovations can again be used as a proxy for more stringent requirements relating to minimum prediction error estimation. Under additionally nonsingularity of the expected Hessian (B.25), we obtain in particular,

1. for the increasing longitudinal sample size $N = n \rightarrow \infty$ situation:

$$\sqrt{n} (\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \left[\frac{k \boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)}{2n} \right]^{-1} \right), \quad (\text{B.26})$$

2. and for the increasing cross-sectional sample size $N = k \rightarrow \infty$ situation:

$$\sqrt{k} (\hat{\boldsymbol{\theta}}_N - \boldsymbol{\theta}_0) \xrightarrow{d} \mathcal{N} \left(\mathbf{0}, \left[\frac{\boldsymbol{\Sigma}_n(\boldsymbol{\theta}_0)}{2} \right]^{-1} \right). \quad (\text{B.27})$$

Note therefore that, under either regime, we obtain the same approximate large sample behavior for $\hat{\theta}_N$:

$$\hat{\theta}_N \approx N\left(\theta_0, \frac{2\Sigma_n(\theta_0)^{-1}}{k}\right).$$

An equivalent expression for the Hessian that emphasizes dependence on second order moments, is to vectorize the data longitudinally by concatenating the series cross-sectionally as $\mathbf{z}_i = (z_{i,1}, \dots, z_{i,n})^T$, $i = 1, \dots, k$. If $\mathbf{z}_i \sim N(\boldsymbol{\mu}_z, \mathbf{M}_z)$ for every $i = 1, \dots, k$, and we take the usual ML estimates for $\boldsymbol{\mu}_z$ and \mathbf{M}_z ² as

$$\bar{\mathbf{z}} = \frac{1}{k} \sum_{i=1}^k \mathbf{z}_i, \quad \text{and} \quad \widehat{\mathbf{M}_z} = \frac{1}{k} \sum_{i=1}^k (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T,$$

we can write a scaled version of the negative log-likelihood as

$$\frac{2\ell(\theta)}{k} = \log |\mathbf{M}_z| + \frac{1}{k} \sum_{i=1}^k (\mathbf{z}_i - \boldsymbol{\mu}_z)^T \mathbf{M}_z^{-1} (\mathbf{z}_i - \boldsymbol{\mu}_z).$$

By “completing the square” on the quadratic form above we obtain,

$$\begin{aligned} & \frac{1}{k} \sum_{i=1}^k (\mathbf{z}_i - \boldsymbol{\mu}_z)^T \mathbf{M}_z^{-1} (\mathbf{z}_i - \boldsymbol{\mu}_z) \\ &= \frac{1}{k} \sum_{i=1}^k [(\mathbf{z}_i - \bar{\mathbf{z}}) + (\bar{\mathbf{z}} - \boldsymbol{\mu}_z)]^T \mathbf{M}_z^{-1} [(\mathbf{z}_i - \bar{\mathbf{z}}) + (\bar{\mathbf{z}} - \boldsymbol{\mu}_z)] \\ &= \frac{1}{k} \sum_{i=1}^k (\mathbf{z}_i - \bar{\mathbf{z}})^T \mathbf{M}_z^{-1} (\mathbf{z}_i - \bar{\mathbf{z}}) + (\bar{\mathbf{z}} - \boldsymbol{\mu}_z)^T \mathbf{M}_z^{-1} (\bar{\mathbf{z}} - \boldsymbol{\mu}_z), \end{aligned}$$

since the cross terms cancel. Finally, noting that each term in the above summand is a scalar (and therefore trivially equal to its trace), we have from basic properties of the trace operator that,

$$\begin{aligned} \frac{1}{k} \sum_{i=1}^k \text{tr} [(\mathbf{z}_i - \bar{\mathbf{z}})^T \mathbf{M}_z^{-1} (\mathbf{z}_i - \bar{\mathbf{z}})] &= \frac{1}{k} \sum_{i=1}^k \text{tr} [\mathbf{M}_z^{-1} (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T] \\ &= \text{tr} \left[\mathbf{M}_z^{-1} \frac{1}{k} \sum_{i=1}^k (\mathbf{z}_i - \bar{\mathbf{z}})(\mathbf{z}_i - \bar{\mathbf{z}})^T \right] \\ &= \text{tr} \left(\mathbf{M}_z^{-1} \widehat{\mathbf{M}_z} \right), \end{aligned}$$

whence we obtain

$$\frac{2\ell(\theta)}{k} = \log |\mathbf{M}_z| + \text{tr} \left(\mathbf{M}_z^{-1} \widehat{\mathbf{M}_z} \right) + (\bar{\mathbf{z}} - \boldsymbol{\mu}_z)^T \mathbf{M}_z^{-1} (\bar{\mathbf{z}} - \boldsymbol{\mu}_z) := \eta(\theta). \quad (\text{B.28})$$

²Noting the dependence on the parameters one should properly write $\boldsymbol{\mu}_z(\theta)$ and $\mathbf{M}_z(\theta)$, but this detail will be sacrificed in favor of notational expedience.

Calculation of the Hessian for $\ell(\boldsymbol{\theta})$ now relies on derivatives for $\eta(\boldsymbol{\theta})$, given in Lemma 2 below. Under the increasing cross-sectional sample size regime, consistency of the sample mean and covariance estimates implies that

$$\operatorname{plim}_{k \rightarrow \infty} \bar{z} = \boldsymbol{\mu}_z, \quad \text{and} \quad \operatorname{plim}_{k \rightarrow \infty} \widehat{\mathbf{M}}_z = \mathbf{M}_z.$$

Taking the limits in k of the Hessian results in substantial simplification, and thus this is the preferred route in obtaining a simple expression for the asymptotic information matrix, i.e.

$$\mathcal{I}(\boldsymbol{\theta}) = \operatorname{plim}_{k \rightarrow \infty} \frac{1}{2} \frac{\partial^2 \eta(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T}.$$

Lemma 2 *For the expression $\eta(\boldsymbol{\theta})$ as defined in (B.28), the (p, q) element of the limiting Hessian matrix is given by*

$$\operatorname{plim}_{k \rightarrow \infty} \frac{\partial^2 \eta(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q} = 2 \frac{\partial \boldsymbol{\mu}_z^T}{\partial \theta_p} \mathbf{M}_z^{-1} \frac{\partial \boldsymbol{\mu}_z}{\partial \theta_q} + \operatorname{tr} \left(\mathbf{M}_z^{-1} \frac{\partial \mathbf{M}_z}{\partial \theta_q} \mathbf{M}_z^{-1} \frac{\partial \mathbf{M}_z}{\partial \theta_p} \right).$$

Proof for Lemma 2

The result follows by routine matrix calculus similar to those used in the proof of Lemma 1. To simplify notation, we drop the z subscripts on $\boldsymbol{\mu}_z$ and \mathbf{M}_z , but introduce p and q subscripts to denote partial derivatives with respect to θ_p and θ_q . Thus, for example, $\boldsymbol{\mu}_p = \partial \boldsymbol{\mu} / \partial \theta_p$, $\boldsymbol{\mu}_{p,q} = \partial^2 \boldsymbol{\mu} / (\partial \theta_p \partial \theta_q)$, etc., and similarly for \mathbf{M} . With this, we obtain the following second derivatives term by term:

$$\begin{aligned} \frac{\partial^2 \eta(\boldsymbol{\theta})}{\partial \theta_p \partial \theta_q} &= \operatorname{tr} \left[\left(\mathbf{M}^{-1} - \mathbf{M}^{-1} \widehat{\mathbf{M}} \mathbf{M}^{-1} \right) \mathbf{M}_{p,q} - \mathbf{M}^{-1} \mathbf{M}_p \mathbf{M}^{-1} \mathbf{M}_q \right] \\ &\quad + \operatorname{tr} \left(\mathbf{M}^{-1} \mathbf{M}_q \mathbf{M}^{-1} \widehat{\mathbf{M}} \mathbf{M}^{-1} \mathbf{M}_p + \mathbf{M}^{-1} \widehat{\mathbf{M}} \mathbf{M}^{-1} \mathbf{M}_q \mathbf{M}^{-1} \mathbf{M}_p \right) \\ &\quad + 2 \boldsymbol{\mu}_p^T \mathbf{M}^{-1} \boldsymbol{\mu}_q - \boldsymbol{\mu}^T \mathbf{M}^{-1} \mathbf{M}_{p,q} \mathbf{M}^{-1} \boldsymbol{\mu} - \bar{z}^T \mathbf{M}^{-1} \mathbf{M}_{p,q} \mathbf{M}^{-1} \bar{z} \\ &\quad + 2(\boldsymbol{\mu}^T + \bar{z}^T) \mathbf{g}_1 + 2\bar{z}^T \mathbf{g}_2, \end{aligned}$$

where the vectors \mathbf{g}_1 and \mathbf{g}_2 are given by:

$$\begin{aligned} \mathbf{g}_1 &= \mathbf{M}^{-1} \boldsymbol{\mu}_{p,q} + \mathbf{M}^{-1} \mathbf{M}_q \mathbf{M}^{-1} \boldsymbol{\mu}_p + \mathbf{M}^{-1} \mathbf{M}_p \mathbf{M}^{-1} \boldsymbol{\mu}_q \\ &\quad - \mathbf{M}^{-1} \mathbf{M}_q \mathbf{M}^{-1} \mathbf{M}_p \mathbf{M}^{-1} \boldsymbol{\mu}, \end{aligned}$$

and

$$\begin{aligned} \mathbf{g}_2 &= \mathbf{M}^{-1} \mathbf{M}_{p,q} \mathbf{M}^{-1} \boldsymbol{\mu} + \mathbf{M}^{-1} \mathbf{M}_q \mathbf{M}^{-1} \mathbf{M}_p \mathbf{M}^{-1} \bar{z} \\ &\quad - \mathbf{M}^{-1} \mathbf{M}_p \mathbf{M}^{-1} \mathbf{M}_q \mathbf{M}^{-1} \boldsymbol{\mu}. \end{aligned}$$

Taking limits as $k \rightarrow \infty$ and noting the consistency of the sample mean and covariance, leads to much cancellation of terms and the simple expression stated in the lemma. ■

With this setup we then obtain, under the $k \rightarrow \infty$ regime,

$$\sqrt{k}(\hat{\theta}_N - \theta_0) \xrightarrow{d} N(\mathbf{0}, \mathcal{I}(\theta)^{-1}),$$

where the (p, q) element of the asymptotic information matrix is

$$\begin{aligned} \mathcal{I}(\theta)_{p,q} &= \frac{\partial \mu_z(\theta)^T}{\partial \theta_p} M_z(\theta)^{-1} \frac{\partial \mu_z(\theta)}{\partial \theta_q} \\ &\quad + \frac{1}{2} \text{tr} \left[M_z(\theta)^{-1} \frac{\partial M_z(\theta)}{\partial \theta_q} M_z(\theta)^{-1} \frac{\partial M_z(\theta)}{\partial \theta_p} \right]. \end{aligned}$$

Appendix C

Software (E^4)

C.1	Models supported in E^4	224
C.1.1	State-Space model	224
C.1.2	The THD format	224
C.1.3	Basic models	226
C.1.3.1	Mathematical definition	226
C.1.3.2	Definition in THD format	227
C.1.4	Models with GARCH errors	228
C.1.4.1	Mathematical definition of the GARCH process	228
C.1.4.2	Defining a model with GARCH errors in THD format	229
C.1.5	Nested models	230
C.2	Overview of computational procedures	233
C.2.1	Standard procedures for time series analysis	233
C.2.2	Signal extraction methods	235
C.2.3	Likelihood and model estimation	238
C.3	Who can benefit from E^4 ?	241

This Appendix describes a MATLAB toolbox for time series modeling. Its name, E^4 , refers to the Spanish: *Estimación de modelos Econométricos en Espacio de los Estados*, meaning “State-Space Estimation of Econometric Models.” It includes ready-to-use functions for model specification, parameter estimation, model forecasting, simulation, and signal extraction. In particular, all the examples in this book (except those of Chapter 11) have been computed with E^4 .

The structure of this Appendix is as follows. Section C.1 reviews the standard models supported (including e.g., SS, VARMAX or transfer functions), describes how to define them, and shows how a model can be created by joining several building blocks through the model combination procedures that described in subsections 3.2.1 and 3.2.2. Section C.2 provides an overview of the analytical and computational procedures employed by E^4 . They fall into three broad categories: standard time series methods, signal extraction procedures, and model estimation algorithms. Last,

Section C.3 provides some remarks about the design priorities of E^4 , and discusses what type of user might benefit from this toolbox.

The source code for all the functions in E^4 is freely provided under the terms of the GNU General Public License. Appendix D indicates how to download E^4 and its related materials, as well as the source code and data required to replicate the examples in this book.

C.1 Models supported in E^4

C.1.1 State-Space model

The most important formulation supported by E^4 is the SS model; the standard representation for which most of its core computational algorithms are devised. The basic SS formulations supported are the MEM and SEM forms defined in Sections 2.1 and 2.2.

On the other hand, SS models are difficult to store and manipulate through software. For this reason, E^4 manages all the models supported using a convenient format called THD (THeta-Din).

C.1.2 The THD format

Defining a model in THD format requires creating two numeric matrices: `theta` and `din`, which contain respectively, the parameter values and a description of the model dynamic structure. Additionally, the model can be documented through an optional character matrix, `lab`, which contains names for the parameters in `theta`. While some analyses may require editing `theta` and `lab`, most users will never need to touch the matrix `din`.

The THD representation of an SS model can be obtained via the function `ss2thd` (SS-to-THD), which translates the matrices characterizing any MEM or SEM form into the THD format. Its syntax is:

```
[theta, din, lab] = ss2thd(Phi, Gam, E, H, D, C, Q, S, R);
```

where the input arguments are the parameter matrices in the SS model, and the outputs are the `theta`, `din`, and `lab` matrices previously described. To illustrate the use of this function, the following Example shows how to define the SS model implied by a Hodrick–Prescott filter (see Harvey and Trimbur [120]):

Example C.1.1.1 Defining a SS model in THD format

```
e4init % Initializes E4 default options and tolerances
% *** Defines the model matrices
Phi= [1 1; NaN 1];
% The Not-A-Number value (NaN) indicates a null parameter
E= [1;NaN];
H=[1 0]; C=[1]; Q=[1/1600]; S=[0]; R=[1];
% *** Obtains the THD representation and displays
```

```
%      the resulting model
[theta, din, lab] = ss2thd(Phi, [], E, H, [], C, Q, S, R);
% The only free parameters in this model are the variances.
% Accordingly, nonzero values in the second column of theta
% indicate which parameters should remain at its current value
theta(1,2)=1; theta(2,2)=1; theta(3,2)=1;
theta(4,2)=1; theta(5,2)=1; theta(6,2)=1;
theta(7,2)=1; theta(9,2)=1;
prtmod(theta, din, lab);
```

and the resulting output is:

```
***** Options set by user *****
Filter. . . . . : KALMAN
Scaled B and M matrices . . . . : NO
Initial state vector. . . . . : AUTOMATIC SELECTION
Initial covariance of state v. . : IDEJONG
Variance or Cholesky factor? . . : VARIANCE
Optimization algorithm. . . . . : BFGS
Maximum step length . . . . . : 0.100000
Stop tolerance. . . . . . : 0.000010
Max. number of iterations . . . : 75
Verbose iterations. . . . . : YES
*****
```

```
***** Model *****
Native SS model
1 endogenous v., 0 exogenous v.
Seasonality: 1
SS vector dimension: 2
Parameters (* denotes constrained parameter):
PHI(1,1)      *    1.0000
PHI(1,2)      *    1.0000
PHI(2,2)      *    1.0000
E(1,1)        *    1.0000
H(1,1)        *    1.0000
H(1,2)        *    0.0000
C(1,1)        *    1.0000
Q(1,1)        0.0006
S(1,1)        *    0.0000
R(1,1)        1.0000
*****
```

Note that asterisks identify which parameters are constrained to remain at their current values in any subsequent estimation.

C.1.3 Basic models

C.1.3.1 Mathematical definition

E^4 supports three basic models: VARMAX, structural econometric models, and single-output transfer functions. The VARMAX model is given by:

$$\phi_p(B) \Phi_P(B^S) z_t = G_n(B) u_t + \theta_q(B) \Theta_Q(B^S) a_t, \quad (C.1)$$

where S denotes the period of the seasonal cycle, B is the (already defined) back-shift operator, the vectors z_t , u_t , and a_t are the model outputs, inputs, and errors, respectively, and the polynomial matrices in (C.1) are given by:

$$\phi_p(B) = I + \phi_1 B + \phi_2 B^2 + \dots + \phi_p B^p \quad (C.2)$$

$$\Phi_P(B^S) = I + \Phi_1 B^S + \Phi_2 B^{2 \cdot S} + \dots + \Phi_P B^{P \cdot S} \quad (C.3)$$

$$G_n(B) = G_1 B + G_2 B^2 + \dots + G_n B^n \quad (C.4)$$

$$\theta_q(B) = I + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (C.5)$$

$$\Theta_Q(B^S) = I + \Theta_1 B^S + \Theta_2 B^{2 \cdot S} + \dots + \Theta_Q B^{Q \cdot S}. \quad (C.6)$$

The structural econometric model is similar to the VARMAX specification (C.1), but allows for contemporary relationships between the endogenous variables and their errors. In econometric terms, it is a “structural form” while the VARMAX is a “reduced form.” It is defined by:

$$\bar{\phi}_p(B) \Phi_P(B^S) z_t = G_n(B) u_t + \bar{\theta}_q(B) \Theta_Q(B^S) a_t, \quad (C.7)$$

with:

$$\bar{\phi}_p(B) = \bar{\phi}_0 + \bar{\phi}_1 B + \bar{\phi}_2 B^2 + \dots + \bar{\phi}_p B^p \quad (C.8)$$

$$\bar{\theta}_q(B) = \bar{\theta}_0 + \bar{\theta}_1 B + \bar{\theta}_2 B^2 + \dots + \bar{\theta}_q B^q. \quad (C.9)$$

(Note that the polynomial matrices $\Phi_P(B^S)$, $G_n(B)$, and $\Theta_Q(B^S)$ in (C.7) were already defined in (C.1).)

Last, the single-output transfer function with r inputs is given by:

$$z_t = \frac{\omega_1(B)}{\delta_1(B)} u_{1,t} + \dots + \frac{\omega_r(B)}{\delta_r(B)} u_{r,t} + \frac{\theta_q(B) \Theta_Q(B^S)}{\phi_p(B) \Phi_P(B^S)} a_t, \quad (C.10)$$

where, for $i = 1, 2, \dots, r$, we have:

$$\omega_i(B) = \omega_0 + \omega_1 B + \dots + \omega_{m_i} B^{m_i} \quad (C.11)$$

$$\delta_i(B) = 1 + \delta_1 B + \dots + \delta_{k_i} B^{k_i}, \quad (C.12)$$

and the polynomials $\theta_q(B)$, $\Theta_Q(B^S)$, $\phi_p(B)$, and $\Phi_P(B^S)$ in (C.10) are scalar versions of the autoregressive and moving average matrix factors in (C.1).

C.1.3.2 Definition in THD format

The E^4 functions `arma2thd`, `str2thd` and `tf2thd` generate the THD formulation for VARMAX, structural econometric models and transfer functions, respectively. The general syntax of these functions is:

```
[theta, din, lab] = arma2thd(FR, FS, AR, AS, V, s, G, r)
[theta, din, lab] = str2thd(FR, FS, AR, AS, V, s, G, r)
[theta, din, lab] = tf2thd(FR, FS, AR, AS, V, s, W, D)
```

where the inputs to these functions are matrices of real numbers containing: (a) the initial values for the regular and seasonal autoregressive factors, `FR` and `FS`, (b) the regular and seasonal moving average factors, `AR` and `AS`, (c) the terms related to the exogenous variables in the transfer function model, `W` and `D`, and in the VARMAX and structural econometric model, `G`, (d) the noise covariance, `V`, as well as (e) scalar integers defining the seasonal period, `s`, and the number of inputs in the VARMAX and structural econometric models, `r`.

The following Example illustrates the use of `arma2thd` by defining a quarterly airline model in THD form, displaying its structure and obtaining the matrices in its SS representation:

Example C.1.1.2 ARIMA specification in THD and SS format

```
e4init % Initializes E4 default options and tolerances
% Defines and displays the non-stationary airline model
% (1-B)(1-B^4) y_t = (1 - .6 B)(1 - .5 B^4) a_t
[theta,din,lab] = arma2thd([-1],[-1],[-.6],[-.5],[.1],4);
theta(1,2)=1; theta(2,2)=1; % Constrains the unit roots
prtmod(theta,din,lab);
% Now obtains the matrices of the equivalent SS representation
[Phi, Gam, E, H, D, C, Q, S, R] = thd2ss(theta, din)
```

and the resulting output is:

```
***** Model *****
VARMAX model (innovations model)
1 endogenous v., 0 exogenous v.
Seasonality: 4
SS vector dimension: 5
Parameters (* denotes constrained parameter):
FR1(1,1)      *     -1.0000
FS1(1,1)      *     -1.0000
AR1(1,1)        -0.6000
AS1(1,1)        -0.5000
V(1,1)          0.1000
*****
Phi =
    1      1      0      0      0
```

0	0	1	0	0
0	0	0	1	0
1	0	0	0	1
-1	0	0	0	0

Gam =
[]

E =
0.4000
0
0
0.5000
-0.7000

H =
1 0 0 0 0

D =
[]

C =
1

Q =
0.1000

S =
0.1000

R =
0.1000

C.1.4 Models with GARCH errors

C.1.4.1 Mathematical definition of the GARCH process

E^4 allows one to combine any SEM, VARMAX, structural model or transfer function with a vector GARCH process, for a survey on this important issue, see Bollerslev, Engle, and Nelson [25].

The general conditional heteroscedastic process is defined by the distributions $a_t \sim \text{IID}(\mathbf{0}, Q)$ and $a_t | \Omega_{t-1} \sim \text{IID}(\mathbf{0}, \Sigma_t)$ where, according to the notation already defined in Chapter 4, Ω_{t-1} denotes the information available up to $t - 1$. In a GARCH model, the conditional covariances, Σ_t , are such that:

$$[\mathbf{I} - \mathbf{B}_m(\mathbf{B})] \text{vech}(\Sigma_t) = \mathbf{w} + \mathbf{A}_k(\mathbf{B}) \text{vech}(a_t a_t^T), \quad (\text{C.13})$$

where $\text{vech}()$ vectorizes the upper triangle and main diagonal of its argument and the polynomial matrices are given by:

$$\mathbf{B}_m(\mathbf{B}) = \mathbf{B}_1\mathbf{B} + \mathbf{B}_2\mathbf{B}^2 + \dots + \mathbf{B}_m\mathbf{B}^m \quad (\text{C.14})$$

$$\mathbf{A}_k(\mathbf{B}) = \mathbf{A}_1\mathbf{B} + \mathbf{A}_2\mathbf{B}^2 + \dots + \mathbf{A}_k\mathbf{B}^k. \quad (\text{C.15})$$

E^4 manages model (C.13) in its alternative VARMAX representation. To derive it, consider the white noise process $\mathbf{v}_t = \text{vech}(\mathbf{a}_t\mathbf{a}_t^T) - \text{vech}(\boldsymbol{\Sigma}_t)$, which has a null conditional mean and a very complex heteroscedastic structure. Then $\text{vech}(\boldsymbol{\Sigma}_t) = \text{vech}(\mathbf{a}_t\mathbf{a}_t^T) - \mathbf{v}_t$, and substituting this expression in (C.13) and rearranging some terms, yields:

$$[\mathbf{I} - \mathbf{A}_k(\mathbf{B}) - \mathbf{B}_m(\mathbf{B})] \text{vech}(\mathbf{a}_t\mathbf{a}_t^T) = \mathbf{w} + [\mathbf{I} - \mathbf{B}_m(\mathbf{B})] \mathbf{v}_t. \quad (\text{C.16})$$

Equation (C.16) defines a VARMAX model for $\text{vech}(\mathbf{a}_t\mathbf{a}_t^T)$, which can be written in the alternative form:

$$\text{vech}(\mathbf{a}_t\mathbf{a}_t^T) = \text{vech}(\mathbf{Q}) + \mathbf{N}_t, \quad (\text{C.17})$$

$$[\mathbf{I} + \bar{\mathbf{A}}_{\max(k,m)}(\mathbf{B})] \mathbf{N}_t = [\mathbf{I} + \bar{\mathbf{B}}_m(\mathbf{B})] \mathbf{v}_t, \quad (\text{C.18})$$

where $\bar{\mathbf{A}}_{\max(k,m)}(\mathbf{B}) = -\mathbf{A}_k(\mathbf{B}) - \mathbf{B}_m(\mathbf{B})$ and $\bar{\mathbf{B}}_m(\mathbf{B}) = -\mathbf{B}_m(\mathbf{B})$. This is the formulation supported by E^4 . Note that this representation allows for IGARCH (Integrated-GARCH) components, which would require specifying some unit roots in the AR factor of (C.18).

C.1.4.2 Defining a model with GARCH errors in THD format

To define a model with GARCH errors in THD format it is necessary to obtain: (a) the THD formulation of the model for the mean, given by `t1, d1` and `lab1`, (b) the THD model corresponding to the model (ARCH, GARCH, or IGARCH) for the variance, given by `t2, d2` and `lab2`, and (c) link both formulations using the `garc2thd` function, which has the following syntax:

```
[theta, din, lab] = garc2thd(t1, d1, t2, d2, lab1, lab2);
```

Example C.1.1.3 Defining a regression model with GARCH errors

The following code generates and displays the THD representation of a regression model with GARCH errors:

```
e4init % Initializes E4 default options and tolerances

% Defines and displays a regression model with GARCH(1,1) errors:
%   y_t = .7 x_t1 + 1 x_t2 + a_t
%   a_t^2 = .1 + N_t
%   (1 - .8 B) N_t = (1 - .7 B) v_t
```

```
[t1, d1, lab1] = arma2thd([], [], [], [], [.1], 1, [.7 1], 2);
[t2, d2, lab2] = arma2thd([- .8], [], [- .7], [], [.1], 1);
[theta, din, lab] = garch2thd(t1, d1, t2, d2, lab1, lab2);
prtmod(theta, din, lab);
```

and the resulting output is:

```
***** Model *****
GARCH model (innovations model)
1 endogenous v., 2 exogenous v.
Seasonality: 1
SS vector dimension: 0
Endogenous variables model:
White noise model (innovations model)
1 endogenous v., 2 exogenous v.
Seasonality: 1
SS vector dimension: 0
Parameters (* denotes constrained parameter):
G0(1,1)          0.7000
G0(1,2)          1.0000
V(1,1)           0.1000
-----
GARCH model of noise:
VARMAX model (innovations model)
1 endogenous v., 0 exogenous v.
Seasonality: 1
SS vector dimension: 1
Parameters (* denotes constrained parameter):
FR1(1,1)         -0.8000
AR1(1,1)         -0.7000
-----
*****
```

C.1.5 Nested models

A “nested” model is built by combining several simple models. E^4 allows for two types of model combination: nesting in inputs and nesting in errors.

Nesting in inputs, or “endogeneization” was defined mathematically in Subsection 3.2.1. It consists in augmenting a SS model with the models for its exogenous variables which, after doing so, become endogenous variables. This operation is useful, for example, when one wants to combine a transfer function and a VARMAX model for its inputs to compute forecasts for all these variables in a single operation. Another situation where nesting is useful is when there are missing values in the inputs. “Endogeneizing” is then required because SS methods only allow for missing values in the endogenous variables; see Section 11.5.

On the other hand, nesting in errors was defined in Subsection 3.3.1. It consists in augmenting a model in SEM form with another model for its disturbances. Intuitively, it can be seen as defining a model by the product of different polynomials. This is handy, e.g., when one wants to separate the factors containing different types of roots, such as unit, complex, or real roots. It is also useful when modeling a series with several seasonal factors, corresponding to different periods.

The following examples illustrate how E^4 implements the definition of nested models.

Example C.1.1.4 Nesting in inputs

Given the models:

$$z_t = \frac{0.3 + 0.6B}{1 - 0.5B} u_t + \frac{1 - 0.8B}{1 - 0.6B + 0.4B^2} a_{1,t}, \quad \sigma_{a1}^2 = 1, \quad (\text{C.19})$$

$$(1 - 0.7B) u_t = a_{2,t}, \quad \sigma_{a2}^2 = 0.3, \quad (\text{C.20})$$

the code required to define and combine both models in inputs is:

```
e4init
```

```
% Obtains the THD representation for both models:
[t1,d1,l1]=tf2thd([- .6 .4],[],[- .8],[],[1.0],1,[ .3 .6],[- .5]);
[t2,d2,l2]=arma2thd(- .7,[],[],[],.3,1);

% Combines the models for the endogenous variable and the input
[theta, din, lab] = stackthd(t1, d1, t2, d2, l1, l2);
[theta, din, lab] = nest2thd(theta, din, 1, lab);
% The 3rd input argument to nest2thd determines the combination:
% 0 : nesting in errors, 1 : nesting in inputs
% and displays the resulting model structure
prtmod(theta,din,lab);
```

and the output from the `prtmod` command is:

```
***** Model *****
Nested model in inputs (innovations model)
2 endogenous v., 0 exogenous v.
Seasonality: 1
SS vector dimension: 4
Submodels:
{
    Transfer function model (innovations model)
    1 endogenous v., 1 exogenous v.
    Seasonality: 1
    SS vector dimension: 3
    Parameters (* denotes constrained parameter):
```

```

FR(1,1)           -0.6000
FR(1,2)           0.4000
AR(1,1)           -0.8000
W1(1,1)           0.3000
W1(2,1)           0.6000
D1(1,1)           -0.5000
V(1,1)            1.0000
-----
VARMAX model (innovations model)
1 endogenous v., 0 exogenous v.
Seasonality: 1
SS vector dimension: 1
Parameters (* denotes constrained parameter):
FR1(1,1)          -0.7000
V(1,1)            0.3000
-----
}

*****

```

Note that the final model has two endogenous variables, z_t and $u_{1,t}$, and no inputs.

Example C.1.1.5 Nesting in errors

Consider now the model:

$$(1 - 0.5B + 0.3B^2)(1 - B)(1 - B^{12})z_t = a_t, \quad \sigma_a^2 = 0.2, \quad (\text{C.21})$$

which has nonstationary

regular and seasonal AR factors. The following code defines and displays model (C.21) through nesting in errors.

```

e4init
% First obtains the THD representation of the factors
[t1, d1, l1] = arma2thd([-1], [-1], [], [], [0], 12);
[t2, d2, l2] = arma2thd([- .5 .3], [], [], [.2], 12);

% and then the stacked and nested models
[ts, ds, ls] = stackthd(t1, d1, t2, d2, l1, l2);
[theta, din, lab] = nest2thd(ts, ds, 0, ls);
% The 3rd input argument to nest2thd determines the combination:
% 0 : nesting in errors, 1 : nesting in inputs

% Constrains the regular and seasonal unit roots
theta(1,2)=1; theta(2,2)=1;
prtmod(theta, din, lab);
% The variance of the model t1-d1 is ignored
% so we can specify any arbitrary value

```

and the output from `prtmod` is:

```
***** Model *****
Nested model in errors (innovations model)
1 endogenous v., 0 exogenous v.
Seasonality: 12
SS vector dimension: 15
Submodels:
{
    VARMAX model (innovations model)
    1 endogenous v., 0 exogenous v.
    Seasonality: 12
    SS vector dimension: 13
    Parameters (* denotes constrained parameter):
    FR1(1,1)      *   -1.0000
    FS1(1,1)      *   -1.0000
    -----
    VARMAX model (innovations model)
    1 endogenous v., 0 exogenous v.
    Seasonality: 12
    SS vector dimension: 2
    Parameters (* denotes constrained parameter):
    FR1(1,1)      -0.5000
    FR2(1,1)      0.3000
    V(1,1)        0.2000
    -----
}
```

C.2 Overview of computational procedures

E^4 includes three basic groups of functions: (a) standard procedures for time series analysis, (b) signal extraction methods and (c) model estimation algorithms.

C.2.1 Standard procedures for time series analysis

The main standard procedures included are detailed in Table C.1. These functions cover basic needs such as plotting a time series, computing its autocorrelations or its descriptive statistics. They are included to make the Toolbox self-contained, but its functionality is not innovative.

The following example illustrates the use of these functions with simulated data. We will not show the outputs resulting from this and other simulation-based examples, because the exact results obtained will differ in each run.

Example C.2.1 ARIMA identification

Table C.1 Standard statistics procedures to specify and simulate a time series model.

Function	Description	References:
<code>augdft</code>	Computes the augmented Dickey-Fuller test	Dickey and Fuller [78]
<code>descser</code>	Descriptive statistics for a vector of time series	
<code>histser</code>	Computes and displays histograms for a vector of time series	
<code>lagser</code>	Generates lags and leads from a data matrix	
<code>midents</code>	Multiple sample autocorrelation and partial autoregression functions	Wei [221]
<code>plotsers</code>	Displays standardized plots for a vector of time series	
<code>rmedser</code>	Displays the mean/std. deviation plot of a vector of time series	Box and Cox [28]
<code>transdif</code>	Computes differences and Box-Cox transforms	Box and Cox [28]
<code>uidents</code>	Sample autocorrelation and partial autocorrelation functions	Box, Jenkins, and Reinsel [30]
<code>simmod</code>	Simulates a homoscedastic model	
<code>simgarch</code>	Simulates a model with GARCH errors	Bollerslev, Engle, and Nelson [25]

Consider the ARMA(2,1) model:

$$(1 - 0.5B + 0.3B^2) z_t = (1 - 0.7B) a_t, \quad \sigma_a^2 = 0.2. \quad (\text{C.22})$$

The following code defines it, simulates a sample, and performs a standard univariate identification process:

```
e4init
% Obtains the THD definition of the model
[theta, din, lab] = arma2thd([- .5 .3], [], [- .7], [], [.2], 1);
prtmod(theta,din,lab)

% Simulates the sample and omits the first 50 observations
z=simmod(theta,din,250); z=z(51:250,1);

% Applies several standard time series analysis tools
% Standard-deviation mean plot
rmedser(z);
% Time series plot
plotsers(z);
```

```
% Augmented Dickey-Fuller test. The second argument (2)
% specifies the number of lags for the dynamic regression
augdft(z,2);
% Histogram
histsers(z);
% Descriptive statistics
descser(z);
% Sample autocorrelation and partial autocorrelation functions
% The second argument specifies the number of lags
uidents(z,10);
```

Example C.2.2 VARMA identification

The following code illustrates the use of the multiple time series identification functions included in E^4 by simulating a VAR(1) process and then identifying it:

```
e4init
% Obtains the THD definition of the model
Phi=[-.5 .3;NaN -.6];
Sigma=[1 .5;.5 1];
[theta,din,lab]=arma2thd([Phi],[],[],[],[Sigma],1);
prtmod(theta,din,lab)

% Simulates the sample and omits the first 50 observations
y=simmod(theta,din,250); y=y(51:250,:);

% Descriptive statistics for multiple time series
descser(y);

% Computes 10 lags of the multiple autocorrelation, partial
% autocorrelation and cross-correlation functions
midents(y,10);
```

C.2.2 Signal extraction methods

As we saw in Chapter 4, the SS literature considers two main signal extraction problems: filtering and smoothing. Both have as their focus the estimation of the sequence of states, but based on different information sets. E^4 computes these sequences for different purposes. Filtered states are typically used to compute the Gaussian likelihood in prediction-error decomposition form (see Chapters 5 and 6), so that filtering algorithms are embedded in the likelihood evaluation functions of E^4 ; see Section C.2.3.

On the other hand, smoothed estimates have many uses, such as interpolating missing in-sample values (Kohn and Ansley [137]), calculating the residuals of a model allowing for missing values (Kohn and Ansley [138]), cleaning noise-contaminated samples (Kohn and Ansley [136]), decomposing a time series into

meaningful unobserved components (Harvey, [118]; Casals, Jerez, and Sotoca [41]), and detecting outliers (De Jong and Penzer [70]). Table C.2 summarizes the main E^4 functions implementing signal-extraction algorithms, with corresponding literature references where relevant.

Table C.2. *Signal extraction functions. These are the main procedures offered for time series disaggregation, structural decomposition, smoothing and forecasting.*

Function	Description	References
<code>aggrmod</code>	Computes smoothed estimates for a high-frequency time series from: (a) a sample with low and high-frequency data, and (b) the high-frequency model for the endogenous variable(s)	Casals, Sotoca and Jerez [46]
<code>e4trend</code>	Decomposes a vector of time series into the trend, seasonal, cycle and irregular components implied by an econometric model in THD format	Casals, Jerez, and Sotoca [41]
<code>fismod</code> , <code>fismiss</code>	Computes smoothed estimates of the state and observable variables of a model in THD form. The function <code>fismiss</code> allows for in-sample missing values	Casals, Jerez, and Sotoca [40]
<code>foremod</code>	Computes forecasts for a homoscedastic model in THD format	
<code>foregarc</code>	Computes forecasts for a model in THD format, allowing for GARCH errors	
<code>residual</code>	Computes the residuals for a model in THD format	

Basic Fixed-Interval Smoothing is addressed by the E^4 function `fismod`, which implements the algorithm of Casals, Jerez, and Sotoca [40]. Its syntax is:

```
[xhat, Px, e] = fismod(theta, din, z);
```

The input arguments of these functions are a THD format specification (`theta`, `din`) and the data matrix (`z`) where the missing values, if any, are coded as NaN (the “not-a-number” MATLAB special symbol). The output arguments of `fismod` are: `xhat`, the sequence of expected values of the state vector conditional on the sample; `Px`, a matrix with the corresponding covariances; and `e`, a matrix of fixed-interval smoothed errors. There is another smoother function, `fismiss`, which allows for missing values in `z`.

Smoothing is often used to decompose a time series into the sum of trend, cycle, seasonal, and irregular components. The function `e4trend` implements the exact decomposition described in Casals, Jerez, and Sotoca [41]. Its simplified syntax is:

```
[trend, season, cycle, irreg] = e4trend(theta, din, z)
```

where the input arguments `theta`, `din`, and `z` are identical to those of `fismod`, and the output arguments are fixed-interval smoothed estimates of the corresponding structural components.

The function `foremod` predicts future values of the endogenous variables for a given model. Its syntax is:

```
[zf, Bf] = foremod(theta, din, z, k, u)
```

where the input arguments `theta`, `din`, and `z` are identical to those of `fismod`, `k` is the number of forecasts to be computed, and `u` is a matrix with the out-of-sample values of the exogenous variables, if any. The forecasts and the corresponding sequence of covariances are returned in the output arguments `zf` and `Bf`. The analogous functions, `foremiss` and `foregarc`, compute forecasts for samples with missing values and for models with conditional heteroscedastic errors, respectively. Last, the residuals for a model are calculated through the function `residual` whose syntax is:

```
[z1, vT, wT, vz1, vvT, vwT] = residual(theta, din, z)
```

where the input arguments `theta`, `din` and `z` are identical to those of `fismod`. On the other hand, the output arguments are `z1`, `vT`, and `wT` which are, respectively, the model residuals, the fixed-interval smoothed estimates of the observation errors, and the smoothed estimates of the state errors. The arguments `vz1`, `vvT`, and `vwT` store the corresponding sequences of covariances.

Example C.2.3 HP filtering

The following code simulates the HP filter model defined in Example C.1.1.1 and extracts the HP trend using different procedures:

```
e4init
% Working with standard deviations improves the model scaling
sete4opt('var','fac');

Phi= [1 1; NaN 1];
E= [1;NaN];
H=[1 0]; C=[1]; Q=[sqrt(1/1600)]; S=[0]; R=[1];
% Obtains the THD representation and displays the model
[theta, din, lab] = ss2thd(Phi, [], E, H, [], C, Q, S, R);
prtmod(theta, din, lab);
% Simulates the data and omits the first 50 observations
y=simmod(theta,din,250); y=y(51:250,:);

% Extracts and plots the trend and error components
[trend,season,cycle,irreg] = e4trend(theta,din,y);
plotsers([trend,irreg]);

% e4trend does not provide variances for the trend component
```

```
% The following commands do this using fismod
[xhat, px, e] = fismod(theta, din, y); varhptrend=px(1:2:400,1);
figure; hold on
plot(varhptrend.^ .5)
hold off
```

C.2.3 Likelihood and model estimation

Table C.3 summarizes the main functions in connection with model estimation. These functions specialize in the computation of likelihood values, information matrices, and parameter estimates.

E^4 includes several functions to compute the Gaussian log-likelihood. The most advanced one is `lfsd`, whose general syntax is:

```
[f, innov, ssvect] = lfsd(theta, din, z)
```

where the input arguments are identical to those defined for previous functions, and the output arguments are: (a) `f`, the value of the log-likelihood, (b) `innov`, a matrix of one-step-ahead forecast errors, and (c) `ssvect`, a matrix of estimates of the state variables organized so that its t -th row contains the filtered estimate of the state vector at time t , conditional on sample values up to $t - 1$. The technical details of this function can be found in Casals, Sotoca and Jerez [47]. Other functions to compute the log-likelihood of a homoscedastic model are `lfmod` (Terceiro [206]), `lffast` (Casals, Sotoca and Jerez [44]), and `lfcfd` (Casals, Sotoca and Jerez [47]). They differ only in minor details, and all have the same inputs and outputs. Last, `lfmiss` and `lfgarch` are two specialized variants of `lfsd`, allowing for missing values in the endogenous variables and GARCH errors, respectively.

As in the case of the likelihood, there are three functions dealing with computation of the information matrix: `imod`, `imiss`, and `igarch`. They all have the same syntax. For example, any call to `imod` has the following structure:

```
[std, corrm, varm, Im] = imod(theta, din, z)
```

This function computes the information matrix of `lfmod` and `lffast` (`Im`) as well as the parameter analytical standard errors (`std`), the correlation matrix of the estimates (`corrm`), and the corresponding covariance matrix (`varm`). The functions `imiss` and `igarch` do the same for `lfmiss` and `lfgarch`, respectively. If the model is misspecified or is non-Gaussian, the function `imodg` computes a robust information matrix, alternative to `imod`; see Ljung and Caines [153] and White [226].

Being able to compute the Gaussian log-likelihood is not enough to estimate the parameters of a model: one needs an iterative procedure to compute the optimal value of the parameters. The E^4 function `e4min` implements two main optimization algorithms, BFGS and Newton–Raphson (Dennis and Schnabel [74]). Its general syntax is:

```
[pnew, iter, fnew, gnew, hessin] = ...
e4min(func, theta, dfunc, P1, P2, P3, P4, P5)
```

Table C.3 *Model estimation functions.* These are the main procedures which can be applied to compute the Gaussian likelihood, its first-order derivatives, and information matrix. These functions can be fed to the standard E⁴ optimizer (`e4min`) to compute ML estimates. Initial estimates can be computed using a specialized function (`e4preest`).

Function	Description	References
Likelihood computation and derivatives		
<code>lfmod</code> ,	Compute the log-likelihood function for	Terceiro [206],
<code>lffast</code> ,	a model in THD form. <code>lfmod</code> , <code>lffast</code> ,	Casals, Sotoca and
<code>lfsd</code> ,	<code>lfsd</code> and <code>lfcd</code> support models with ho-	Jerez [44], Casals,
<code>lfcd</code> ,	moscedastic errors. <code>lfper</code> supports peri-	Sotoca and Jerez
<code>lfper</code> ,	odic models. Last, <code>lfgarch</code> admits for	[47]
<code>lfgarch</code> ,	GARCH errors and <code>lfmiss</code> allows for in-	
<code>lfmiss</code>	sample missing values	
<code>gmod</code> ,	Computes the analytical gradient for	Casals, Sotoca and
<code>gmiss</code> ,	<code>lfmod</code> and <code>lffast</code> (<code>gmod</code>), <code>lfmiss</code>	Jerez [45], Casals,
<code>ggarch</code>	(<code>gmiss</code>) and <code>lfgarch</code> (<code>ggarch</code>)	and Sotoca [43]
<code>imod</code> ,	Compute the Gaussian analytical infor-	Terceiro [206],
<code>imiss</code> ,	mation matrix for <code>lfmod</code> , <code>lffast</code> , <code>lfsd</code>	Ljung and Caines
<code>igarch</code> ,	and <code>lfcd</code> (<code>imod</code>), <code>lfmiss</code> (<code>imiss</code>) and	[153], White [226]
<code>imodg</code>	<code>lfgarch</code> (<code>igarch</code>). The function <code>imodg</code>	
	computes the robustified information ma-	
	trix for <code>lfmod</code> and <code>lffast</code>	
Model estimation		
<code>e4min</code>	Computes the unconstrained numerical	Dennis and Schn-
	minimum of a nonlinear function using the	abel [74]
	BFGS or the Newton-Raphson algorithms	
<code>e4preest</code>	Computes a fast estimate of the parameters	Garcia-Hiernaux,
	for a model in THD format	Jerez, and Casals
		[104]
<code>prtest</code>	Displays estimation results	
Toolbox defaults and options		
<code>e4init</code>	Initializes the global variables and defaults	
<code>sete4opt</code>	Modifies the toolbox options	

The operation of `e4min` is as follows. Starting from the initial estimate of the parameters in `theta`, it iterates on the objective function whose name is stored in the argument `func`. The iteration can be based on the analytical gradient or on a numerical approximation, depending on whether `dfunc` contains the name of the analytical gradient function or is an empty string, '''. Finally, the stopping criterion takes into account the relative changes in the values of the parameters and/or the size of the gradient vector. The parameters `P1`, ..., `P5` are optional. If they are specified, its values are fed as additional input arguments to the objective function `func`.

Once the iterative process is stopped, the function returns: pnew, the value of the parameters; iter, the number of iterations performed; fnew, the value of the objective function at pnew; gnew, the analytical or numerical gradient, depending on the contents of dfunc; and finally hessin, a numerical approximation to the Hessian of the objective function.

The auxiliary function `sete4opt` manages different options controlling the performance of the optimizer, including type of algorithm, tolerance for the stopping criteria, or the maximum number of iterations allowed.

Another important function for model estimation is `e4preest`. It computes fast and consistent estimates of the parameters in theta which, in most cases, are adequate starting values for likelihood optimization. The syntax of this function is:

```
thetanew = e4preest(theta, din, z)
```

where the input arguments are identical to those of `lffast`, and the estimates are returned in `thetanew`. The algorithm implemented in `e4preest` is described by Garcia-Hiernaux, Jerez, and Casals [104].

The following code simulates, estimates and applies different signal-extraction procedures to the nonstationary

airline model structure defined in Example C.1.1.2:

Example C.2.4 Simulation, estimation and signal extraction

```
e4init
% Defines and displays the non-stationary airline model
% (1-B)(1-B^12) y_t = (1 - .6 B)(1 - .5 B^12) a_t
[theta,din,lab] = arma2thd([-1],[-1],[-.6],[-.5],.1,12);
theta(1,2)=1; theta(2,2)=1; % Constrains the unit roots

% Simulates the sample and omits the first 50 observations
z=simmod(theta,din,250); z=z(51:250,1);

% Computes preliminary estimates
theta=e4preest(theta,din,z);
% ... and then ML estimates
[thopt, it, lval, g, h] = e4min('lffast', theta, '', din, z);
% Computes the analytical standard errors
[std, corrm, varm, Im] = imod(thopt, din, z);
% Summary of estimation results
prtest(thopt, din, lab, z, it, lval, g, h, std, corrm)

% Residual diagnostics
ehat=residual(thopt, din, z);
descser(ehat);
plotsers(ehat);
uidents(ehat,39);
```

```
% Structural decomposition
[trend,season,cycle,irreg] = e4trend(thopt,din,z);
plotsers([trend,season,irreg]);

% Computes and displays 10 forecasts and their standard errors
[zfor, Bfor] = foremod(thopt, din, z, 10);
[zfor Bfor.^.5]

% Marks obs. 10 and 40 as missing and estimates them
z1=z; z1(10)=NaN; z1(40)=NaN;
[zhat, pz] = fismiss(thopt, din, z1);
[z(1:50) z1(1:50) zhat(1:50)]
```

C.3 Who can benefit from E^4 ?

E^4 combines the intellectual input from many people who faced special needs that commercial software could not satisfy. If these needs coincide with yours, perhaps you should consider using E^4 for your work.

First of all, we needed **transparency**. Statistical software often hides relevant computational outputs, such as the likelihood, gradient, or the information matrix conditioning number. This happens because these results convey one of two messages: either that everything worked as expected, or that there is a problem with the model, the data, or the software. Providing this information may not be the best way to do “business” but, for some users at least, they are important pieces of the puzzle.

Second, we needed **reliability**. Mainstream software tends to prioritize computational speed at the expense of numerical accuracy. This can be done in different ways such as, e.g., choosing fast but sloppy computational algorithms or easy-to-satisfy default convergence criteria. Speed is crucial for a real-time computational system. However, for academic work and many professional applications, speed is attractive but not so valuable when compared to precision or computational reliability.

Third, we needed **consistency**. Common time series packages implement different procedures to perform the same task for different models, and this may produce substantial inconsistencies. These inconsistencies are often due to the intrinsic difficulty of making a clean distinction between models and procedures. For example, a standard time series package may have a specific procedure to forecast an ARIMA model and another one for VARMA, without any guarantee that their outputs will be mutually consistent. E^4 avoids this shortcoming by translating internally all the models supported to the SS form, so that, for example, in the ARIMA and VARMA case, a unique forecasting procedure is employed.

Our fourth and last need was strictly personal: we wanted to exploit the software generated by research. Academic work often produces neat pieces of code which do a specific task well. This software could be useful to other people in an adequate framework, but if this framework does not exist, it will probably vanish forever into the “limbo” reserved for neat pieces of code after the paper is published. On the other

hand, a coordinated effort to put together, organize, and document these scientific by-products, may be academically profitable for both authors and end-users.

Our personal answer to these questions was E^4 , a MATLAB toolbox intended to harbor the collection of results from a research line in SS methods, whose origins extend back into the early 1980s.

E^4 is our main research tool and, accordingly, it prioritizes academic needs. As a consequence, its functions provide a transparent, well-documented, and user-configurable suite of outputs that provides many clues about the problem such as lack of convergence or ill-conditioning. Also, the code has been carefully optimized for numerical accuracy and robustness, making an intensive use of stability-inducing methods, such as factorization and scaling. Finally, computational tolerances are transparent, user-configurable, and have strict default values.

E^4 does all the important calculations using the equivalent SS form of the models employed. As noted before, this feature is important in enforcing consistency, since all the core algorithms are developed for a SS model. Therefore, when an E^4 function receives an input model, the first thing it does is obtain the equivalent SS representation, and only then does it apply the corresponding procedure. This approach cleanly separates models from procedures, assures a high degree of coherence in the results, simplifies code development, and, last but not least, makes the toolbox very easy to extend, because implementing new models with equivalent SS representations only requires coding new user interface functions. Conversely, a new procedure only needs to support the SS model, because, thanks to the bridge functions provided, it can be immediately applied to all the models supported.

E^4 has been an effective repository to organize and distribute the code resulting from our research. Every new research project we undertake adds new functionality to E^4 , which is documented not only in the user manual, but also, with exceptional depth, in the corresponding papers. In 2000 we launched a website (www.ucm.es/e-4/), containing the source code for the Toolbox as well as complete documentation. Appendix D provides more information about these issues.

Appendix D

Downloading E^4 and the examples in this book

D.1	The E^4 website	243
D.2	Downloading and installing E^4	243
D.3	Downloading the code for the examples in this book	245

D.1 The E^4 website

You can find the E^4 website at the URL: <https://www.ucm.es/e-4/>. This website distributes many E^4 -related materials, including the complete toolbox code, a user manual, as well as all the code and data required to replicate the examples in this book.

D.2 Downloading and installing E^4

To download E^4 you need to:

1. Access <https://www.ucm.es/e-4> using your WEB browser.
2. Click the “Downloads” item in the left-hand-side menu (alternatively, you can type the address <https://www.ucm.es/e-4/downloads> in the navigation bar of your browser).
3. Download the file `E4.zip` in a folder of your choice and extract the files contained in it. By doing so, you will obtain 150+ text files, most of them with the MATLAB `.m` extension, which constitute the core of E^4 code.
4. Optionally, download other optional materials such as: (a) a user manual for E^4 , (b) source code and data for the examples in the user manual, (c) source code and data for the examples in this book, as well as (d) the documentation for several groups of specialized functions which are distributed with E^4 but not covered by the user manual, and (e) beta-versions of some functions which are not yet included in the core of E^4 .

The only installation required consists of including the folder(s) with E^4 code in the MATLAB search path¹. To this end, you will need to select “Set Path” from the File menu in the initial MATLAB window and add the E^4 folder using the menu options. Alternatively, you can access the same menu by typing `path` at the MATLAB Command Window prompt. To edit this search path you may need to execute MATLAB with full administrative rights. Refer to your Operating System documentation, or to specialized Internet resources, to see how to do this.

To use the E^4 functions, you will also need to type the command:

```
e4init;
```

at least once at the beginning of each E^4 session. This command initializes the toolbox options and parameters, and its output is similar to:

```
EEEEEEEEEE 444 444
EEEEEEEEEE 444 444
EEE        44444444
EEE        44444444
EEEEEE     444
EEEEEE    444
EEE
EEE
EEEEEEEEEE
EEEEEEEEE
```

Toolbox for State Space Estimation of Econometric Models
Version APR-2013

Web: www.ucm.es/e-4/

```
***** Options set by user *****
Filter. . . . . : KALMAN
Scaled B and M matrices . . . . : NO
Initial state vector. . . . . : AUTOMATIC SELECTION
Initial covariance of state v. . : IDEJONG
Variance or Cholesky factor? . . : VARIANCE
Optimization algorithm. . . . . : BFGS
Maximum step length . . . . . : 0.100000
Stop tolerance. . . . . . . : 0.000010
Max. number of iterations . . . : 75
Verbose iterations. . . . . : YES
*****
```

Bear in mind that E^4 will not work properly if this `e4init` is not run.

¹This search path is a list of folders where MATLAB will look for the code of any function which is not part of its core.

After these installation and initialization steps have been completed, MATLAB is able to use the E^4 library.

D.3 Downloading the code for the examples in this book

In the Downloads section of [<https://www.ucm.es/e-4/downloads>] you will find an item titled “Source code and data for the examples in the book: State-Space Methods for Time Series Analysis: Theory, Applications and Software.” By clicking on this item you will download a file titled `SSBookExamples.zip`. You should extract its content to any folder of your choice. By doing this, you will create the folders: `ExamplesChapter3`, `ExamplesChapter4`,... and so on until `ExamplesChapter11`. These folders contain the source code and data required to replicate the examples in the corresponding Chapter.

In each of these folders you will find either code files or new folders. In both cases, the name of the file/folder refers to a specific example, e.g. the files contained in `ExamplesChapter3` are:

```
Example0311.m  
Example0312.m  
Example0313.m  
Example0321.m  
Example0322.m  
Example0331.m
```

where the first nine characters, `Example03`, refer to the Chapter, the tenth character refers to the Section containing the example, and the last character is the sequential number of the example. Therefore, the name `Example0311.m` refers to the first example in Chapter 3, Section 1, `Example0312.m` corresponds to the second example in the same Chapter and Section, and so on.

In other cases the main folder will contain several sub-folders. For example, `ExamplesChapter8` includes the sub-folders:

```
Example0851 (Simulated data)  
Example0852 (Wheat price)  
Example0853 (Lydia Pinkham data)
```

where the first part of the folder name follows the naming convention described above, and the free text within parentheses refers to the specific dataset employed. Each of these folder contains MATLAB code (in `.m` files) and data (in text files with the extension `.dat`). In some cases you will also find and Excel worksheets with additional information about the dataset and the results.

The `ExamplesChapter11` folder is a little different since the examples were coded in R, and it contains just three files:

```
Example1161.R  
Example1162.R  
whale.txt
```

where `whale.txt` is the dataset required by the second example in Section 6 (`Example1162.R`). The first two files are chunks of R code meant to be executed

sequentially, with accompanying comments to indicate what exactly is being computed. Each of these may require the user to load specific R libraries². In particular, both examples make liberal use of the "astsa" library that accompanies the book by Shumway and Stoffer [196]

²See the website <https://www.r-project.org/> for detailed information on R.

Bibliography

- [1] S. Ahn and G. Reinsel. Estimation for partially nonstationary multivariate autoregressive models. *Journal of the American Statistical Association*, 85:813–823, 1990.
- [2] H. Akaike. *Canonical Correlation Analysis of Time Series and the Use of an Information Criterion*. New York: Academic Press, 1976.
- [3] H. Akaike. Covariance matrix computation of the state variable of a stationary gaussian process. *Annals of the Institute of Statistics and Mathematics*, Series B, 30:499–504, 1978.
- [4] G. Alexander and P. Benson. More on beta as a random coefficient. *Journal of Financial and Quantitative Analysis*, XVII, 1:27–36, 1982.
- [5] T. Amemiya and R. Wu. The effect of aggregation on prediction in the autoregressive model. *Journal of the American Statistical Association*, 339:628–632, 1972.
- [6] B. Anderson and J. Moore. *Optimal Filtering*. Englewood Cliffs (N.J.): Prentice Hall, 1979.
- [7] C. Ansley and R. Kohn. Exact likelihood of vector autoregressive-moving average process with missing or aggregated data. *Biometrika*, 70:275–278, 1983.
- [8] C. Ansley and R. Kohn. Estimation, filtering and smoothing in state space models with incompletely specified initial conditions. *The Annals of Statistics*, 13:1286–1316, 1985.
- [9] C. Ansley and R. Kohn. Filtering and smoothing in state space models with partially diffuse initial conditions. *Journal Time Series Analysis*, 11:275–293, 1990.
- [10] C. Ansley and P. Newbold. Finite sample properties of estimators for autoregression moving average models. *Journal of Econometrics*, 13:159–183, 1980.
- [11] M. Aoki. *State Space Modelling of Time Series*. Springer Verlag, New York, 1990.
- [12] K. Arun and S. Kung. Balanced approximation of stochastic systems. *SIAM Journal of Matrix Analysis and Applications*, 11(1):42–68, 1990.
- [13] G. Baiocchi and W. Distaso. GRETL: Econometric software for the GNU generation. *Journal of Applied Econometrics*, 18:105–110, 2003.

- [14] R. Barnard, A. Trindade, and R. Wickramasinghe. Autoregressive moving average models under exponential power distributions. *ProbStat Forum*, 7:65–77, 2014.
- [15] R. Bartels and G. Stewart. Solution of the equation $ax + xb = c$. *Communications ACM*, 15:820–826, 1972.
- [16] D. Bauer. Order estimation for subspace methods. *Automatica*, 37:1561–1573, 2001.
- [17] D. Bauer and L. Ljung. Some facts about the choice of the weighting matrices in Larimore type of subspace algorithms. *Automatica*, 38:763–773, 2002.
- [18] D. Bauer and M. Wagner. Estimating cointegrated systems using subspace algorithms. *Journal of Econometrics*, 111:47–84, 2002.
- [19] W. Bell and S. Hillmer. Initializing the Kalman filter for nonstationary time series models. *Journal of Time Series Analysis*, 12:283–300, 1991.
- [20] T. Bengtsson and J. Cavanaugh. An improved Akaike information criterion for state-space model selection. *Computational Statistics and Data Analysis*, 50(10):2635–2654, 2006.
- [21] M. Bentarzi and M. Hallin. On the invertibility of periodic moving-average models. *Journal of Time Series Analysis*, 15, 3:263–268, 1994.
- [22] A. Bera, M. Higgins, and S. Lee. Interaction between autocorrelation and conditional heteroscedasticity: a random-coefficient approach. *Journal of Business and Economic Statistics*, 10, 2:133–142, 1992.
- [23] R. Bitmead, M. Gevers, I. Petersen, and R. Kaye. Monotonicity and stabilizability properties of the solutions of the Riccati difference equation: Propositions, lemmas, theorems, fallacious conjectures and counterexamples. *System Control Letters*, 5:309–315, 1985.
- [24] K. Bollen. *Structural Equations With Latent Variables*. Wiley, New York, 1989.
- [25] T. Bollerslev, R. Engle, and D. Nelson. ARCH models. In R. Engle and D. McFadden, editors, *Handbook of Econometrics*, volume IV. North-Holland, 1994.
- [26] J. Boot, W. Feibes, and J. Lisman. Further methods of derivation of quarterly figures from annual data. *Applied Statistics*, 16:65–75, 1967.
- [27] P. Boswijk and P. Franses. Unit roots in periodic autoregressions. *Journal of Time Series Analysis*, 17, 3:221–245, 1995.
- [28] G. Box and D. Cox. An analysis of transformations. *Journal of the Royal Statistical Society, Series B*, 26:211–243, 1964.
- [29] G. Box and G. Jenkins. *Time Series Analysis, Forecasting and Control*. Holden-Day, San Francisco, 1970.
- [30] G. Box, G. Jenkins, and G. Reinsel. *Time Series Analysis, Forecasting and Control*. Prentice Hall, Englewood Cliffs, 1994.

- [31] G. Box and G. Tiao. Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70(349):70–79, 1975.
- [32] P. Brockwell and R. Davis. *Time Series: Theory and Methods*. Springer-Verlag, New York, 2nd edition, 1991.
- [33] M. Bujosa, A. Garcia-Ferrer, and P. Young. Linear dynamic harmonic regression. *Computational Statistics and Data Analysis*, 52(2):999–1024, 2007.
- [34] J. Burman. Seasonal adjustment by signal extraction. *Journal of the Royal Statistical Society, Series A*, 143:321–337, 1980.
- [35] I. Burr. *Statistical Quality Control Methods*. Marcel Dekker, New York, 1976.
- [36] Z. Cai, J. Fan, and Q. Yao. Functional-coefficient regression models for nonlinear time series. *Journal of the American Statistical Association*, 95, 451:941–956, 2000.
- [37] P. Caines. *Linear Stochastic Systems*. John Wiley, New York, 1988.
- [38] J. Casals. *Métodos de Subespacios en Econometría*. PhD thesis, Universidad Complutense de Madrid, 1997.
- [39] J. Casals, A. García-Hiernaux, and M. Jerez. From general State-Space to VARMAX models. *Mathematics and Computers in Simulation*, 80(5):924–936, 2012.
- [40] J. Casals, M. Jerez, and S. Sotoca. Exact smoothing for stationary and non-stationary time series. *International Journal of Forecasting*, 16(1):199–209, 2000.
- [41] J. Casals, M. Jerez, and S. Sotoca. An exact multivariate model-based structural decomposition. *Journal of the American Statistical Association*, 97(458):553–564, 2002.
- [42] J. Casals, M. Jerez, and S. Sotoca. Decomposition of a state-space model with inputs. *Journal of Statistical Computation and Simulation*, 80, 9:979–992, 2010.
- [43] J. Casals and S. Sotoca. The exact likelihood for a state space model with stochastic inputs. *Computers and Mathematics with Applications*, 42:199–209, 2001.
- [44] J. Casals, S. Sotoca, and M. Jerez. Un algoritmo rápido para evaluar la verosimilitud exacta de modelos VARMAX periódicos. *Estadística Española*, 40, 143:269–291, 1998.
- [45] J. Casals, S. Sotoca, and M. Jerez. A fast and stable method to compute the likelihood of time invariant state space models. *Economics Letters*, 65(3):329–337, 1999.
- [46] J. Casals, S. Sotoca, and M. Jerez. Modeling and forecasting time series sampled at different frequencies. *Journal of Forecasting*, 28, 4:316–342, 2009.

- [47] J. Casals, S. Sotoca, and M. Jerez. Minimally conditioned likelihood for a nonstationary state space model. *Mathematics and Computers in Simulation*, 100C:24–40, 2014.
- [48] G. Casella and R. Berger. *Statistical Inference*. Duxbury, Pacific Grove, USA., 2nd edition, 2002.
- [49] S. Chan, G. Goodwin, and K. Sin. Convergence properties of the Riccati difference equation in optimal filtering of nonstabilizable systems. *IEEE Transactions on Automatic Control*, AC-29, 2:110–118, 1984.
- [50] H. Chang and Y. Hsing. The demand for residential electricity: New evidence on time-varying elasticities. *Applied Economics*, 23, 7:1251–56, 1991.
- [51] S. Chatterjee, M. Handcock, and J. Simonoff. *A Casebook for a First Course in Statistics and Data Analysis*. Wiley, New York, 1995.
- [52] G. Chow and A. Lin. Best linear unbiased interpolation, distribution and extrapolation of time series by related series. *The Review of Economics and Statistics*, 53:372–375, 1971.
- [53] S. Chow, M. Ho, E. Hamaker, and C. Dolan. Equivalence and differences between structural equation modeling and state-space modeling techniques. *Structural Equation Modeling*, 17:303—332, 2010.
- [54] W. Cleveland and S. Devlin. Calendar effects in monthly time series: Modeling and adjustment. *Journal of the American Statistical Association*, 77, 379:520–528, 1982.
- [55] J. Commandeur and S. Koopman. *An Introduction to State Space Time Series Analysis*. Oxford University Press, 2007.
- [56] T. Cooley and E. Prescott. An adaptive regression model. *International Economic Review*, 14, 2:364–371, 1973.
- [57] D. M. Cooper and E. F. Wood. Identifying multivariate time series models. *Journal of Time Series Analysis*, 3:277–293, 1982.
- [58] H. Cramer. *Mathematical Methods of Statistics*. Princeton University Press, Princeton (N.J.), 1946.
- [59] M. J. Crowder. Maximum likelihood estimation for dependent observations. *Journal of the Royal Statistical Society. Series B (Methodological)*, 38, 1:45–53, 1976.
- [60] B. Datta. *Numerical Linear Algebra and Applications*. Brooks/ Cole, 1995.
- [61] R. Davidson and J. MacKinnon. *Econometric Theory and Methods*. Oxford University Press, New York, 2004.
- [62] R. Davidson and J. G. MacKinnon. *Estimation and Inference in Econometrics*. Oxford University Press, New York, 1993.
- [63] R. Davis, M. Chen, and W. Dunsmuir. Inference for MA(1) processes with a unit root on or near the unit circle. *Probability and Mathematical Statistics*, 15:227–242, 1995.

- [64] R. Davis and W. Dunsmuir. Least absolute deviation estimation for regression with ARMA errors. *Journal of Theoretical Probability*, 10(2):481–497, 1997.
- [65] R. Davis and L. Song. Unit roots in moving averages beyond first order. *The Annals of Statistics*, 39(6):3062–3091, 2011.
- [66] P. de Jong. The likelihood for a state space model. *Biometrika*, 75, 1:165–169, 1988.
- [67] P. de Jong. Smoothing and interpolation with the state-space model. *Journal of the American Statistical Association*, 84:1085–1088, 1989.
- [68] P. de Jong. The diffuse Kalman filter. *Annals of Statistics*, 2:1073–1083, 1991.
- [69] P. de Jong and S. Chu-Chun-Lin. Fast likelihood evaluation and prediction for nonstationary state space models. *Biometrika*, 81:133–142, 1994.
- [70] P. de Jong and J. Penzer. Diagnosing shocks in time series. *Journal of the American Statistical Association*, 93, 442:796–806, 1998.
- [71] C. De Souza, M. Gevers, and G. Goodwin. Riccati equations in optimal filtering of nonstabilizable systems having singular state transition matrices. *IEEE Transactions on Automatic Control*, 31:831–838, 1986.
- [72] J. S. Demetry. A note on the nature of optimality in the discrete Kalman filter. *IEEE Transactions on Automatic Control*, AC-15:603–604, 1970.
- [73] A. P. Dempster, N. M. Laird, and D. B. Rubin. Maximum likelihood from incomplete data via the EM algorithm. *Journal of the Royal Statistical Society B*, 39:1–38, 1977.
- [74] J. E. Dennis and R. B. Schnabel. *Numerical methods for unconstrained optimization and nonlinear equations*. Prentice Hall, Englewood Cliffs, NJ, 1996.
- [75] F. Denton. Adjustment of monthly or quarterly series to annual totals: An approach based on quadratic minimization. *Journal of the American Statistical Association*, 66:99–102, 1971.
- [76] T. Di Fonzo. The estimation of m disaggregate time series when contemporaneous and temporal aggregates are known. *The Review of Economics and Statistics*, 72:178–182, 1990.
- [77] D. A. Dickey and W. A. Fuller. Distribution of the estimators for autoregressive time series with unit root. *Journal of the American Statistical Association*, 74:427–431, 1979.
- [78] D. A. Dickey and W. A. Fuller. Likelihood ratio statistics for autoregressive time series with a unit root. *Econometrica*, 49:1057–1072, 1981.
- [79] B. Dickinson, M. Morf, and T. Kailath. Canonical matrix fraction and state space descriptions for deterministic and stochastic linear systems. *IEEE Transactions on Automatic Control*, AC-19:656–667, 1974.
- [80] P. J. Diggle, P. Heagerty, K. Liang, and S. L. Zeger. *Analysis of Longitudinal Data*. Oxford University Press, New York, 2nd edition, 2002.
- [81] J. M. Dufour and D. Pelletier. Linear estimation of weak VARMA models with macroeconomic applications. Technical report, Centre Interuniversitaire

- de Recherche en Economie Quantitative (CIREQ), Université de Montréal, CANADA, 2004.
- [82] J. Durbin and S. Koopman. *Time Series Analysis by State Space Methods*. Oxford University Press, 2012.
 - [83] J. Durbin and B. Quenneville. Benchmarking by state space models. *International Statistical Review / Revue Internationale de Statistique*, 65:23–48, 1997.
 - [84] R. Engle and T. Bollerslev. Modelling the persistence of condicinal variances. *Econometric Reviews*, 5:1–50, 1986.
 - [85] R. Engle and K. Kroner. Multivariate simultaneous generalized ARCH. *Econometric Theory*, 11:122–150, 1995.
 - [86] R. F. Engle. Autoregressive conditional heteroskedasticity with estimates of the variance of united kingdom inflations. *Econometrica*, 50:987–1007, 1982.
 - [87] R. F. Engle and S. Kozicki. Testing for common features. *Journal of Business and Economic Statistics*, 11(4):369–395, 1993.
 - [88] P. Evans, B. Kim, and K. Oh. Capital mobility in saving and investment: a time-varying coefficients approach. *Journal of International Money and Finance*, 27:806–815, 2008.
 - [89] F. Fabozzi and J. Francis. Beta as a random coefficient. *Journal of Financial and Quantitative Analysis*, 13:101–116, 1978.
 - [90] L. Fahrmeir and G. Tutz. *Multivariate Statistical Modelling Based on Generalized Linear Models*. Springer, New York, 2nd edition, 2001.
 - [91] J. Fan, J. Jiang, C. Zhang, and Z. Zhou. Time-dependent difussion models for term structure dynamics and the stock price volatility. *Statistica Sinica*, 13:965–992, 2003.
 - [92] J. Fan and W. Zhang. Statistical estimation in varying coefficient models. *Annals of Statistics*, 27, 5:1491–1518, 1999.
 - [93] J. Fan and W. Zhang. Statistical methods with varying coefficient models. *Statistics and Its Interface*, 1:179–195, 2008.
 - [94] W. Favoreel, B. De Moor, and P. Van Overschelle. Subspace state space system identification for industrial processes. *Journal of Process Control*, 10:149–155, 2000.
 - [95] R. Fernandez. A methodological note on the estimation of time series. *The Review of Economics and Statistics*, 63:471–476, 1981.
 - [96] R. Fisher. On the mathematical foundations of theoretical statistics. *Philosophical Transactions of the Royal Society of London. Series A*, 222:309–368, 1921.
 - [97] R. Flores and G. Serrano. A generalized least squares estimation method for VARMA models. *Statistics*, 36(4):303–316, 2002.
 - [98] D. Y. T. Fong. *State Space Models and Filtering Methods in Longitudinal Studies*. PhD thesis, University of Waterloo, 1997.

- [99] C. Francq and J. M. Zakoian. Estimating weak GARCH representations. *Econometric Theory*, 16:692–728, 2000.
- [100] P. Franses. *Periodicity and Stochastic Trends in Economic Time Series*. Oxford University Press, 1996.
- [101] P. Franses and R. Paap. Model selection in periodic autoregressions. *Oxford Bulletin of Economics and Statistics*, 56, 4:421–439, 1994.
- [102] A. Garcia-Hiernaux. *Identificación de Modelos para Series Temporales mediante Métodos de Subespacios*. PhD thesis, Universidad Complutense de Madrid, Febrero 2005.
- [103] A. Garcia-Hiernaux. Forecasting linear dynamical systems using subspace methods. *Journal of Time Series Analysis*, 32(5):462–468, 2011.
- [104] A. Garcia-Hiernaux, M. Jerez, and J. Casals. Fast estimation methods for time series models in state-space form. *Journal of Statistical Computation and Simulation*, 79(2):121–134, 2009.
- [105] A. Garcia-Hiernaux, M. Jerez, and J. Casals. Unit roots and cointegration modeling through a family of flexible information criteria. *Journal of Statistical Computation and Simulation*, 80(2):173–189, 2010.
- [106] A. Garcia-Hiernaux, M. Jerez, and J. Casals. Estimating the system order by subspace methods. *Computational Statistics*, 27(3):411–425, 2012.
- [107] E. Ghysels and D. Osborn. *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press, 2001.
- [108] K. Glover and J. C. Willems. Parameterizations of linear dynamical systems: Canonical forms and identifiability. *IEEE Transactions on Automatic Control*, 19, 6:640–646, 1974.
- [109] G. Golub and C. Van Loan. *Matrix Computations*. John Hopkins University Press, Baltimore, 1996.
- [110] V. Gomez and A. Maravall. Estimation prediction and interpolation for non-stationary series with the Kalman filter. *Journal of the American Statistical Association*, 89:611–624, 1994.
- [111] V. Gomez., A. Maravall, and D. Pena. Missing observations in ARIMA models: Skipping approach versus additive outlier approach. *Journal of Econometrics*, 88, 2:341–363, 1998.
- [112] R. L. Goodrich and P. E. Caines. Linear system identification from nonstationary cross-sectional data. *IEEE Transactions on Automatic Control*, 24:403—411, 1979.
- [113] C. Granger and P. Siklos. Systematic sampling, temporal aggregation, seasonal adjustment and cointegration: Theory and evidence. *Journal of Econometrics*, 66:357–369, 1995.
- [114] P. Hackl. Demand for international telecommunication: time-varying price elasticity. *Journal of Econometrics*, 70:243–260, 1996.
- [115] H. Hamilton. *Time Series Analysis*. Princeton University Press, 1993.

- [116] E. J. Hannan and M. Deistler. *The Statistical Theory of Linear Systems*. John Wiley, New York, 1988.
- [117] E. J. Hannan and B. G. Quinn. The determination of the order of an autoregression. *Journal of the Royal Statistical Association, Series B*, 41:713–723, 1979.
- [118] A. Harvey. *Forecasting, structural time series models and the Kalman Filter*. Cambridge University Press, 1989.
- [119] A. Harvey, E. Ruiz, and N. Shephard. Multivariate stochastic variance models. *Review of Economic Studies*, 61, 2:247–264, 1994.
- [120] A. Harvey and T. Trimbur. Trend estimation and the hodrick-prescott filter. *Journal of the Japan Statistical Society*, 38:41–49, 2008.
- [121] D. A. Harville. *Matrix algebra from a statistician's perspective*. Springer, New York, 1997.
- [122] T. Hastie and R. Tibshirani. Varying-coefficient models. *Journal of the Royal Statistical Society. Series B (Methodological)*, 55, 4:757–796, 1993.
- [123] G. Hewer. An iterative technique for the computation of the steady state gains for the discrete optimal regulator. *IEEE Transaction on Automatic Control*, AC-16, 4:382–384, 1971.
- [124] S. Hillmer and G. Tiao. Likelihood function of stationary multiple autoregressive moving average models. *Journal of the American Statistical Association*, 74:652–660, 1979.
- [125] S. Hillmer and G. Tiao. An ARIMA-model-based approach to seasonal adjustment. *Journal of the American Statistical Association*, 77:63–70, 1982.
- [126] R. Hodrick and E. Prescott. Post-war U.S. business cycles: An empirical investigation. *Journal of Money, Credit and Banking*, 29:1–16, 1997.
- [127] C. Hsiao. *Analysis of Panel Data*. Cambridge University Press, New York, 2014.
- [128] C. Ionescu, V. Oara and M. Weiss. General matrix pencil techniques for the solution of algebraic Riccati equations: A unified approach. *IEEE Transactions on Automatic Control*, 42, 8:1085–1097, 1997.
- [129] M. Jamshidian and R. I. Jennrich. Standard errors for EM estimation. *Journal of the Royal Statistical Society. Series B, Statistical Methodology*, 62, 2:257–270, 2000.
- [130] A. Jazwinsky. *Stochastic Processes and Filtering Theory*. Academic Press, New York, 1970.
- [131] G. Jenkins and A. Alavi. Some aspects of modelling and forecasting multivariate time series. *Journal of Time Series Analysis*, 2(1):1–47, 1981.
- [132] R. Jones. Maximum likelihood fitting of ARMA models to time series with missing observations. *Technometrics*, 22, 3:389–395, 1980.
- [133] R. H. Jones. *Longitudinal Data with Serial Correlation: A State-Space Approach*. Chapman & Hall/CRC, Boca Raton, 1993.

- [134] R. Kalman. A new approach to linear filtering and prediction problems. *Transactions of the ASME-Journal of Basic Engineering (Series D)*, 82:35–45, 1960.
- [135] G. Kauermann and G. Tutz. On model diagnostics using varying coefficient models. *Biometrika*, 86, 1:119–128, 1999.
- [136] R. Kohn and C. Ansley. Signal extraction for finite nonstationary time series. *Biometrika*, 74, 2:411–421, 1987.
- [137] R. Kohn and C. F. Ansley. Estimation, prediction and interpolation for ARIMA models with missing data. *Journal of the American Statistical Association*, 81:751–761, 1986.
- [138] R. Kohn and C. F. Ansley. A fast algorithm for signal extraction, influence and cross-validation in state space models. *Biometrika*, 76, 1:65–79, 1989.
- [139] S. Koopman. Exact initial Kalman filtering and smoothing for nonstationary time series models. *Journal of the American Statistical Association*, 92:1630–1638, 1997.
- [140] S. Koopman and N. Shephard. Exact score for time series models in state space form. *Biometrika*, 79, 4:823–826, 1992.
- [141] S. Koopman, N. Shephard, and J. Doornik. *SsfPack 3.0: Statistical Algorithms for Models in State Space Form*. Timberlake Consultants Press, London, 2008.
- [142] S. J. Koopman and J. Durbin. Filtering and smoothing of state vector for diffuse state-space models. *Journal of Time Series Analysis*, 24, 1:85–98, 2003.
- [143] S. G. Koreisha and T. H. Pukkila. Fast linear estimation methods for vector autoregressive moving average models. *Journal of Time Series Analysis*, 10:325–339, 1989.
- [144] S. G. Koreisha and T. H. Pukkila. A generalized least-squares approach for estimation of autoregressive moving-average models. *Journal of Time Series Analysis*, 11:139–151, 1990.
- [145] S. Y. Kung. A new identification and model reduction algorithm via singular value decompositions. In *Proceedings of the 12th Asilomar Conference on Circuits, Systems and Computers, Pacific Grove*, pages 705–714, 1978.
- [146] N. M. Laird and J. H. Ware. Random-effects models for longitudinal data. *Biometrics*, 38:963—974, 1982.
- [147] W. E. Larimore. Canonical variate analysis in identification, filtering and adaptive control. *Proceedings of the 29th Conference on Decision and Control, Hawaii*, pages 596–604, 1990.
- [148] W. K. Li. *Diagnostic Checks in Time Series*. Chapman and Hall/CRC, Florida, 2004.
- [149] R. Litterman. A random walk, markov model for the distribution of time series. *Journal of Business and Economic Statistics*, 1:169–173, 1983.

- [150] T. D. Little. *Longitudinal Structural Equation Modeling*. Guilford Press, New York, 2013.
- [151] G. M. Ljung and G. E. P. Box. On a measure of lack of fit in time series models. *Biometrika*, 65:297–303, 1978.
- [152] L. Ljung. *System Identification, Theory for the User*. PTR Prentice Hall, New Jersey, 1999.
- [153] L. Ljung and P. Caines. Asymptotic normality of prediction error estimators for approximate system models. *Stochastic*, 3:29–46, 1979.
- [154] R. Lucas. Econometric policy evaluation: A critique. *Carnegie-Rochester Conference Series on Public Policy*, 1, 1:19–46, 1976.
- [155] H. Lutkepohl. *Forecasting Aggregated Vector ARMA Processes*. Springer-Verlag, Berlin, 1987.
- [156] H. Lutkepohl. *Introduction to Multiple Time Series Analysis*. Springer Verlag, Berlin, 1993.
- [157] H. Lutkepohl. *Handbook of Matrices*. Wiley, Chichester, 1996.
- [158] H. Lutkepohl. *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin, 2005.
- [159] H. Lutkepohl and D. S. Poskitt. Specification of echelon form VARMA models. *Journal of Business and Economic Statistics*, 14(1):69–79, 1996.
- [160] S. G. Makridakis, S. C. Wheelwright, and V. E. McGee. *Forecasting: Methods and Applications*. John Wiley & Sons, 1983.
- [161] M. Marcellino. Some consequences of temporal aggregation in empirical analysis. *Journal of Business and Economic Statistics*, 1:129–136, 1999.
- [162] D. Margaritis. A time-varying model of rational learning. *Economics Letters*, 33:309–314, 1990.
- [163] R. K. Mehra. Identification and estimation of the error-in-variables model (EVM) in structural form. In R. Web, editor, *Stochastic Systems: Modelling Identification and Optimization*, volume 1. North-Holland, Amsterdam, 1976.
- [164] G. Mélard, R. Roy, and A. Saidi. Exact maximum likelihood estimation of structured or unit root multivariate time series models. *Computational Statistics and Data Analysis*, 50,11:2958–2986, 2006.
- [165] A. Naranjo, A. A. Trindade, and G. Casella. Extending the state-space model to accommodate missing values in responses and covariates. *Journal of the American Statistical Association*, 108, 501:202–216, 2013.
- [166] M. Nerlove, D. M. Grether, and J. L. Carvalho. *Analysis of Economic Time Series: A Synthesis*. Academic Press, New York, 1995.
- [167] D. Nicholls and B. Quinn. *Random Coefficient Autoregressive Models: An Introduction*. John Wiley & Sons, 1982.

- [168] A. Novales and R. Flores. Forecasting with periodic models a comparison with time invariant coefficient models. *International Journal of Forecasting*, 13, 3:393–405, 1997.
- [169] L. Nunes. Nowcasting quarterly GDP growth in a monthly coincident indicator model. *Journal of Forecasting*, 24:575–592, 2005.
- [170] T. M. O'Donovan. *Short Term Forecasting*. John Wiley & Sons, 1983.
- [171] D. Osborn and J. Smith. The performance of periodic autoregressive models in forecasting seasonal u.k. consumption. *Journal of Business & Economic Statistics*, 7, 1:117–127, 1989.
- [172] J. H. L. Oud and R. A. R. Jansen. Continuous time state space modeling of panel data by means of SEM. *Psychometrika*, 65:199–215, 2000.
- [173] A. Pagan. Some identification and estimation results for regression models with stochastically varying coefficients. *Journal of Econometrics*, 13, 3:341–363, 1980.
- [174] R. Paige, A. Trindade, and R. Wickramasinghe. Extensions of saddlepoint-based bootstrap inference. *Annals of the Institute of Statistical Mathematics*, 66(5):961–981, 2014.
- [175] K. Palda. *The Measurement of Cumulative Advertising Effects*. Prentice-Hall, Englewood Cliffs (N.J.), 1964.
- [176] T. Pappas, A. Laub, and N. Sandell. On the numerical solution of the algebraic Riccati equation. *IEEE Transactions on Automatic Control*, AC-25, 4:631–641, 1980.
- [177] B. U. Park, E. Mammen, Y. K. Lee, and E. Lee. Varying coefficient regression models: A review and new developments. *International Statistical Review*, 83, 1:36–64, 2014.
- [178] S. Park and G. Zhao. An estimation of U.S. gasoline demand: A smooth time-varying cointegration approach. *Energy Economics*, 32, 1:110–120, 2010.
- [179] M. M. Pelagatti. State space methods in Ox/SsfPack. *Journal of Statistical Software*, 41, 3:1–25, 2011.
- [180] K. Peternell. *Identification of Linear Dynamical Systems by Subspace and Realization-Based Algorithms*. PhD thesis, TU Wien, 1995.
- [181] P. Petkov, N. Christov, and M. Konstantinov. *Computational Methods for Linear Control Systems*. Prentice-Hall, Englewood Cliffs, New Jersey, 1991.
- [182] G. Petris and S. Petrone. State space models in R. *Journal of Statistical Software*, 41:1–25, 2011.
- [183] R. Pierse and A. Snell. Temporal aggregation and the power of tests for a unit root. *Journal of Econometrics*, 65:333–345, 1995.
- [184] J. Pinheiro and D. Bates. *Mixed-Effects Models in S and S-PLUS*. Springer, New York, 2000.
- [185] S. J. Qin. An overview of subspace identification. *Computers and Chemical Engineering*, 30:1502–1513, 2006.

- [186] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2014.
- [187] C. G. Reinsel. *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York, 2nd edition, 1997.
- [188] B. Rosenberg. *Varying Parameter Estimation*. PhD thesis, Harvard University, 1968.
- [189] B. Rosenberg. The analysis of a cross section of time series by stochastically convergent parameter regression. *Annals of Economic and Social Measurement*, 2:399–428, 1973.
- [190] H. H. Rosenbrock. *State Space and Multivariable Theory*. John Wiley, New York, 1970.
- [191] G. Schwarz. Estimating the dimension of a model. *Annals of Statistics*, 6:461–464, 1978.
- [192] F. Schweppe. Evaluation of likelihood functions for gaussian signals. *IEEE Transactions on Information Theory*, IT-11, 1:61–70, 1965.
- [193] R. Selukar. State space modeling using sas. *Journal of Statistical Software*, 41, 12:1–13, 2011.
- [194] R. H. Shumway and D. S. Stoffer. An approach to time series smoothing and forecasting using the EM algorithm. *Journal of Time Series Analysis*, 3:253–264, 1982.
- [195] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer, New York, 2nd edition, 2006.
- [196] R. H. Shumway and D. S. Stoffer. *Time Series Analysis and Its Applications: With R Examples*. Springer, New York, 3rd edition, 2011.
- [197] J. Sorelius. *Subspace-Based Parameter Estimation Problems in Signal Processing*. PhD thesis, Uppsala University, 1999.
- [198] H. Spliid. A fast estimation method for the vector autoregressive moving average model with exogenous variables. *Journal of the American Statistical Association*, 78(384):843–849, 1983.
- [199] J. Stock and M. Watson. Why has U.S. inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39:3–33, 2007.
- [200] D. S. Stoffer. *Estimation of Parameters in a Linear Dynamic System with Missing Observations*. PhD thesis, University of California, Davis, 1982.
- [201] D. S. Stoffer and K. Wall. Bootstrapping state space models: Gaussian maximum likelihood estimation and the Kalman filter. *Journal of the American Statistical Association*, 86:1024–1033, 1991.
- [202] P. Stoica, Y. Selen, and Jian Li. On information criteria and the generalized likelihood ratio test of model order selection. *Signal Processing Letters, IEEE*, 11(10):794–797, 2004.
- [203] P. Stoica and T. Soderstrom. On nonsingular information matrices and local identifiability. *International Journal of Control*, 36:323–329, 1982.

- [204] D. Stram and W. Wei. Temporal aggregation in the ARIMA process. *Journal of Time Series Analysis*, 7:279–292, 1986.
- [205] P. Swamy and G. Tavlas. Random coefficient models: Theory and applications. *Journal of Economic Surveys*, 9:165–196, 1995.
- [206] J. Terceiro. *Estimation of Dynamics Econometric Models with Errors in Variables*. Springer-Verlag, Berlin, 1990.
- [207] G. Tiao. Asymptotic behavior of time series aggregates. *Biometrika*, 59:521–531, 1972.
- [208] G. Tiao. *Vector ARMA Models*. John Wiley & Sons, 2001.
- [209] G. C. Tiao and R. S. Tsay. Model specification in multivariate time series. *Journal of the Royal Statistical Society, Series B*, 51(2):157–213, 1989.
- [210] H. Tong. *Nonlinear Time Series: A Dynamical System Approach*. Oxford University Press, Oxford, 1990.
- [211] E. M. M. Toscano and V. A. Reisen. The use of canonical correlation analysis to identify the order of multivariate ARMA models: Simulation and application. *Journal of Forecasting*, 19:441–455, 2000.
- [212] R. S. Tsay. Identifying multivariate time series models. *Journal of Time Series Analysis*, 10(4):357–372, 1989.
- [213] F. Tusell. Kalman filtering in r. *Journal of Statistical Software*, 39:1–27, 2011.
- [214] P. Van Dooren. A generalized eigenvalue approach for solving Riccati equations. *SIAM Journal of Scientific Statistic Computation*, 2:121–135, 1981.
- [215] P. Van Overschee and B. De Moor. N4SID: Subspace algorithms for the identification of combined deterministic-stochastic systems. *Automatica*, 30(1):75–93, 1994.
- [216] P. Van Overschee and B. De Moor. A unifying theorem for three subspace system identification algorithms. *Automatica*, 31(12):1853–1864, 1995.
- [217] M. Verhaegen. Identification of the deterministic part of MIMO state space models given in innovations form from input-output data. *Automatica*, 30(1):61–74, 1994.
- [218] A. Wald. Note on the consistency of the maximum likelihood estimate. *The Annals of Mathematical Statistics*, 20, 4:595–601, 1949.
- [219] M. Watson and R. Engle. Alternative algorithms for the estimation of dynamic factor, MIMIC and varying coefficient regression models. *Journal of Econometrics*, 23, 3:385–400, 1983.
- [220] W. Wei. *Some Consequences of Temporal Aggregation in Seasonal Time Series Models*. In *Seasonal Analysis of Economic Time Series*. Bureau of the Census, Washington D.C., 1978.
- [221] W. Wei. *Time Series Analysis*. Addison Wesley, 2006.
- [222] W. Wei and D. Stram. Disaggregation of time series models. *Journal of the Royal Statistical Society, Series B*, 52:453–467, 1990.

- [223] M. West. Time series decomposition and analysis in a study of oxygen isotope records. *Biometrika*, 84:489–494, 1997.
- [224] M. West and P. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York, 1989.
- [225] M. West and P. Harrison. *Bayesian Forecasting and Dynamic Models*. Springer-Verlag, New York, 2nd edition, 1997.
- [226] H. White. Maximum likelihood estimation of misspecified models. *Econometrica*, 50, 1:1–25, 1982.
- [227] H. O. A. Wold. *A Study in the Analysis of Stationary Time Series*. Almqvist and Wiksell, Uppsala, 1964.
- [228] C. M. Woodside. Estimation of the order of linear systems. *Automatica*, 7:727–733, 1971.

Statistics

The state-space approach provides a formal framework where any result or procedure developed for a basic model can be seamlessly applied to a standard formulation written in state-space form. Moreover, it can accommodate with a reasonable effort nonstandard situations, such as observation errors, aggregation constraints, or missing in-sample values.

Exploring the advantages of this approach, **State-Space Methods for Time Series Analysis: Theory, Applications and Software** presents many computational procedures that can be applied to a previously specified linear model in state-space form.

After discussing the formulation of the state-space model, the book illustrates the flexibility of the state-space representation and covers the main state estimation algorithms: filtering and smoothing. It then shows how to compute the Gaussian likelihood for unknown coefficients in the state-space matrices of a given model before introducing subspace methods and their application. It also discusses signal extraction, describes two algorithms to obtain the VARMAX matrices corresponding to any linear state-space model, and addresses several issues relating to the aggregation and disaggregation of time series. The book concludes with a cross-sectional extension to the classical state-space formulation in order to accommodate longitudinal or panel data. Missing data is a common occurrence here, and the book explains imputation procedures necessary to treat missingness in both exogenous and endogenous variables.

The authors' *E⁴* MATLAB® toolbox (available online) offers all the computational procedures, administrative and analytical functions, and related materials for time series analysis. This flexible, powerful, and free software tool enables you to replicate the practical examples in the text and apply the procedures to your own work.

Features

- Focuses on the computational processes that can be applied to a specified linear model
- Discusses two classical topics in the state-space literature: model estimation and signal extraction
- Describes many procedures to combine, decompose, aggregate, and disaggregate a state-space form
- Covers the connection between mainstream time series models and the state-space representation



CRC Press

Taylor & Francis Group
an informa business
www.crcpress.com

6000 Broken Sound Parkway, NW
Suite 300, Boca Raton, FL 33487
711 Third Avenue
New York, NY 10017
2 Park Square, Milton Park
Abingdon, Oxon OX14 4RN, UK

K22164

ISBN: 978-1-4822-1959-3
9 0000

9 781482 219593

WWW.CRCPRESS.COM