

Structural Vector Autoregressive Analysis

Lutz Kilian and Helmut Lütkepohl

Structural Vector Autoregressive Analysis

Structural vector autoregressive (VAR) models are important tools for empirical work in macroeconomics, finance, and related fields. This book not only reviews the many alternative structural VAR approaches discussed in the literature, but also highlights their pros and cons in practice. It provides guidance to empirical researchers as to the most appropriate modeling choices and methods of estimating and evaluating structural VAR models. The book traces the evolution of the structural VAR methodology and contrasts it with other common methodologies, including dynamic stochastic general equilibrium (DSGE) models. It is intended as a bridge between the often quite technical econometric literature on structural VAR modeling and the needs of empirical researchers. The focus is not on providing the most rigorous theoretical arguments, but on enhancing the reader's understanding of the methods in question and their assumptions. Empirical examples are provided for illustration.

Lutz Kilian is Professor of Economics at the University of Michigan. Between 2001 and 2003 he served as an adviser to the European Central Bank in Frankfurt am Main, Germany. Professor Kilian has been a research visitor at the Federal Reserve Board, the Bank of Canada, the European Central Bank, and the International Monetary Fund. His work has appeared in *Econometrica*, the *American Economic Review*, and the *Journal of Political Economy*, among other leading journals. He has served as associate editor of the *Journal of Business and Economic Statistics*, the *Journal of Development Economics*, and the *Journal of Economic Dynamics and Control*.

Helmut Lütkepohl has held professorial positions at the University of Hamburg, the University of Kiel, the Humboldt University in Berlin, the European University Institute in Florence, and the Free University of Berlin. He has served as Dean of the Graduate Center of the German Institute for Economic Research in Berlin. He has published in *Econometrica*, the *Journal of Econometrics*, the *Journal of Business and Economic Statistics*, *Econometric Theory*, and the *Journal of Applied Econometrics*. He has also served as associate editor of the *Journal of Econometrics*, *Econometric Theory*, *Macroeconomic Dynamics*, the *Journal of Applied Econometrics*, and *Econometric Reviews*. He is the author of *New Introduction to Multiple Time Series Analysis*.

Themes in Modern Econometrics

Series Editor

PETER C. B. PHILLIPS, *Sterling Professor of Economics, Yale University*

Themes in Modern Econometrics provides an organized sequence of advanced textbooks in econometrics aimed directly at the student population and is the first series in the discipline to have this as its express aim. Written at a level accessible to those who have completed an introductory course in econometrics, each book addresses topics and themes that students and researchers encounter daily. All areas of econometrics are covered within the series. Particular emphasis is given to theory fundamentals and practical implementation in breaking research areas that are relevant to empirical applications. Each book stands alone as an authoritative survey in its own right. The distinct emphasis throughout is on pedagogic excellence and accessibility.

Books in the Series

Almost All about Unit Roots: Foundations, Developments, and Applications (2015) IN CHOI

Granularity Theory with Applications to Finance and Insurance (2014)
PATRICK GAGLIARDINI and CHRISTIAN GOURIÉROUX

Econometric Modeling with Time Series (2012) VANCE MARTIN,
STAN HURN, and DAVID HARRIS

Economic Modeling and Inference (2007) JEAN-PIERRE FLORENSE,
VELAYOUDOM MARIMOUTOU, and ANNE PEGUIN-FEISSOLLE
Translated by JOSEF PERKTOLD and MARINE CARRASCO

*Introduction to the Mathematical and Statistical Foundations of
Econometrics* (2004) HERMAN J. BIERENS

Applied Time Series Econometrics (2004) HELMUT LÜTKEPOHL and
MARKUS KRÄTZIG

Semiparametric Regression for the Applied Econometrician (2003)
ADONIS YATCHEW

The Econometric Analysis of Seasonal Time Series (2001)
ERIC GHYSELS and DENISE R. OSBORN

Econometrics of Qualitative Dependent Variables (2000)
CHRISTIAN GOURIEROUX Translated by PAUL B. KLASSEN

Nonparametric Econometrics (1999) ADRIAN PAGAN and AMAN ULLAH
Generalized Method of Moments Estimation (1999)

Edited by LÁSZLÓ MÁTYÁS
Unit Roots, Cointegration, and Structural Change (1999)

G.S. MADDALA and IN-MOO KIM
Time Series and Dynamic Models (1997) CHRISTIAN GOURIEROUX and
ALAIN MONFORT Translated and edited by GIAMPIERO GALLO

Statistics and Econometric Models: Volumes 1 and 2 (1995)
CHRISTIAN GOURIEROUX and ALAIN MONFORT

Translated by QUANG VUONG

STRUCTURAL VECTOR AUTOREGRESSIVE ANALYSIS

LUTZ KILIAN

University of Michigan

HELMUT LÜTKEPOHL

DIW and Freie Universität Berlin



CAMBRIDGE
UNIVERSITY PRESS

CAMBRIDGE

UNIVERSITY PRESS

University Printing House, Cambridge CB2 8BS, United Kingdom

One Liberty Plaza, 20th Floor, New York, NY 10006, USA

477 Williamstown Road, Port Melbourne, VIC 3207, Australia

4843/24, 2nd Floor, Ansari Road, Daryaganj, Delhi - 110002, India

79 Anson Road, #06-04/06, Singapore 079906

Cambridge University Press is part of the University of Cambridge.

It furthers the University's mission by disseminating knowledge in the pursuit of education, learning, and research at the highest international levels of excellence.

www.cambridge.org

Information on this title: www.cambridge.org/9781107196575

DOI: 10.1017/9781108164818

© Lutz Kilian and Helmut Lütkepohl 2017

This publication is in copyright. Subject to statutory exception and to the provisions of relevant collective licensing agreements, no reproduction of any part may take place without the written permission of Cambridge University Press.

First published 2017

Printed in The United States of America by Sheridan Books, Inc.

A catalogue record for this publication is available from the British Library.

ISBN 978-1-107-19657-5 Hardback

ISBN 978-1-316-64733-2 Paperback

Cambridge University Press has no responsibility for the persistence or accuracy of URLs for external or third-party Internet Web sites referred to in this publication and does not guarantee that any content on such Web sites is, or will remain, accurate or appropriate.

Contents

<i>Preface</i>	<i>page</i> xvii
1 Introduction	1
1.1 Overview	1
1.2 Outline of the Book	4
2 Vector Autoregressive Models	19
2.1 Stationary and Trending Processes	19
2.2 Linear VAR Processes	23
2.2.1 The Basic Model	23
2.2.2 The Moving Average Representation	26
2.2.3 VAR Models as an Approximation to VARMA Processes	27
2.2.4 Marginal Processes, Measurement Errors, Aggregation, Variable Transformations	28
2.3 Estimation of VAR Models	30
2.3.1 Least-Squares Estimation	31
2.3.2 Restricted Generalized Least Squares	34
2.3.3 Bias-Corrected LS	35
2.3.4 Maximum Likelihood Estimation	38
2.3.5 VAR Processes in Levels with Integrated Variables	41
2.3.6 Sieve Autoregressions	43
2.4 Prediction	46
2.4.1 Predicting from Known VAR Processes	46
2.4.2 Predicting from Estimated VAR Processes	47
2.5 Granger Causality Analysis	48
2.6 Lag-Order Selection Procedures	51
2.6.1 Top-Down Sequential Testing	51
2.6.2 Bottom-Up Sequential Testing	52
2.6.3 Information Criteria	54

2.6.4	Recursive Mean-Squared Prediction Error Rankings	57
2.6.5	The Relative Merits of Alternative Lag-Order Selection Tools	58
2.7	Model Diagnostics	66
2.7.1	Tests for Autocorrelation in the Innovations	67
2.7.2	Tests for Nonnormality	67
2.7.3	Residual ARCH Tests	68
2.7.4	Time Invariance	69
2.8	Subset VAR Models, AVAR Models, and VARX Models	72
2.8.1	Subset VAR Models	72
2.8.2	Asymmetric VAR Models	73
2.8.3	VARX Models	74
3	Vector Error Correction Models	75
3.1	Cointegrated Variables and Vector Error Correction Models	75
3.1.1	Common Trends and Cointegration	75
3.1.2	Deterministic Terms in Cointegrated Processes	80
3.2	Estimation of VARs with Integrated Variables	82
3.2.1	The VAR(1) Case	82
3.2.2	Estimation of VECMs	86
3.2.3	Estimation of Levels VAR Models with Integrated Variables	95
3.3	Model Specification	99
3.3.1	Choosing the Lag Order	99
3.3.2	Specifying the Cointegrating Rank	100
3.4	Diagnostic Tests	104
3.5	The Benefits of the VECM Representation	105
3.6	Practical Issues	105
3.6.1	Limitations of Tests for Unit Roots and Cointegration	106
3.6.2	Alternative Approaches	106
4	Structural VAR Tools	109
4.1	Structural Impulse Responses	110
4.2	Forecast Error Variance Decompositions	113
4.3	Historical Decompositions	116
4.4	Forecast Scenarios	123
4.4.1	Conditional Forecasts Expressed in Terms of Sequences of Structural Shocks	124
4.4.2	Conditional Forecasts Expressed in Terms of Sequences of Observables	130
4.5	Simulating Counterfactual Outcomes	131
4.6	Policy Counterfactuals	136

5 Bayesian VAR Analysis	140
5.1 Basic Terms and Notation	141
5.1.1 Prior, Likelihood, Posterior	141
5.1.2 Bayesian Estimation and Inference	142
5.1.3 Simulating the Posterior Distribution	145
5.2 Priors for Reduced-Form VAR Parameters	149
5.2.1 General Procedures for Choosing the Parameters of Prior Densities	150
5.2.2 Normal Prior for the VAR Parameters for Given Σ_u	151
5.2.3 The Original Minnesota Prior	155
5.2.4 The Natural Conjugate Gaussian-Inverse Wishart Prior	162
5.2.5 The Independent Gaussian-Inverse Wishart Prior	166
5.3 Extensions and Related Issues	169
6 The Relationship between VAR Models and Other Macroeconometric Models	171
6.1 The Relationship between VAR Models and Traditional Dynamic Simultaneous Equations Models	171
6.1.1 The VAR Representation of Traditional DSEMs	172
6.1.2 Incredible Restrictions in Traditional DSEMs	174
6.1.3 Structural VAR Models as an Alternative to Traditional DSEMs	176
6.2 The Relationship between VAR Models and DSGE Models	177
6.2.1 Basics	177
6.2.2 The Role of Data Transformations	180
6.2.3 Why Not Use VARMA Models?	180
6.2.4 Autoregressive Sieve Approximations of $\text{VAR}(\infty)$ Processes	181
6.2.5 Summary of Potential Problems in Approximating DSGE Models with VAR Models	182
6.3 DSGE Models as an Alternative to VAR Models?	183
6.3.1 Calibrated DSGE Models	184
6.3.2 Estimated DSGE Models	185
6.3.3 Calibration versus Bayesian Estimation	186
6.3.4 Are Structural VAR Models Less Credible than DSGE Models?	187
6.3.5 Are DSGE Models More Accurate than VAR Models?	189
6.3.6 Policy Analysis in DSGE Models and SVAR Models	191
6.4 An Overview of Alternative Structural Macroeconometric Models	193

6.4.1 Combining DSEMs and SVAR Models	193
6.4.2 Combining DSGE and SVAR Models	194
7 A Historical Perspective on Causal Inference in Macroeconomics	196
7.1 A Motivating Example	196
7.2 Granger Causality Tests for Covariance Stationary VAR Models	197
7.3 Granger Causality, Predeterminedness, and Exogeneity	199
7.3.1 Basic Concepts	199
7.3.2 Granger Causality and Forward-Looking Behavior	201
7.3.3 Strict Exogeneity in Modern Macroeconomic Models	203
7.4 The Demise of Granger Causality Tests in Macroeconomics	204
7.5 Responses to Unanticipated Changes in Money Growth	205
7.5.1 The Narrative Approach	205
7.5.2 Exogenous Shocks Derived from Data-Based Counterfactuals	208
7.5.3 News Shocks	209
7.5.4 Shocks to Financial Market Expectations	210
7.5.5 Summary	211
7.6 Structural VAR Shocks	211
7.6.1 The Identification Problem	212
7.6.2 The Relationship between Structural VAR Shocks and Direct Shock Measures	213
7.6.3 Causality in Structural VAR Models	214
8 Identification by Short-Run Restrictions	216
8.1 Introduction	216
8.2 Recursively Identified Models	219
8.3 Sources of Identifying Restrictions	221
8.4 Examples of Recursively Identified Models	224
8.4.1 A Simple Macroeconomic Model	224
8.4.2 A Model of the Global Market for Crude Oil	225
8.4.3 Oil Price Shocks and Stock Returns	226
8.4.4 Models of the Transmission of Energy Price Shocks	227
8.4.5 Semistructural Models of Monetary Policy	228
8.4.6 The Permanent Income Model of Consumption	234
8.5 Examples of Nonrecursively Identified Models	235
8.5.1 Fiscal Policy Shocks	236
8.5.2 An Alternative Simple Macroeconomic Model	237
8.5.3 Discussion	237
8.5.4 The Graph-Theoretic Approach	238
8.6 Summary	239

9 Estimation Subject to Short-Run Restrictions	241
9.1 Model Setup	241
9.2 Method-of-Moments Estimation	242
9.2.1 Recursively Identified Models	242
9.2.2 Nonrecursively Identified Models	250
9.2.3 GMM Estimation of Overidentified Models	253
9.3 Instrumental Variable Estimation	258
9.4 Full Information Maximum Likelihood Estimation	262
9.5 Bayesian Estimation	265
9.6 Summary	268
10 Identification by Long-Run Restrictions	269
10.1 The Traditional Framework for Imposing Long-Run Restrictions	269
10.2 A General Framework for Imposing Long-Run Restrictions	272
10.2.1 The Long-Run Multiplier Matrix	272
10.2.2 Identification of Structural Shocks	275
10.3 Examples of Long-Run Restrictions	278
10.3.1 A Real Business Cycle Model with and without Nominal Variables	278
10.3.2 A Model of Neutral and Investment-Specific Technology Shocks	282
10.3.3 A Model of Real and Nominal Exchange Rate Shocks	284
10.3.4 A Model of Expectations about Future Productivity	284
10.4 Examples of Models Combining Long-Run and Short-Run Zero Restrictions	287
10.4.1 The IS-LM Model Revisited	287
10.4.2 A Model of the Neoclassical Synthesis	289
10.4.3 A U.S. Macroeconomic Model	290
10.5 Limitations of Long-Run Restrictions	292
10.5.1 Long-Run Restrictions Require Exact Unit Roots	292
10.5.2 Sensitivity to Omitted Variables	293
10.5.3 Lack of Robustness at Lower Data Frequencies	294
10.5.4 Nonuniqueness Problems without Additional Sign Restrictions	294
10.5.5 Sensitivity to Data Transformations	296
11 Estimation Subject to Long-Run Restrictions	297
11.1 Model Setup	297
11.2 Models Subject to Long-Run Restrictions Only	299
11.2.1 Method-of-Moments Estimation	301

11.2.2	Full Information Maximum Likelihood Estimation	306
11.2.3	Instrumental Variable Estimation	307
11.3	Models Subject to Long-Run and Short-Run Restrictions	310
11.3.1	Estimating the Model in VAR Representation	310
11.3.2	Estimating the Model in VECM Representation	316
11.4	Practical Limitations of Long-Run Restrictions	320
11.4.1	Estimators of the Long-Run Multiplier Matrix May Be Unreliable	321
11.4.2	Lack of Power	321
11.4.3	Near-Observational Equivalence of Shocks with Permanent Effects and Shocks with Persistent Effects	322
11.4.4	Weak Instrument Problems	322
11.5	Can Structural VAR Models Recover Responses in DSGE Models?	323
11.5.1	The Origin of This Controversy	323
11.5.2	The Position of Chari et al. (2008)	325
11.5.3	The Position of Christiano et al. (2006)	327
11.5.4	Understanding the Simulation Evidence	328
11.5.5	Summary	331
12	Inference in Models Identified by Short-Run or Long-Run Restrictions	334
12.1	Delta Method Intervals for Structural Impulse Responses	335
12.1.1	Finite-Order VAR Models	336
12.1.2	Infinite-Order VAR Models	338
12.1.3	Discussion	339
12.1.4	Extensions to Other Statistics	339
12.1.5	On the Choice of the Significance Level	340
12.2	Bootstrap Intervals for Structural Impulse Responses	340
12.2.1	The Standard Residual-Based Recursive-Design Bootstrap	341
12.2.2	The Standard Residual-Based Fixed-Design Bootstrap	345
12.2.3	The Residual-Based Wild Bootstrap	345
12.2.4	Bootstrapping Tuples of Regressands and Regressors	347
12.2.5	Block Bootstrap Methods	348
12.2.6	Alternative Bootstrap Confidence Intervals	356
12.3	Bootstrap Intervals Based on Bias-Adjusted Estimators	363
12.4	Potential Pitfalls in Impulse Response Inference	365
12.5	Finite-Sample Properties of Bootstrap Confidence Intervals	368

12.6	Inference for Integrated and Cointegrated VAR Processes	369
12.6.1	VAR Models in Differences	369
12.6.2	Vector Error Correction Models	370
12.6.3	Integrated and/or Cointegrated VAR Models in Levels	373
12.7	Inference in Local-to-Unity VAR Processes	377
12.7.1	Local-to-Unity Asymptotics	378
12.7.2	Inference in Levels for Local-to-Unity VAR Models	381
12.7.3	The Grid Bootstrap Method	382
12.7.4	A Hybrid Method	384
12.7.5	Implications for Second-Stage Inference after Pretesting	385
12.8	Local Projections	389
12.9	Synthesis	393
12.10	Bayesian Regions of Highest Posterior Density	394
12.10.1	Pointwise Inference on Structural Impulse Responses	395
12.11	Joint Inference on Structural Impulse Responses	398
12.11.1	Joint Confidence Sets for Structural Impulse Responses	399
12.11.2	Joint Credible Sets	406
12.12	Other Bootstrap Applications	410
12.12.1	Bootstrap Prediction	410
12.12.2	Bootstrapping the Critical Values of Test Statistics	411
12.13	Examples of Impulse Response Confidence Intervals	412
12.13.1	An Exactly Identified Model	412
12.13.2	Guarding against Conditional Heteroskedasticity	415
12.13.3	Extensions to Overidentified Models	416
13	Identification by Sign Restrictions	421
13.1	A Model of Demand and Supply	421
13.2	How to Impose Static Sign Restrictions	424
13.2.1	Givens Rotation Matrices	426
13.2.2	The Householder Transformation	427
13.2.3	The Ouliaris-Pagan Approach	428
13.3	Partially Identified VAR Models	430
13.4	Beyond Static Sign Restrictions	432

13.4.1	Dynamic Sign Restrictions	432
13.4.2	Elasticity Bounds	432
13.4.3	Shape Restrictions	435
13.5	Can Sign Restrictions Be Verified?	435
13.6	Estimation and Inference in Sign-Identified VAR Models	437
13.6.1	Frequentist Approaches	438
13.6.2	Bayesian Approaches	440
13.6.3	Evaluating the Posterior of the Structural Impulse Responses	442
13.6.4	The Penalty Function Approach	448
13.6.5	Using Historical Information to Narrow the Set of Admissible Models	451
13.7	The Role of the Prior for the Rotation Matrix	452
13.7.1	An Approach Based on Explicit Bayesian Priors for B_θ	453
13.7.2	An Approach Based on Explicit Bayesian Priors for the Structural Impulse Responses	459
13.7.3	A Robust Bayesian Approach	461
13.7.4	An Agnostic Bayesian Approach	462
13.7.5	A Non-Bayesian Approach	463
13.8	Examples of Models Identified by Sign Restrictions	464
13.8.1	A Small-Scale Macroeconomic Model	464
13.8.2	A Slightly Larger Macroeconomic Model	465
13.8.3	A Model of Unemployment and Vacancies	466
13.8.4	An Extended Model of Unemployment and Vacancies	466
13.8.5	A Model of Technology Shocks	467
13.8.6	A Model of Exchange Rate Responses to Monetary Policy Shocks	467
13.8.7	A Medium-Scale Macroeconomic Model	468
13.8.8	A Model of Speculation in the Global Oil Market	469
13.9	Mixing Sign and Exclusion Restrictions	471
13.9.1	Examples of Models Mixing Sign and Short-Run Zero Restrictions	471
13.9.2	How to Combine Sign Restrictions and Exclusion Restrictions	474
13.9.3	Discussion	482
13.10	Empirical Illustrations	483
13.10.1	A Model of the Global Oil Market	483
13.10.2	A Model of Monetary Policy	485
13.11	Concluding Remarks	488

14 Identification by Heteroskedasticity or Non-Gaussianity	491
14.1 Introduction	491
14.2 The Model Setup	492
14.2.1 The Baseline Model	492
14.2.2 An Illustrative Example	494
14.2.3 The General Model	495
14.3 Alternative Volatility Models	496
14.3.1 Structural VAR Models with Extraneously Specified Volatility Changes	496
14.3.2 Structural VAR Models with Markov Switching in the Variances	505
14.3.3 Structural VAR Models with Smooth Transitions in the Variances	511
14.3.4 Structural VAR Models with GARCH Errors	517
14.4 Alternative Approaches Using Heteroskedasticity	524
14.4.1 Time-Varying Instantaneous Effects	525
14.4.2 Correlated Shocks	525
14.5 Identification by Non-Gaussianity	526
14.5.1 Independent Shocks	526
14.5.2 Uncorrelated Shocks	528
14.6 Discussion	530
15 Identification Based on Extraneous Data	532
15.1 Identification Based on High-Frequency Futures Prices	532
15.1.1 A Set-Identified Approach	534
15.1.2 A Point-Identified Approach	538
15.1.3 Discussion	542
15.2 Identification Based on External Instruments	542
15.2.1 Estimation and Inference	544
15.2.2 Discussion	548
16 Structural VAR Analysis in a Data-Rich Environment	549
16.1 Factor Models	551
16.1.1 Static Factor Models	551
16.1.2 Dynamic Factor Models	555
16.1.3 Selecting the Number of Factors	562
16.1.4 Structural Change	565
16.2 Factor-Augmented Structural VAR Models and Related Techniques	565
16.2.1 Structural FAVAR Models	565
16.2.2 Structural Analysis with DFM	571
16.2.3 Empirical Examples of FAVAR Models and DFM	577

16.3	Large Bayesian VAR Models	579
16.3.1	Priors for Large Bayesian VARs	580
16.3.2	Structural Identification in Large BVARs	583
16.4	Alternative Large-Dimensional VAR Models	584
16.4.1	Panel VARs	584
16.4.2	Global VARs	586
16.4.3	Spatial Models	587
16.5	Discussion	587
17	Nonfundamental Shocks	590
17.1	Introduction	590
17.2	Fundamental and Nonfundamental Moving Average Representations	592
17.3	Fundamental versus Nonfundamental Representations	594
17.3.1	Nonfundamental Shocks in Economic Models	594
17.3.2	Nonfundamentalness Due to MA Roots in the Unit Circle	596
17.3.3	Nonfundamentalness Due to Omitted Variables	597
17.3.4	Avoiding Nonfundamentalness by Using Factor-Augmented or Large Bayesian VARs	601
17.3.5	Other Approaches to Dealing with Anticipation	603
17.4	Conclusions	607
18	Nonlinear Structural VAR Models	609
18.1	Motivation	609
18.2	Nonlinear VAR Analysis	612
18.2.1	General Setup	612
18.2.2	Structural Analysis	614
18.3	Threshold and Smooth-Transition VAR Models	619
18.3.1	Model Setup	619
18.3.2	Example: A TVAR Model of U.S. Monetary Policy	621
18.4	Markov-Switching VAR Models	622
18.4.1	Model Setup	623
18.4.2	Identification	625
18.4.3	Estimation	626
18.4.4	Model Selection	628
18.4.5	Example: An MS-VAR Model of U.S. Monetary Policy	629
18.5	Time-Varying Coefficient VAR Models	630
18.5.1	Model Setup	631
18.5.2	Estimation	633

18.5.3	Example: A TVC-VAR Model of U.S. Monetary Policy	635
18.6	VAR Models with GARCH-in-Mean	636
18.6.1	Model Setup	636
18.6.2	Estimation	637
18.6.3	Example: The Effect of Oil Price Uncertainty on U.S. Real Output	638
18.7	Other Nonlinear Models	640
18.7.1	Nonparametric VAR Analysis	640
18.7.2	Noncausal VAR Models	645
18.8	Discussion of Nonlinear VAR Modeling	648
18.9	Linear Structural Models with Nonlinear Transformations of the Variables	650
18.9.1	The Censored Oil Price VAR Model	651
18.9.2	A Nonlinear Structural Model Allowing for Asymmetric Responses	652
18.9.3	Quantifying Nonlinear Responses to Oil Price Shocks	653
18.9.4	Testing the Null of Unconditionally Symmetric Response Functions	654
18.9.5	Testing the Null of Conditionally Symmetric Response Functions	655
18.9.6	Testing the Null of No Time Dependence	656
18.9.7	Conditional Prediction Error Decompositions	657
18.9.8	Extensions	658
19	Practical Issues Related to Trends, Seasonality, and Structural Change	659
19.1	Alternative Trend Models	659
19.1.1	Hodrick-Prescott (HP) Filter	659
19.1.2	Band-Pass Filters	660
19.1.3	Potential Shortcomings of Trend Filters	661
19.1.4	Trend-Filtered Variables in VAR Models	661
19.1.5	Choosing between Different Trend Models	662
19.1.6	Combining Different Trend Specifications	662
19.2	Seasonality	663
19.2.1	Deterministic Seasonal Variation in VAR Models	663
19.2.2	Stochastic Seasonal Variation in VAR Models	664
19.2.3	Synthesis	666
19.2.4	Periodic Seasonal VAR Models	666
19.2.5	Seasonal TVC-VAR Models	667

19.2.6	Seasonally Filtered Data in VAR Models	667
19.2.7	Combining Seasonally Adjusted and Unadjusted Data in the same VAR Model	668
19.2.8	Summary	668
19.3	Structural Change in the Stochastic Component of the VAR Model	669
19.3.1	Breaks in the Stochastic Component	669
19.3.2	Smooth Structural Change in the Stochastic Component	671
<i>Bibliography</i>		673
<i>Notation and Abbreviations</i>		713
<i>Author Index</i>		721
<i>Subject Index</i>		729

Preface

Objectives of the Book

Since the seminal work of Sims (1980a), structural vector autoregressions have evolved into one of the most widely used models in empirical research using time series data. They are used in macroeconomics and in empirical finance, but also in many other fields including agricultural economics and energy economics. The evolution of the structural vector autoregressive (VAR) methodology since 1980 has not always been smooth. Over time many new ideas have been explored, sometimes uncritically applied or misunderstood by practitioners, then questioned, and later refined or replaced by alternative methods. The development of new methods of identification, estimation, and inference for structural VAR models continues at a rapid pace even today. One of the objectives of this book is to summarize these new developments and to put them in perspective. The other is to take stock of what we have learned about more traditional structural VAR models and to interpret these models from today's perspective. The profession's understanding of these models has evolved substantially, becoming more nuanced in recent years and allowing us to understand better some of the methodological debates of the past.

In this book, we not only review the ever-increasing range of structural VAR tools and methods discussed in the literature; we also highlight their pros and cons in practice and provide guidance to empirical researchers as to the most appropriate modeling choices. In addition, we trace the evolution of the structural VAR methodology and contrast it with other common methodologies including the narrative approach to identification and the use of calibrated or estimated dynamic stochastic general equilibrium (DSGE) models. We stress that structural VAR models should be viewed as one of several econometric tools used in empirical work, each of which has its own strengths and weaknesses.

The book is intended as a bridge between the often quite technical econometric literature on structural VAR modeling and the needs of empirical

researchers. The focus of the book is not on providing the most rigorous theoretical arguments, but on enhancing the reader's understanding of the methods in question and their assumptions, allowing him or her to decide on the most suitable methods for applied work. In many cases, empirical examples are provided for illustration. References to articles in academic journals are provided for readers with an interest in the more technical aspects of the discussion.

Audience and Uses of the Book

The target audience includes graduate students in economics departments and in business schools as well as practitioners interested in a comprehensive yet accessible review of the literature. It also includes consumers of empirical studies using the structural VAR methodology.

The book is intended for a semester-long course on structural VAR analysis, but the material may be adapted to the time available to the instructor and the focus of the course. Parts of the book may also be used in teaching a course in macroeometrics or in applied time series analysis, if preceded by a general introduction to univariate time series analysis in the first part of the course. Alternatively, selected chapters may be used in teaching a graduate-level quantitative methods course that focuses on the use of DSGE models and structural VAR models in macroeconomics.

For example, an instructor only interested in structural VAR models subject to short-run identifying restrictions would focus on Chapters 2, 4, 8, 9, and possibly 12, with the material in Chapters 6 and 7 providing additional motivation as needed. If one wanted to cover structural VAR models subject to long-run restrictions, one would add Chapters 3, 10, and 11. An extension to sign-identified structural VAR models instead would involve adding Chapters 5 and 13. Chapters 14 and 15 cover more advanced identification methods. Chapters 16, 17, and 18 deal with special topics such as large-dimensional VAR processes, nonfundamental shocks, and nonlinear structural VAR models that are more technically challenging and would only be covered in a full-semester graduate-level course. Chapter 19 covers topics such as trend adjustment, seasonality, and structural change. It relates to material that could be skipped at a first reading or used only selectively, but provides useful background material at a later stage, once the remainder of the book has been absorbed.

The book may not only form the basis of a graduate-level course, but it may also be used for self-study. Although structural VAR models are routinely relied on in teaching empirical macroeconomics and in published work, not every department provides instruction in the use of these methods. There is typically no room for teaching structural VAR analysis either in the econometric theory courses or in the macroeconomics courses of a Ph.D. program,

as these fields have greatly expanded in recent years. Our book provides a self-contained resource for students wishing to complement the material on solving DSGE models typically provided as part of the first-year macroeconomics sequence by a review of the structural VAR methodology.

The need for such a book is self-evident. It has become increasingly difficult for students, practitioners, and even academic researchers to keep up with the proliferation of new methods and econometric results discussed in the literature. Not only are these results widely scattered across academic journals but there is no up-to-date treatise even of the traditional structural VAR literature. For example, the introductory discussion in Amisano and Giannini (1997) is not only terse but incomplete and outdated at this point. The more recent macroeometrics textbook of Favero (2001) contains only one chapter on structural VAR models, which focuses primarily on VAR models of monetary policy. Even the textbook of Lütkepohl (2005) focuses mainly on reduced-form VAR analysis and devotes only one chapter to structural VAR analysis. Finally, the focus of Canova (2007), DeJong and Dave (2011), and Herbst and Schorfheide (2016) is more on the empirical evaluation of DSGE models than on structural VAR analysis, and the standard time series textbook by Hamilton (1994) only devotes 13 of 800 pages to structural vector autoregressions. Nor do handbook chapters such as Watson (1994) or Kilian (2013) provide a review of the structural VAR literature as comprehensive as this book.

Prerequisites

The book takes for granted that the reader is familiar with regression analysis and with asymptotic reasoning. It is also assumed that the reader has been exposed to univariate time series methods at the undergraduate level and is familiar with the concepts of stationarity and invertibility and with the Wold representation, white noise, unit roots, the estimation of univariate AR, MA, and ARMA models, GARCH models, and univariate forecasts. As a rule, the discussion is not more mathematical than it has to be to appreciate the material. An overview of the notation is provided at the end of the book. The reader is assumed to be familiar with vectors and matrices. A useful resource summarizing key mathematical and statistical results is appendix A-C in Lütkepohl (2005).

Acknowledgments

We especially thank Christiane Baumeister, Silvia Gonçalves, Nikolai Gospodinov, Atsushi Inoue, Markku Lanne, Aleksei Netsunajev, and José Montiel Olea, who provided in-depth reviews of several draft chapters and spent more time than one could have reasonably expected to answer our questions. We have also benefited from detailed advice and feedback on

specific technical issues provided by Jonas Arias, Jörg Breitung, Fabio Canova, Eric Leeper, Adrian Pagan, Elena Pesavento, Barbara Rossi, Pentti Saikkonen, Frank Schorfheide, and Dan Waggoner.

In addition, a large number of colleagues and students has commented on parts of the book, correcting errors, providing suggestions, and helping us improve the exposition. We cannot mention all of them, but we would like to single out Niels Aka, Benjamin Beckers, Martin Bruns, Wenjuan Chen, Frank Diebold, Michael Hachula, Christoph Hanck, Ana Maria Herrera, Simon Jurkatis, Tatsiana Kliatskova, Hans-Martin Krolzig, Sorin Pascu, Peter Phillips, Michele Piffer, Max Podstawska, Thore Schlaak, Annika Schnücker, Gregor von Schweinitz, Lars Winkelmann, and Tomasz Wozniak.

Finally, we would like to thank Peter Phillips for his encouragement and persistent efforts to have the book included in the *Themes in Modern Econometrics* series. Much of this book was completed while the second author was a Bundesbank Professor at the Freie Universität Berlin. Financial support was provided by the Deutsche Forschungsgemeinschaft through the SFB 649 *Economic Risk*.

Ann Arbor and Berlin,

Lutz Kilian and Helmut Lütkepohl

1 Introduction

1.1 Overview

The vector autoregressive (VAR) model is a widely used model for multivariate time series analysis. It consists of a system of regression equations. VAR models are estimated by regressing each model variable on lags of its own as well as lags of the other model variables up to some prespecified maximum lag order, p . A VAR model with p autoregressive lags is referred to as a VAR(p) model. VAR models are based on the notion that every model variable depends on its own lags as well as the lags of every other model variable, rendering exclusion restrictions on the interaction of lagged model variables not credible.

VAR models are typically based on monthly or quarterly data. For example, let y_t denote a K -dimensional vector of time series data, consisting of U.S. real GNP growth (Δgnp_t), the U.S. rate of inflation (π_t), and the U.S. short-term nominal interest rate (i_t), for $t = 1, \dots, T$. Then, suppressing the intercept, a quarterly VAR(2) model for these three variables may be written as a system of three equations

$$\begin{aligned}\Delta gnp_t &= a_{11,1} \Delta gnp_{t-1} + a_{12,1} \pi_{t-1} + a_{13,1} i_{t-1} \\ &\quad + a_{11,2} \Delta gnp_{t-2} + a_{12,2} \pi_{t-2} + a_{13,2} i_{t-2} + u_{1t} \\ \pi_t &= a_{21,1} \Delta gnp_{t-1} + a_{22,1} \pi_{t-1} + a_{23,1} i_{t-1} \\ &\quad + a_{21,2} \Delta gnp_{t-2} + a_{22,2} \pi_{t-2} + a_{23,2} i_{t-2} + u_{2t} \\ i_t &= a_{31,1} \Delta gnp_{t-1} + a_{32,1} \pi_{t-1} + a_{33,1} i_{t-1} \\ &\quad + a_{31,2} \Delta gnp_{t-2} + a_{32,2} \pi_{t-2} + a_{33,2} i_{t-2} + u_{3t},\end{aligned}\tag{1.1.1}$$

where $K = 3$ and the zero mean innovations u_{it} , $i = 1, 2, 3$, are serially uncorrelated if the maximum lag order has been chosen appropriately. The model allows these innovations to be mutually correlated with covariance matrix Σ_u . More compactly, we can express this VAR(2) model as

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + u_t,\tag{1.1.2}$$

where

$$y_t = \begin{pmatrix} \Delta gnp_t \\ \pi_t \\ i_t \end{pmatrix}, \quad A_i = \begin{bmatrix} a_{11,i} & a_{12,i} & a_{13,i} \\ a_{21,i} & a_{22,i} & a_{23,i} \\ a_{31,i} & a_{32,i} & a_{33,i} \end{bmatrix}, \quad i = 1, 2,$$

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \end{pmatrix}.$$

The innovation vector u_t is the linearly unpredictable component of y_t , given an information set consisting of the lagged values of all three model variables.

In the language of the literature on dynamic simultaneous equations models such a model is known as a reduced form, defined as a model that expresses the current values of the data as a linear function only of its own lagged values and lagged values of the other model variables. The reduced-form VAR model may be viewed as a finite-order approximation to a general linear process. This model has proved useful for summarizing the properties of the data, for forecasting, for testing for the existence of equilibrium relationships tying together two or more economic variables, and for quantifying the speed with which the model variables revert back to the equilibrium following a disturbance.

This book instead focuses on the use of VAR models for structural modeling. The premise is that we can think of the reduced-form VAR(p) model as representing data generated from the structural VAR(p) model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t, \quad (1.1.3)$$

where y_t is the $K \times 1$ vector of observed time series data, $t = 1, \dots, T$, and the deterministic terms have been suppressed for convenience. Furthermore, B_i , $i = 1, \dots, p$, is a $K \times K$ matrix of autoregressive slope coefficients, the $K \times K$ matrix B_0 reflects the instantaneous relations among the model variables, and the $K \times 1$ vector of mean zero structural shocks w_t is serially uncorrelated with a diagonal covariance matrix Σ_w of full rank such that the number of shocks coincides with the number of variables. The $K \times K$ matrix B_0^{-1} captures the impact effects of each of the structural shocks on each of the model variables. Model (1.1.3) is structural in that the shocks are postulated to be mutually uncorrelated with each element of w_t having a distinct economic interpretation. This fact allows one to interpret movements in the data caused by any one element of w_t as being caused by that shock. The structural shocks in general are not directly observable, but under suitable conditions they may be recovered from the reduced-form representation of model (1.1.3).

The reduced-form representation of model (1.1.3) can be obtained by premultiplying both sides of (1.1.3) by B_0^{-1} , resulting in the model

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad (1.1.4)$$

where $A_i = B_0^{-1}B_i$ and $u_t = B_0^{-1}w_t$, and may be estimated by unrestricted least-squares (LS) or maximum likelihood (ML) estimation methods. Model (1.1.1) is easily recognized as an example of such a model. It is readily apparent that, given an estimate of this reduced form, all that is required for recovering the structural model (1.1.3) is knowledge of the structural impact multiplier matrix B_0^{-1} (or, equivalently, of its inverse B_0).

Estimation of the matrix B_0 requires additional restrictions on the data generating process (DGP). If the matrix B_0 can be solved for, given these restrictions and the data, we say that the structural VAR model parameters, $(B_0, B_1, \dots, B_p, \Sigma_w)$, are identified or, equivalently, that the structural shocks $w_t = B_0 u_t$ are identified. The problem of finding suitable economically credible restrictions on B_0^{-1} or B_0 is known as the identification problem in structural VAR analysis. Much of this book is concerned with alternative strategies for achieving identification. As with any structural econometric model, the validity of structural VAR analysis rests on the credibility of these identifying restrictions. Finding a credible set of restrictions can be challenging. Depending on the identifying assumptions, the structural VAR model may or may not be unique. In the latter case, there is a range of structural models that are observationally equivalent in that they have the same reduced-form representation.

The existence of a structural VAR model allows us to think of the variation in the data as being driven by the cumulative effects of economically interpretable structural shocks. Current observations of the data may be viewed as a weighted average of current and past structural shocks. This insight is important because it helps researchers quantify causal relationships in the data that are obscured in reduced-form VAR analysis. For expository purposes and without loss of generality, suppose that the structural shock of interest involves changes in monetary policy not in response to macroeconomic conditions. After expressing the estimate of the VAR model in a suitable form, one may answer a range of questions about the causal effects of this shock.

- One may ask by how much an unexpected monetary policy tightening in the current quarter will reduce output growth over the next two years, when that policy change occurs all else equal and is not followed by any further monetary policy shocks after the current quarter. The response of output growth to this shock over time can be quantified in the form of an impulse response function.
- One may ask how much of the variability of output growth on average is accounted for by shocks to monetary policy as opposed to other structural shocks. This question can be answered by a forecast error variance decomposition.
- One may ask how much of the recession of 1982, for example, is explained by the cumulative effects of earlier monetary policy

shocks. This question can be answered by constructing a historical decomposition.

- One may also ask by how much the recession of 1982 would have deepened had monetary policymakers not responded to output growth at all. This question, under suitable conditions, may be answered by a policy counterfactual.

This set of examples illustrates why the structural VAR framework is often useful for economic analysis. The chief advantage of the structural VAR model compared with alternative econometric approaches is that it tends to fit the data well and only involves minimal identifying restrictions. In particular, it does not impose cross-equation restrictions or exclusion restrictions on the reduced form that tend not to be robust across alternative specifications of the underlying economic model. The remainder of this book provides a more detailed discussion of the historical evolution of the structural VAR approach, of its implementation, of its properties, and of its pros and cons compared with alternative structural econometric models.

1.2 Outline of the Book

Chapter 2

Chapter 2 of the book deals with the reduced-form specification of VAR models. It introduces stochastic and deterministic trends and shows that, under weak assumptions, the purely stochastic component of a multivariate time series process may be modeled as a finite-order VAR model with white noise errors. This VAR model may be viewed either as the DGP or, more plausibly, as an approximation to a more general linear DGP. If the VAR model is stable, it may equivalently be represented as a moving average (MA) process that expresses the model variables as a weighted average of past regression errors. This MA representation is a good point of departure for studying the implications of temporal or cross-sectional aggregation of the VAR model variables, which can be shown to have important effects on the MA structure and hence on the implied reduced-form VAR representation.

In practice, the parameters of a given VAR model of finite lag order are not known and must be estimated from the data. We review not only the conventional LS and conditional ML estimators of unrestricted stationary reduced-form VAR models but also the bias-corrected LS estimator of unrestricted stationary VAR models and the generalized least-squares (GLS) estimator of restricted stationary VAR models. We then discuss how the presence of integrated variables may affect the convergence rate and asymptotic distribution of the LS and ML estimator, when estimating the model in levels, and what assumptions are required for these estimators to remain asymptotically valid

when the finite-order VAR model is only an approximation to a more general linear process such as an invertible vector autoregressive moving average (VARMA) process.

Next, we explain how to form predictions from the estimated VAR process, and derive the one-step ahead mean-squared prediction error (MSPE) matrix, which allows a formal discussion of the notion of Granger causality (or linear predictability) in the VAR model. The one-step ahead MSPE matrix also plays an important role in the design of commonly used data-based lag-order selection criteria for VAR models. Among the latter methods, we discuss top-down and bottom-up sequential testing procedures, information criteria, and recursive MSPE rankings. The latter approach is conceptually closely related to the use of information criteria. We note that data-based lag-order selection methods not only tend to select lag orders that are too low in small samples, but that they undermine the asymptotic validity of inference about the parameters of the implied VAR model. We make the case that the use of a fixed conservative lag order may circumvent these small-sample and asymptotic problems. Chapter 2 also briefly reviews standard diagnostic tests for non-normality, for serial correlation and for conditional heteroskedasticity in the regression errors, and for the time invariance of VAR model parameters.

Although unrestricted reduced-form VAR models are most common in empirical work, there have been proposals for restricting the lag structure of reduced-form VAR models on economic grounds or on statistical grounds. We conclude with a brief discussion of three classes of restricted reduced-form VAR models, including (1) subset VAR models, which allow for the maximum lag order to differ across model equations; (2) asymmetric VAR (AVAR) models, in which the maximum lag order is the same for each variable in all equations but differs across model variables; and (3) VARX models, where the X refers to the inclusion of one or more exogenous variables. Loosely speaking, a variable is exogenous if it is determined outside of the system of equations under consideration. In other words, an exogenous variable is not subject to current or lagged feedback from other VAR model variables, but depends only on its own lags (or possibly lags of other exogenous variables). For example, in a small open economy, the world interest rate may be considered exogenous with respect to the domestic economic variables.

Chapter 3

Many economic variables can be thought of as exhibiting stochastic trends. Such variables are integrated of order 1, which means that only their first difference is stationary. If two or more integrated variables share a common stochastic trend, they are referred to as cointegrated. More generally, cointegration arises when linear combinations of integrated variables are stationary. Although integrated and cointegrated models may be estimated in levels, as

discussed in Chapter 2, imposing integration and cointegration restrictions on the reduced-form VAR model improves the finite-sample accuracy of the VAR model estimates, if these restrictions are correct. Such restricted VAR parameterizations are known as vector error correction models (VECMs). If all VAR model variables are integrated, but none are cointegrated, the VECM reduces to a VAR model in first differences.

VECMs are discussed in Chapter 3. They are of special interest for structural VAR analysis because they allow economists to identify structural economic shocks by imposing restrictions on the long-run behavior of selected variables, as discussed in Chapter 10, which would not be possible if the level representation of the VAR model were stationary. Moreover, common trends implied by cointegration restrictions may have natural economic interpretations as equilibrium relationships.

We first define cointegration and then show how a reduced-form VAR model in levels may be reparameterized as a VECM with special attention to the role of deterministic terms in cointegrated processes. We discuss how VECMs may be estimated by ML methods or by feasible GLS methods, possibly subject to restrictions on the VECM parameters. Because VECMs are reparameterizations of VAR models in levels, they may alternatively be viewed as the true model or as approximations to a more general linear DGP. In the latter case, additional assumptions are needed to ensure the asymptotic validity of the VECM estimator.

Chapter 3 reexamines the question of how to choose the lag order and how to conduct diagnostic tests in the context of the VECM, stressing that many of the results in Chapter 2 are robust to the presence of integrated and cointegrated variables in the model. We also address the important question of how to specify the cointegrating rank when the nature of the cointegration relationships is not already pinned down by economic theory.

Finally, we examine the costs of estimating VECMs in levels without imposing integration and cointegration restrictions, and discuss the question of how to choose between these two specifications. Given the uncertainty about the validity of integration and cointegration restrictions in practice, and our inability to discriminate between alternative VAR and VEC model specifications reliably in small samples, a number of alternative asymptotic thought experiments have been proposed including local-to-unity asymptotics and asymptotics for fractionally integrated processes. In Chapter 3, we briefly introduce these ideas, which will come up again in Chapters 11 and 12.

Chapter 4

The structural VAR representation expresses the reduced-form VAR errors as a linear combination of structural shocks with economic interpretation. If

we know the structural impact multiplier matrix, which describes the weights attached to each structural shock contributing to the reduced-form error, we can always recover the structural VAR representation from the reduced-form VAR representation, as discussed in Chapter 4. Knowledge of the structural representation of the VAR model (or of the VECM) allows users to construct the responses of each model variable to each structural shock (known as structural impulse responses), to assess the extent to which each structural shock contributes to the variability in the model variables (known as a forecast error variance decomposition), and to assess how the data would have evolved in the absence of one or more of the structural shocks (known as a historical decomposition). The latter decomposition also allows users to simulate counterfactual outcomes, it can be used for the construction of policy counterfactuals which examine how hypothetical changes in policy rules affect economic outcomes, and it facilitates the construction of forecast scenarios which measure the extent to which a baseline forecast would change in response to certain hypothetical future events, expressed as sequences of future structural shocks. Chapter 4 reviews and illustrates each of these tools, highlighting alternative representations used in the literature. The question of how to obtain the structural impact multiplier matrix is deferred to Chapters 8, 9, 10, 11, 13, 14, 15, and 17.

Chapter 5

A substantial part of the VAR literature relies on Bayesian methods. Indeed, many of the leading contributors to the VAR literature have been Bayesian econometricians. For example, the method of using sign restrictions for identification discussed in Chapter 13 was originally developed within a Bayesian framework. It therefore is important for users of structural VAR models to understand these methods, if only to be able to interpret Bayesian estimation results reported in the literature.

Chapter 5 contrasts the central premises of the Bayesian estimation approach with those of the frequentist estimation approach discussed in Chapters 2 and 3. We review the role of the prior density of the model parameters, the likelihood of the model, and the posterior density of the model parameters in Bayesian analysis. We discuss how a Bayesian would construct a point estimate from the posterior distribution, how Bayesian credible sets differ from frequentist confidence intervals, how Bayesian model comparisons differ from classical hypothesis testing, and how model averaging may be used as an alternative to model selection.

Because of the central role played by the posterior distribution of the VAR model parameters in Bayesian estimation and inference, we provide a brief overview of methods that may be used to sample from this distribution

including direct sampling from a known posterior distribution, acceptance sampling, importance sampling, Markov Chain Monte Carlo methods, the Metropolis-Hastings algorithm, and the Gibbs sampler.

We then review the leading methods of specifying priors for the reduced-form VAR parameters in empirical work that are all based on the premise of a Gaussian VAR process. These methods include (1) a Gaussian prior for the slope parameters for a given estimated error covariance matrix, (2) the natural conjugate Gaussian-inverse Wishart prior, and (3) the independent Gaussian-inverse Wishart prior. Making these methods operational is not straightforward without further assumptions. A popular approach known as the Minnesota or Litterman prior reduces the problem of specifying a high-dimensional Gaussian prior distribution for all VAR slope parameters to one of specifying a much smaller set of hyperparameters at the cost of imposing additional parametric structure. Alternatively, priors may be imposed on the VEC representation of a VAR model, allowing for cointegration or for near unit roots. The latter approach is also known as the sum-of-coefficients prior.

Chapter 6

Chapter 6 puts the development of VAR models in historical context and clarifies their relationship with other modeling frameworks used in macroeconomics. We discuss what structural VAR models have in common with dynamic simultaneous equations models (DSEMs) and how they differ from traditional DSEMs of the type widely used in empirical macroeconomics until the 1970s. We also examine the conditions under which DSGE models, which have been popular since the 1980s, have a reduced-form VAR representation, highlighting the fact that the VAR representation of DSGE model variables, if it exists, typically will not be of finite order. Even more stringent conditions are required for DSGE models to have a structural VAR representation.

DSGE models today are the leading alternative to structural VAR models in macroeconomics. We briefly review alternative approaches of evaluating DSGE models by calibration, by frequentist, and by Bayesian estimation, stressing the commonalities and differences between calibration methods and GMM estimation, on the one hand, and calibration and Bayesian estimation, on the other. We then contrast the pros and cons of DSGE models compared with structural VAR models. As part of this discussion, we also address common misperceptions about structural VAR models not being “structural” and about DSGE models not requiring auxiliary assumptions about data transformations and lag orders. The implications of the Lucas Critique for policy analysis in DSGE models and in structural VAR models are also discussed. We conclude that DSGE models and structural VAR models are complementary with each approach having its own strengths and weaknesses. The chapter ends with a brief overview of efforts to combine elements of structural VAR models

with traditional DSEMs on the one hand and with DSGE models on the other.

Chapter 7

The central objective in structural VAR analysis is to quantify causal relationships in the data. Chapter 7 studies the precursors of structural VAR models, some of which continue to be used in empirical work to this day. Using the debate over money-income causality as our motivating example, we trace the evolution of the profession's thinking about causality from the narrative approach of Friedman and Schwartz (1963) to Granger causality tests in the 1970s with special attention to the concepts of strict exogeneity and predeterminedness. We explain why the profession lost interest in questions of Granger causality in the 1980s and began to focus on unanticipated changes in economic variables instead. The chapter focuses on the development of direct measures of exogenous monetary policy shocks, of fiscal policy shocks, of OPEC oil supply shocks, of news shocks based on macroeconomic announcements, and of shocks to financial market expectations, for example. We trace the further evolution of this literature from distributed-lag models of the impact of directly observable exogenous shocks to VAR models driven by unobserved exogenous shocks that can only be recovered with the help of additional identifying assumptions.

Chapter 8

Although there have been important advances in how one determines the specification of structural VAR models, in how structural VAR estimates are presented, and in how the estimation uncertainty is captured, the question of the identification of structural economic shocks has always been central in this literature, and this question appropriately receives the most weight in our book, starting with Chapter 8.

The chapter starts by contrasting the structural representation of VAR models with the reduced-form specification discussed in Chapters 2, 3, and 5. We discuss the nature of the identification problem in structural VAR modeling and illustrate how alternative normalizing assumptions affect the type of additional restrictions required for the order and rank conditions for exact identification to hold. Identification of unique structural shocks in a VAR model may be achieved by imposing restrictions on the structural impact multiplier matrix of the model (or alternatively on its inverse). These matrices govern the contemporaneous interaction of the model variables and/or of the structural shocks, conditional on the lagged model variables. Hence, restrictions on these matrices are commonly referred to as short-run identifying restrictions. There are several ways of reducing the number of free parameters in the structural VAR

model to be estimated, but the most common approach is to impose exclusion restrictions that limit the contemporaneous feedback between some of the model variables.

We emphasize that imposing a recursive ordering on the model variables in the impact period, as is common in applied work, amounts to imposing a particular causal chain that results in economically meaningless measures of structural shocks, unless this ordering can be economically motivated. Having reviewed common sources of economically meaningful identifying restrictions, we consider several examples of recursively identified structural VAR models with careful attention to the economic content of their identifying assumptions. We stress that credibly identifying all structural shocks by recursive orderings is feasible only in rare cases, but note that sometimes recursive orderings may be used to identify one of the structural shocks with the other structural shocks remaining unidentified from an economic point of view. We also discuss examples of nonrecursively identified structural VAR models. The chapter concludes with a brief discussion of the graph-theoretic approach to identification. We point out that such data-based approaches to identification are not designed to uncover economically meaningful structures and hence are no substitute for economic reasoning in the construction of structural VAR models.

Chapter 9

Having decided on the identifying restrictions, the question arises of how to estimate the structural VAR model. There are three common approaches. We can estimate the structural VAR model (1) by the method of moments or by instrumental-variable (IV) methods, (2) by full information maximum likelihood (FIML) methods, or (3) by Bayesian methods.

Perhaps the most common approach in applied work is the method of moments. For exactly identified models, one proceeds in two steps. One first estimates the reduced-form VAR model as described in Chapters 2 and 3. One then solves for the structural impact multiplier matrix. If the structural model is recursive, this may be accomplished by simply applying a Cholesky decomposition to the covariance matrix of the reduced-form residuals. More generally, we may use a nonlinear equation solver to solve the system of equations linking the unique elements of the covariance matrix of the reduced-form residuals to the unknown elements of the structural impact multiplier matrix. A third option that is computationally less demanding than a nonlinear equation solver (and hence particularly appealing when working with nonrecursive models) is to use the algorithm proposed by Rubio-Ramírez, Waggoner, and Zha (2010). If there are more restrictions than required for exact identification, rendering the model overidentified, we may solve the model in one step by numerically minimizing the GMM objective function. In some cases, the

method-of-moments estimator may also be constructed using traditional IV regression techniques.

Another common approach is FIML estimation, which also accommodates overidentified models, but in the latter case (like the GMM estimator) requires the use of numerical methods. Finally, Bayesian estimation is common in applied work. There are two alternative approaches. For exactly identified models it is standard to rely on conventional reduced-form priors, as discussed in Chapter 5, to generate draws from the posterior of the VAR model, from each of which an estimate of the structural impact multiplier matrix may be obtained by applying the second step of the method of moments. An alternative to this widely used approach is to specify a prior directly on the structural VAR representation, which also accommodates the overidentified case, as discussed at the end of Chapter 9.

Chapter 10

Finding enough short-run restrictions for identifying the structural shocks of interest can be a challenge in practice. One alternative idea in the literature has been to impose restrictions on the long-run response of the model variables to selected shocks. In the presence of unit roots in some variables, but not in others, this approach may allow us to identify at least some structural shocks. The use of long-run restrictions has been appealing because many economists find it easier to agree on long-run restrictions than on the short-run behavior of the economy. It is not without important drawbacks, however.

Chapter 10 introduces a general framework for imposing both short-run and long-run restrictions. We show how alternative specifications of the same model affect how long-run restrictions are imposed. We discuss a range of empirical examples of the use of long-run restrictions both in isolation and in conjunction with short-run restrictions. We draw attention to the fact that special care is needed in specifying such models to avoid some structural shocks in the model having unintended permanent effects.

The chapter concludes with an overview of the limitations of models based on long-run identifying restrictions including the fact that they require exact unit roots in some model variables, their sensitivity to omitted variables, their lack of robustness at lower data frequencies, their sensitivity to data transformations, and the fact that they yield nonunique solutions without additional normalizations.

Chapter 11

As in the case of models identified by short-run restrictions, we may estimate models identified by long-run restrictions (or by a combination of long-run and short-run restrictions) by the method of moments, by IV methods, or by FIML

methods, depending on the nature of the restrictions. Chapter 11 highlights differences in the estimation of models subject to long-run restrictions, depending on whether the model is expressed as a stationary VAR model or in VECM representation. We also review a number of practical problems with the estimation of models identified by long-run restrictions, including that the estimator of the long-run multiplier matrix may be unreliable, the near-observational equivalence of shocks with permanent effects and shocks with persistent effects, and weak instrument problems in implementing the estimator.

The chapter concludes with a review of an ongoing debate among macroeconomists over the ability of structural VAR models to recover the structural impulse responses implied by DSGE models based on synthetic data generated by these DSGE models. This debate was triggered by Galí (1999) who suggested that evidence based on structural VAR models identified by long-run exclusion restrictions had important implications for DSGE modeling. We reexamine this controversy, drawing on the insights provided in Chapters 10 and 11.

Chapter 12

Estimates of structural VAR models are subject to uncertainty. In practice, users of VAR models are typically not interested in the estimation uncertainty about the model parameters themselves, but in the uncertainty about the implied estimates of the structural impulse responses and forecast error variance decompositions discussed in Chapter 4. For expository purposes, Chapter 12 focuses on inference about structural impulse response estimates. It is understood that the discussion (with some exceptions noted in the chapter) also generalizes to related statistics such as forecast error variance decompositions. The chapter starts with a review of impulse response confidence intervals based on the delta method in stationary VAR models, followed by a detailed review of how to generate bootstrap approximations to the sampling distribution of VAR estimators and of how to use that information for the construction of bootstrap confidence intervals for structural impulse responses.

Next we discuss potential limitations of both delta method and bootstrap approaches in the VAR model and in the VECM context. The chapter highlights the special problems that arise in the possible presence of unit roots and cointegration. We stress that pretesting for unit roots and cointegration undermines the validity of inference in structural VAR and VEC models. We first discuss how standard methods of inference designed for stationary models can be made robust to the possible presence of unit roots and cointegration, at least asymptotically. We then review the use of local-to-unity asymptotics as an alternative asymptotic approximation for the VAR model in levels when the data are persistent. We also discuss nonstandard bootstrap methods designed for local-to-unity processes and other methods designed to improve inference

on structural impulse responses in that framework. Finally, we explain why Bayesian methods of inference are not immune to the possible presence of stochastic trends.

Standard methods of inference for structural impulse responses and related statistics are pointwise. From an economic point of view we are often interested not so much in the value of a given response function at a given horizon as in the overall shape and pattern of sets of structural impulse response functions. Assessing these features requires the user to do joint inference. Chapter 12 reviews recently proposed frequentist and Bayesian methods of conducting joint inference about structural impulse responses and illustrates the importance of these methods by example.

We also briefly explain how the implementation of the bootstrap approach depends on whether we are interested in constructing bootstrap confidence intervals, in conducting predictive inference, or in constructing bootstrap critical values for a test statistic. The chapter concludes with a range of empirical examples that illustrate the implementation of commonly used methods of constructing confidence intervals for structural impulse responses.

Chapter 13

Although there are situations in which exclusion restrictions may be motivated by economic theory or institutional features, it is rare for economists to be able to make a strong case for such restrictions. In many situations, economic theory only speaks to the sign of structural impulse responses on impact. For example, in a typical bivariate model of demand and supply, we expect a negative supply shock to lower quantity and to raise the price, whereas a positive demand shock would raise both quantity and the price. In other words, economic theory has implications for the sign of the impact responses. Only if the short-run supply curve were vertical would imposing a zero restriction on the contemporaneous effect of a demand shock on the quantity make economic sense.

This observation has motivated the idea of identifying structural VAR shocks based on sign restrictions. Because sign restrictions are inequality restrictions, the resulting structural models are no longer exactly identified, but only set identified. Put differently, even with an infinite amount of data, we can only narrow down each structural parameter estimate to a range of values rather than to a point in the parameter space. This property also is shared by the implied structural impulse responses and related statistics. As a result, the methods of inference discussed in Chapter 12 cannot be used for sign-identified models.

Chapter 13 examines in detail methods for approximating the set of identified structural impulse responses. Such approximations are typically constructed using Bayesian methods. We show how sign restrictions on impact responses (“static sign restrictions”) may be complemented by sign restrictions

on responses at longer horizons (“dynamic sign restrictions”) and by inequality restrictions on linear and nonlinear transformations of structural impulse responses. The chapter reviews several methods designed to evaluate the posterior distribution of set-identified impulse responses with special attention to the role of priors. We highlight the challenges in summarizing the posterior information from sign-identified models, discuss the difficulties in interpreting so-called median response functions, and demonstrate how this problem may be overcome in practice.

A central concern with sign-identified VAR models is that the prior remains asymptotically informative about the structural impulse responses. We discuss alternative frequentist and Bayesian methods recently proposed in the literature designed to make the role of the priors more explicit, to attenuate the role of priors, or to inject more economic information into the priors. The chapter concludes with a range of empirical examples and additional discussion on how to combine sign restrictions and more conventional short-run and long-run restrictions within the same structural VAR model.

Chapter 14

As we stressed in earlier chapters, the identification of structural shocks typically relies on economically motivated restrictions that are imposed on the data. An alternative strand of the literature exploits properties of the data for identification. In particular, changes in the conditional or unconditional volatility of the VAR errors may be used for identification. Chapter 14 discusses these approaches in some detail. We stress that this data-based approach does not provide any guidance as to the economic interpretation of the resulting “structural” shocks, however. It only achieves statistical identification in that it produces a unique set of mutually uncorrelated shocks. Thus, it is not a genuine alternative to the approaches discussed in earlier chapters. It is nevertheless attractive because it allows us to formally test traditional economically motivated exclusion restrictions by means of a test of overidentifying restrictions. Chapter 14 examines this approach in a variety of contexts including models with extraneous volatility changes, with Markov switching in the variances, with smooth transitions in the variances, and with generalized autoregressive conditionally heteroskedastic (GARCH) errors. We note that in some cases, this approach supports conventional identifying restrictions, whereas in others these restrictions can be rejected.

Chapter 14 also discusses an alternative data-based approach to identification that exploits the non-Gaussianity of the VAR errors in many applications. In this case, statistical identification may be achieved by insisting that the structural shocks be stochastically independent rather than just uncorrelated. In the Gaussian model, in contrast, uncorrelatedness implies independence. As in the heteroskedastic model, this approach provides a means for testing

more conventional identifying restrictions, if we are willing to postulate independence.

Chapter 15

Yet another option for identifying structural VAR shocks is to rely on additional extraneous data not already included among the VAR variables. Chapter 15 discusses two such approaches. The first approach relies on high-frequency interest rate futures prices. The change in these prices around the time of a monetary policy shift, suitably scaled to monthly frequency, is interpreted as the policy shock. Responses of the variable of interest to these extraneous shocks may be estimated outside of the VAR model and later imposed when estimating the structural VAR model.

The second approach uses directly observed measures of exogenous shocks as instruments in identifying exogenous variation in the VAR variable of interest. This approach allows us to use the exogenous shock measures discussed in Chapter 7 in the VAR context and is not limited to modeling monetary policy. For example, direct measures of exogenous fiscal policy shocks may be used as instruments to isolate the exogenous variation in fiscal variables.

Chapter 16

Typical VAR models include only a comparatively small number of variables. In recent years, public access to time series data has improved to the point that large panels of time series data are now available, often including dozens or hundreds of variables. A reasonable presumption is that all of these variables potentially include information relevant to forming expectations and to identifying structural shocks. Clearly, such large-dimensional models cannot be estimated without imposing additional structure in estimation, however.

Two popular approaches to this problem have been structural factor-augmented VAR (FAVAR) models and large-scale Bayesian structural VAR models. FAVAR models reduce the dimensionality of the estimation problem by imposing an approximate factor structure on the data, allowing us to approximate any variable as a linear combination of the most important factors contained in the panel of data. The advantage of this approach is that the set of factors is of much lower dimension than the original set of model variables. Its disadvantage is that it is not clear that the economically relevant structural shocks can be identified based on factors or linear combinations of factors. Large Bayesian structural VAR models, in contrast, solve the problem of dimensionality by imposing priors on the estimation problem. An obvious concern with the latter approach is that it is difficult to know how influential that prior is, given that we are not able to estimate the model without the prior.

One reason why such models are used in applied work is that they allow us to trace the responses of a larger set of variables to a given structural shock than would be possible using a conventional VAR model, providing an alternative to traditional DSEMs which had the same ability but at the cost of imposing stronger dynamic restrictions. Another reason is that such models greatly increase the information set used in measuring structural shocks, helping us address concerns about the informational structure of many small-scale structural VAR models. Chapter 16 reviews the derivation, specification, identification, and estimation of these models, highlighting potential problems with each approach. We also briefly discuss related approaches such as panel VAR models and global VAR (GVAR) models.

Chapter 17

Structural VAR analysis is based on the premise that the structural VAR shocks can be recovered from the reduced-form prediction errors. The corresponding reduced-form MA representation is called a fundamental representation. If the shocks in the reduced-form MA representation of the DGP are not the VAR prediction errors, in contrast, the reduced-form MA representation of the VAR model is nonfundamental, and it will in general not be possible to recover the true structural shocks even asymptotically. Such a situation arises when the econometrician's model does not have all the information that economic agents in the real world use in forming expectations. In other words, the reduced-form VAR model is informationally deficient. An important special case of this situation would be a model in which agents have forward-looking expectations that cannot be captured based on the information set of the VAR model.

Most commonly, nonfundamental representations are associated with an omitted-variable problem. The obvious response is to recognize that the root of this problem is the omission of relevant variables and to extend the information set, if possible. Note that the use of large-scale VAR models as discussed in Chapter 16 does not necessarily solve this problem because some of the variables relevant to agents' expectations may simply not be contained in any database. For example, oil market participants may anticipate rising oil prices because of fears about ethnic unrest in the Middle East, yet no database includes time series capturing the determinants of these fears. There are creative solutions to this type of problem in specific cases, however, some of which are discussed in Chapter 17.

Chapter 18

The standard VAR model is linear. In some cases, we may wish to allow the model variables to depend nonlinearly on past observations of the model variables rather than just linearly. Such models are collectively referred to as

nonlinear VAR models. Examples of nonlinear dynamics include models with smoothly evolving time-varying coefficients and models with coefficients that change with the state of the economy. Nonlinear VAR models allow economists to model target zones, stochastically switching regimes in the economy, gradual transitions to new economic regimes, thresholds induced by transaction costs, asymmetries in the responses of model variables to positive and negative shocks, and many other economically relevant phenomena.

An important difference compared with linear VAR models is that nonlinear structural impulse responses depend on the history of the data prior to the time period in which a structural shock occurs as well as on the magnitude and sign of this structural shock. This means that the structural impulse responses must be evaluated by numerical methods. Although the effect of alternative histories may be integrated out to arrive at an unconditional impulse response function, the nature of the structural shock remains important even in that case. Put differently, there is no unique set of structural impulse responses in nonlinear models, but rather a family of responses indexed by the magnitude and sign of the structural shock.

A common simplifying assumption in the literature has been that the nonlinear model is linear conditional on past values of the model variables, allowing the use of standard short-run exclusion restrictions. The use of sign restrictions is feasible as well, although it is not clear how to summarize the posterior distribution of the set of structural impulse responses in that case. In contrast, the use of long-run restrictions is not straightforward. The reason is that the closed-form solutions for the construction of structural impulse responses in linear models are not valid for nonlinear models. Rather, structural impulse responses must be constructed by Monte Carlo integration. It is not clear how to impose long-run restrictions on these numerical estimates. Neither do standard methods of constructing forecast error variance decompositions or historical decompositions apply in nonlinear VAR models.

Chapter 18 discusses the specification of nonlinear VAR models, their estimation, the identification of structural shocks, and inference about statistics such as structural impulse responses. We illustrate how existing methods for VAR models must be adapted in the nonlinear context. We also discuss alternative proposals for constructing impulse responses in nonlinear models such as the generalized impulse response function (GIRF). In addition, the chapter reviews related nonlinear models in the literature, such as nonparametric VAR models that allow for more flexible functional forms and noncausal VAR models that have been used to model nonfundamental representations of the type discussed in Chapter 17.

Finally, we examine models that are linear in the parameters yet imply nonlinear impulse response functions. A case in point is recently proposed structural models that allow for asymmetries in the response to positive and negative shocks, even in the impact period. For example, it is widely thought

that unexpected declines in the price of oil cause the gasoline price to fall less quickly than an unexpected increase in the price of oil of the same magnitude would cause the price of gasoline to increase. Using a common analogy, in the former case, gasoline prices fall slowly like a feather; in the latter, they shoot up like a rocket. We stress that the precise specification of the model matters when modeling asymmetries. For example, widely used censored oil price VAR models are invalid.

Chapter 19

The last chapter discusses practical issues related to trends, seasonality, and structural change. We review alternative more flexible trend models such as the HP filter and band-pass filters. We discuss how to combine variables with different trend specifications within the same model. We summarize in some detail the options for modeling seasonality in VAR models, and we discuss the implications of structural breaks for the specification of VAR models.

2 Vector Autoregressive Models

Structural VAR analysis is based on the premise that the DGP is well approximated by a reduced-form VAR model. In applied work, it is therefore important to choose a suitable VAR specification, taking account of the properties of the data. This chapter is devoted to the question of how to specify and estimate reduced-form VAR models. In Section 2.1 stochastic and deterministic trends in the data are discussed. Section 2.2 outlines the basic linear VAR model and its properties. Section 2.3 examines the estimation of reduced-form VAR models. Section 2.4 discusses how to generate predictions from VAR models, and Section 2.5 introduces the concept of Granger causality. Lag-order selection and model diagnostics are discussed in Sections 2.6 and 2.7. Section 2.8 briefly reviews three classes of restricted reduced-form VAR models.

Given that the linear VAR model is one of the standard tools for empirical research in macroeconomics and finance, there are many previous good expositions of the topics covered in this chapter. Our discussion draws heavily on the material in Lütkepohl (2005, 2006, 2009, 2013).

2.1 Stationary and Trending Processes

We call a stochastic process covariance stationary or simply stationary if it has time invariant first and second moments. Similarly, an economic variable is referred to as covariance stationary if the underlying DGP is covariance stationary. More formally, the scalar process y_t , $t \in \mathbb{N}$ or $t \in \mathbb{Z}$, is covariance stationary if

$$\mathbb{E}(y_t) = \mu \quad \text{and} \quad \text{Cov}(y_t, y_{t+h}) = \gamma_h, \quad \forall t, h.$$

Note that μ and γ_h are constants that do not depend on t . This property is also known as second-order stationarity. If the joint distribution of y_t, \dots, y_{t+h} is time invariant, the process y_t is strictly stationary.

In practice, an economic variable being stationary is the exception rather than the rule. For example, often the raw data have to be transformed prior

to the analysis by taking natural logs to stabilize the variance of the variable. In addition, there are many variables that have trends that have to be removed or modeled explicitly to ensure stationarity. A trend in a time series variable is thought of as a systematic upward or downward movement over time. For example, a variable y_t may vary about a linear trend line of the form $y_t = \mu_0 + \mu_1 t + x_t$, where x_t is a zero mean stationary stochastic process. The straight line, $\mu_0 + \mu_1 t$, represents a simple deterministic trend function that captures the systematic upward or downward movement of many economic variables reasonably well.

Alternatively, a variable may be viewed as being driven by a stochastic trend. A simple example of a process with a stochastic trend is the univariate AR(1) process

$$y_t = ay_{t-1} + u_t$$

with coefficient $a = 1$ such that

$$y_t = y_{t-1} + u_t.$$

This process is called a random walk. Its AR polynomial has a unit root, i.e.,

$$1 - az = 0 \quad \text{for } z = 1.$$

Its stochastic error u_t (also known as the innovation) is assumed to be a white noise process with mean 0 and variance σ_u^2 . In other words, u_t and u_s are uncorrelated for $s \neq t$, $\mathbb{E}(u_t) = 0$, and $\mathbb{E}(u_t^2) = \sigma_u^2$. Given that $y_t - y_{t-1} = u_t$, it is easily seen that the effect of a random change in u_t on future values of y_t is not reversed in expectation. Thus, the effect of u_t on future values of y_t is permanent.

Successive substitution for lagged y_t variables in the defining equation of the random walk, $y_t = y_{t-1} + u_t$, yields

$$y_t = y_0 + \sum_{i=1}^t u_i. \tag{2.1.1}$$

Hence, assuming that the process is defined for $t \in \mathbb{N}$, we have

$$\mathbb{E}(y_t) = \mathbb{E}(y_0) \quad \text{and} \quad \text{Var}(y_t) = t\sigma_u^2 + \text{Var}(y_0).$$

In other words, even though $\text{Var}(y_0)$ is finite, the variance of a random walk tends to infinity. Moreover, the correlation

$$\text{Corr}(y_t, y_{t+h}) = \frac{\mathbb{E}\left[\left(\sum_{i=1}^t u_i\right)\left(\sum_{i=1}^{t+h} u_i\right)\right]}{[t\sigma_u^2(t+h)\sigma_u^2]^{1/2}} = \frac{t}{(t^2 + th)^{1/2}} \xrightarrow[t \rightarrow \infty]{} 1 \tag{2.1.2}$$

for any given integer h . Due to this property, even random variables y_t and y_s of the process far apart in time (such that s is much greater than t) are strongly correlated. This property indicates a strong persistence in the time series process. In fact, it turns out that the expected time between two crossings of zero is infinite. Such behaviour is associated with a trend in the data. Clearly, since u_t is stochastic, so is the trend.

A univariate AR(1) process with unit coefficient and a constant term,

$$y_t = \nu + y_{t-1} + u_t,$$

is called a random walk with drift. Successive substitution of lags of y_t shows that in this case

$$y_t = y_0 + t\nu + \sum_{i=1}^t u_i$$

and, hence, the process has a linear trend in the mean:

$$\mathbb{E}(y_t) = \mathbb{E}(y_0) + t\nu.$$

Higher-order AR processes such as

$$y_t = \nu + a_1 y_{t-1} + \cdots + a_p y_{t-p} + u_t,$$

where u_t is white noise as before, have stochastic trending properties similar to random walks if the AR polynomial $1 - a_1 z - \cdots - a_p z^p$ has a root for $z = 1$. The AR polynomial can be decomposed as

$$1 - a_1 z - \cdots - a_p z^p = (1 - \lambda_1 z) \times \cdots \times (1 - \lambda_p z), \quad (2.1.3)$$

where $\lambda_1, \dots, \lambda_p$ are the reciprocals of the roots of the polynomial. If the process has only one unit root or, equivalently, only one of the λ_i roots is 1 and all the others are smaller than 1, the process behaves similarly to a random walk in that it follows a stochastic trend. More precisely, y_t can be decomposed into a random walk and a stationary component such that y_t varies about a stochastic trend generated by its random walk component.

The representation of the AR polynomial shows that the unit root can be removed by taking first differences of the process. Let $\Delta y_t \equiv (1 - L)y_t \equiv y_t - y_{t-1}$, where L is the lag operator such that $Ly_t \equiv y_{t-1}$, and Δ is the difference operator such that $\Delta \equiv 1 - L$ and hence $\Delta y_t = y_t - y_{t-1}$.

An AR(p) process with AR polynomial satisfying the condition

$$1 - a_1 z - \cdots - a_p z^p \neq 0 \quad \forall z \in \mathbb{C}, |z| \leq 1, \quad (2.1.4)$$

is called stable. Here $|z|$ denotes the modulus of the complex number z . Put differently, $|z|$ is the distance from the origin of the complex plane. If, in addition, the mean of the AR process does not change over time deterministically, as would be the case in the presence of a deterministic time trend, if the error

term u_t has time-invariant variance σ_u^2 , and if its first and second moments are bounded, then the AR process is stationary. Sometimes in the literature, condition (2.1.4) is rather imprecisely viewed as a condition ensuring stationarity. Of course, interpreting (2.1.4) as a stationarity condition implicitly assumes that there are no other deviations from stationarity such as a linear deterministic trend in the mean or an innovation variance changing over time.

To ensure finite moments, AR processes with unit roots are assumed to start at some fixed time period, say t_0 , if not explicitly stated otherwise. For example, in the foregoing discussion we have assumed that $t_0 = 0$. In contrast, stable AR processes without unit roots are typically assumed to have started in the infinite past to ensure stationarity. Without that assumption they may only be asymptotically stationary in that the moments are not time-invariant, but converge to their limit values only for $t \rightarrow \infty$.

If the AR polynomial has $d \in \mathbb{N}$ unit roots and, hence, d of the λ_i roots in (2.1.3) are equal to 1, the process is called integrated of order d ($I(d)$). In that case, the process can be made stable by differencing it d times. For example, if $d = 1$, $\Delta y_t = y_t - y_{t-1}$ is stable. If $d = 2$, $\Delta^2 y_t = (1 - L)^2 y_t = y_t - 2y_{t-1} + y_{t-2}$ is stable, and so forth. If $d = 2$, the original y_t must be differenced twice. For example, if the log price level p_t is $I(2)$, then the inflation rate $\pi_t = \Delta p_t = p_t - p_{t-1}$ is $I(1)$, and the change in the inflation rate $\Delta \pi_t = \pi_t - \pi_{t-1} = p_t - p_{t-1} - (p_{t-1} - p_{t-2}) = p_t - 2p_{t-1} + p_{t-2}$ is $I(0)$. As before, initial values can be chosen such that $\Delta^d y_t = (1 - L)^d y_t$ is stationary, provided the conditions for the mean and for the innovation variance required for stationarity are satisfied.

Stable, stationary processes are referred to as $I(0)$ processes. Generally, for $d \in \mathbb{N}$, a stochastic process y_t is called $I(d)$, if $\Delta^d y_t \equiv z_t$ is a stationary process with infinite-order moving average (MA) representation, $z_t = \sum_{j=0}^{\infty} \theta_j u_{t-j} = \theta(L)u_t$, where the MA coefficients satisfy the condition $\sum_{j=0}^{\infty} j|\theta_j| < \infty$, $\theta(1) = \sum_{j=0}^{\infty} \theta_j \neq 0$, and $u_t \sim (0, \sigma_u^2)$ is white noise. For example, in the case of an $I(1)$ process, this condition implies that $y_t = y_{t-1} + z_t$ has the representation

$$\begin{aligned} y_t &= y_0 + z_1 + \cdots + z_t \\ &= y_0 + \theta(1)(u_1 + \cdots + u_t) + \sum_{j=0}^{\infty} \theta_j^* u_{t-j} - z_0^*, \end{aligned} \tag{2.1.5}$$

where $\theta_j^* = -\sum_{i=j+1}^{\infty} \theta_i$, $j = 0, 1, \dots$, and $z_0^* = \sum_{j=0}^{\infty} \theta_j^* u_{t-j}$ contains initial values. The variable y_t is decomposed into the sum of a random walk, $\theta(1)(u_1 + \cdots + u_t)$, a stationary process, $\sum_{j=0}^{\infty} \theta_j^* u_{t-j}$, and initial values, $y_0 - z_0^*$. The decomposition (2.1.5) is known as the Beveridge-Nelson decomposition (see Beveridge and Nelson 1981).

Of course, our primary interest is in systems of variables. Hence, it is useful to extend the $I(d)$ terminology to that setting as well. Accordingly, we call a vector process $y_t = (y_{1t}, \dots, y_{Kt})'$ $I(d)$ if stochastic trends can be removed by

differencing y_t d times and if differencing $d - 1$ times is not enough for trend removal. It is important to note, however, that in systems of variables even if only one of the variables is $I(d)$ individually, the whole system is viewed as $I(d)$. Moreover, it is possible that a single stochastic trend drives several of the variables jointly. This is the important case of cointegrated variables to be discussed in Chapter 3.

The $I(d)$ terminology has also been extended to non-integer, real numbers d . For general $d \in \mathbb{R}$ the so-called fractional differencing operator Δ^d is defined as a binomial expansion,

$$\begin{aligned}\Delta^d = (1 - L)^d &= 1 - dL - \frac{d(1-d)}{2}L^2 - \frac{d(1-d)(2-d)}{6}L^3 - \dots \\ &= \sum_{i=0}^{\infty} (-1)^i \binom{d}{i} L^i \\ &= \sum_{i=0}^{\infty} (-1)^i \frac{d(d-1) \times \dots \times (d-i+1)}{1 \times 2 \times \dots \times i} L^i.\end{aligned}$$

The infinite sum reduces to a finite sum for $d \in \mathbb{N}$. The process y_t is called fractional or fractionally integrated of order d if $\Delta^d y_t = z_t$ is $I(0)$ with MA representation $z_t = \theta(L)u_t$, $\theta(1) \neq 0$ (see, e.g., Johansen and Nielsen 2012). Such processes were introduced to the time series econometrics literature by Granger and Joyeux (1980) and Hosking (1981b). Fractionally integrated processes are often referred to as long-memory processes because for $d > 0$ they are more persistent and their autocorrelations taper off to zero more slowly than for $I(0)$ processes. Although fractionally integrated processes are not $I(0)$, they may be stationary. Stationarity of a fractionally integrated process requires $|d| < 0.5$.

Integer-valued differences often have a natural interpretation. For example, first differences of the logs of a variable represent growth rates. Such an easy interpretation is lost for fractionally differenced variables. Thus, it is perhaps not surprising that the concept of fractional integration to date has not been used much in structural VAR analysis. More importantly, reliable estimation of fractionally integrated processes requires larger samples than typically available in macroeconomics. Fractional processes therefore do not play an important role in this volume. In the remainder of this book, when we refer to $I(d)$ variables, we always mean non-negative integers d unless explicitly stated otherwise.

2.2 Linear VAR Processes

2.2.1 The Basic Model

Suppose that the relationship between a set of K time series variables, $y_t = (y_{1t}, \dots, y_{Kt})'$, is of interest and that the DGP can be represented as the sum of

a deterministic part μ_t and a purely stochastic part x_t with mean zero such that

$$y_t = \mu_t + x_t. \quad (2.2.1)$$

In other words, the expected value of y_t is $\mathbb{E}(y_t) = \mu_t$. The deterministic term may contain a constant, polynomial trend terms, deterministic seasonal terms, and other dummy variables. For simplicity, μ_t is usually assumed to contain only a constant such that $\mu_t = \mu_0$. Occasionally a linear trend of the form $\mu_t = \mu_0 + \mu_1 t$ is considered. Generally the additive setup (2.2.1) makes it necessary to think about the deterministic terms at the beginning of the analysis and to allow for the appropriate polynomial order. In some applications trend adjustments are performed prior to a VAR analysis. This approach must be taken, for example, when the detrending procedure cannot be incorporated into the VAR specification. An example is the use of HP-filtered data. Further discussion of these alternative detrending methods can be found in Chapter 19. In that case there may be no deterministic term in the VAR representation in levels, i.e., $\mu_t = 0$ in expression (2.2.1) and $y_t = x_t$.

The purely stochastic part, x_t , of the DGP is assumed to follow a linear VAR process of order p (referred to as a VAR(p) model) of the form

$$x_t = A_1 x_{t-1} + \cdots + A_p x_{t-p} + u_t, \quad (2.2.2)$$

where the A_i , $i = 1, \dots, p$, are $K \times K$ parameter matrices and the error process $u_t = (u_{1t}, \dots, u_{Kt})'$ is a K -dimensional zero mean white noise process with covariance matrix $\mathbb{E}(u_t u_t') = \Sigma_u$ such that $u_t \sim (0, \Sigma_u)$. The white noise assumption rules out serial correlation in the errors but allows for conditional variance dynamics such as generalized autoregressive conditionally heteroskedastic (GARCH) errors (see e.g. Chapter 14). Sometimes it is useful to strengthen this assumption, for example, by postulating independent and identically distributed (iid) errors or by postulating that u_t is a martingale difference sequence.¹

Expression (2.2.2) defines a system of equations. Each model variable in y_t is regressed on its own lags as well as lags of the other model variables up to a lag order p (see Chapter 1). To economize on notation, it is convenient to define the matrix polynomial in the lag operator $A(L) = I_K - A_1 L - \cdots - A_p L^p$ and write the process (2.2.2) as

$$A(L)x_t = u_t. \quad (2.2.3)$$

The observed variables y_t inherit the VAR structure of x_t . This can be seen easily by pre-multiplying (2.2.1) by $A(L)$ and considering $A(L)y_t = A(L)\mu_t +$

¹ The stochastic process v_t is called a martingale sequence if $\mathbb{E}(v_t | v_{t-1}, v_{t-2}, \dots) = v_{t-1} \forall t$. Then $u_t \equiv \Delta v_t$ is called a martingale difference if it has expectation $\mathbb{E}(u_t | v_{t-1}, v_{t-2}, \dots) = 0 \forall t$. Unlike an iid white noise process, a white noise process that is a martingale difference sequence allows for conditional heteroskedasticity.

u_t . For instance, if the deterministic term is just a constant, i.e., $\mu_t = \mu_0$, then

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad (2.2.4)$$

where $v = A(L)\mu_0 = A(1)\mu_0 = (I_K - \sum_{j=1}^p A_j)\mu_0$. In the terminology of the literature on simultaneous equations models, model (2.2.4) is a reduced form because all right-hand side variables are lagged and hence predetermined.

The VAR process x_t and, hence, y_t is stable if all roots of the determinantal polynomial of the VAR operator are outside the complex unit circle, i.e.,

$$\det(A(z)) = \det(I_K - A_1 z - \cdots - A_p z^p) \neq 0 \quad \forall z \in \mathbb{C}, |z| \leq 1, \quad (2.2.5)$$

where \mathbb{C} denotes the set of complex numbers. Under common assumptions such as a constant mean and white noise innovations with time-invariant covariance matrix, a stable VAR process has time-invariant means, variances, and covariance structure and hence is stationary, as will be seen in the next subsection. Thus, condition (2.2.5) generalizes the stability condition (2.1.4) to the multivariate case.

For later reference we note that the K -dimensional VAR(p) process (2.2.4) can be written as a pK -dimensional VAR(1) process by stacking p consecutive y_t variables in a pK -dimensional vector, $Y_t = (y'_t, \dots, y'_{t-p+1})'$, and noting that

$$Y_t = v + \mathbf{A} Y_{t-1} + U_t, \quad (2.2.6)$$

where

$$v \equiv \begin{bmatrix} v \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{Kp \times 1}, \quad \mathbf{A} \equiv \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_K & 0 & \cdots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{bmatrix}_{Kp \times Kp}, \quad \text{and} \quad U_t \equiv \begin{bmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{Kp \times 1}.$$

The matrix \mathbf{A} is referred to as the companion matrix of the VAR(p) process. Using the stability condition (2.2.5), Y_t is stable if

$$\det(I_{Kp} - \mathbf{A}z) \neq 0 \quad \forall z \in \mathbb{C}, |z| \leq 1, \quad (2.2.7)$$

which, of course, is equivalent to condition (2.2.5). It is easy to see that this condition is equivalent to all eigenvalues of \mathbf{A} having modulus less than 1, which provides a convenient tool for assessing the stability of a VAR model and for computing the autoregressive roots. By construction, the eigenvalues of \mathbf{A} are the reciprocals of the roots of the VAR lag polynomial (2.2.5).

2.2.2 The Moving Average Representation

A stable VAR(p) process y_t can be represented as the weighted sum of past and present innovations. This is easily seen for a VAR(1) process,

$$y_t = v + A_1 y_{t-1} + u_t.$$

Successive substitution implies

$$y_t = \sum_{i=0}^{\infty} A_1^i v + \sum_{i=0}^{\infty} A_1^i u_{t-i} = (I_K - A_1)^{-1} v + \sum_{i=0}^{\infty} A_1^i u_{t-i}.$$

The sum on the right-hand side of this infinite-order representation exists if the eigenvalues of A_1 are all less than 1 in modulus. Similarly, a representation in terms of past and present innovations of a VAR(p) model can be obtained via the corresponding VAR(1) representation, resulting in

$$\begin{aligned} y_t &= A(L)^{-1} v + A(L)^{-1} u_t \\ &= A(1)^{-1} v + \sum_{i=0}^{\infty} J \mathbf{A}^i J' J U_{t-i} \\ &= \mu + \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \end{aligned} \tag{2.2.8}$$

where $J \equiv [I_K, 0_{K \times K(p-1)}]$ is a $K \times Kp$ matrix, $\mu = A(1)^{-1} v$ and the $K \times K$ coefficient matrices of the inverse VAR operator $A(L)^{-1} = \sum_{i=0}^{\infty} \Phi_i L^i$ are equal to $\Phi_i = J \mathbf{A}^i J'$, $i = 0, 1, \dots$. These matrices can also be obtained recursively as

$$\Phi_0 = I_K, \quad \text{and} \quad \Phi_i = \sum_{j=1}^i \Phi_{i-j} A_j, \quad i = 1, 2, \dots,$$

with $A_j = 0$ for $j > p$ (see Lütkepohl 2005, chapter 2).

The existence of the inverse VAR operator is ensured by the stability of the process. The representation (2.2.8) is known as the moving average (MA) representation or more precisely the Wold MA representation or the prediction error MA representation. This qualifier is important because there are infinitely many MA representations of y_t . In fact, any nonsingular linear transformation of the white noise process u_t , say $v_t = Qu_t$, gives rise to a white noise process and can be used for an MA representation of y_t ,

$$y_t = \mu + \sum_{i=0}^{\infty} \Theta_i v_{t-i}, \tag{2.2.9}$$

with $\Theta_i = \Phi_i Q^{-1}$, $i = 0, 1, \dots$. A distinguishing feature of the Wold MA representation is that the weighting matrix Φ_0 of the unlagged error term

is the identity matrix, while Θ_0 is not an identity matrix for nontrivial transformations.

It follows immediately from the Wold MA representation that

$$\mathbb{E}(y_t) = \mu$$

and that

$$\Gamma_y(h) \equiv \text{Cov}(y_t, y_{t-h}) = \mathbb{E}[(y_t - \mu)(y_{t-h} - \mu)'] = \sum_{i=0}^{\infty} \Phi_{h+i} \Sigma_u \Phi_i'. \quad (2.2.10)$$

Hence, the first and second moments of this VAR process are time invariant and the process is stationary (see Lütkepohl 2005, chapter 2).

2.2.3 VAR Models as an Approximation to VARMA Processes

An important result in this context is due to Wold (1938) who showed that every K -dimensional nondeterministic zero mean stationary process y_t has an MA representation

$$y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i}, \quad (2.2.11)$$

where $\Phi_0 = I_K$. This result follows from the Wold Decomposition Theorem and motivates the terminology used for the MA representation (2.2.8). This result is important because it illustrates the generality of the VAR model. Suppose the Φ_i are absolutely summable and that there exists an operator $A(L)$ with absolutely summable coefficient matrices satisfying $A(L)\Phi(L) = I_K$. Then $\Phi(L)$ is invertible [$A(L) = \Phi(L)^{-1}$] and y_t has a VAR representation of possibly infinite order that can be approximated arbitrarily well by a finite-order VAR(p) if p is sufficiently large.

In particular, under suitable conditions, a VAR(p) process may be used to approximate time series generated from vector autoregressive moving average (VARMA) models of the form

$$y_t = v + A_1 y_{t-1} + \cdots + A_{p_0} y_{t-p_0} + u_t + M_1 u_{t-1} + \cdots + M_{q_0} u_{t-q_0},$$

where p_0 and q_0 denote the true autoregressive and moving average lag orders, provided the VAR lag order p is sufficiently large. If the VAR operator $A(z) = I_K - A_1 z - \cdots - A_{p_0} z^{p_0}$ of the VARMA process satisfies the stability condition (2.2.5) and, thus, the VAR operator has no roots in or on the complex unit circle, the VARMA process has a possibly infinite-order MA representation (2.2.11).

Moreover, if the determinant of the MA operator of the VARMA process has all its roots outside the unit circle, i.e.,

$$\det(M(z)) = \det(I_K + M_1 z + \cdots + M_{q_0} z^{q_0}) \neq 0 \quad \forall z \in \mathbb{C}, |z| \leq 1,$$

the process also has an equivalent pure VAR representation of possibly infinite order.² Unlike in the univariate case, the inverse of a finite-order operator may also be a finite-order operator in the multivariate case. In other words, $M(z)^{-1}$ may be a finite order operator if $M(z)$ has finite order. Hence, it is possible in the multivariate case that a finite-order MA process has an equivalent finite-order VAR representation and vice versa.

A detailed introductory exposition of VARMA processes is provided by Lütkepohl (2005), and a more advanced treatment can be found in Hannan and Deistler (1988). Since VARMA processes are much more difficult to deal with in practice, we focus on VAR models in the remainder of this book.

If the VAR process of interest has a unit root and, hence, the stability condition is not satisfied, the infinite-order MA representation (2.2.8) does not exist. However, we can still think of the process as starting from $Y_0 = (y'_0, \dots, y'_{-p+1})'$ and obtain a representation

$$y_t = \mu_t + \sum_{i=0}^{t-1} \Phi_i u_{t-i} + J\mathbf{A}^i Y_0$$

by successive substitution. For some purposes this representation is useful, but not for all. In particular, it obscures the long-run properties of the process. These are more easily understood using the so-called Granger representation discussed in Chapter 3.

2.2.4 Marginal Processes, Measurement Errors, Aggregation, Variable Transformations

The reduced-form MA representation is also a good point of departure for studying the implications of dropping variables from a VAR process. Consider a bivariate stationary process for two variables,

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} + \begin{bmatrix} \phi_{11}(L) & \phi_{12}(L) \\ \phi_{21}(L) & \phi_{22}(L) \end{bmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}. \quad (2.2.12)$$

Thus, the first variable has the representation

$$y_{1t} = \mu_1 + \phi_{11}(L)u_{1t} + \phi_{12}(L)u_{2t}$$

² An MA representation with MA operator satisfying this invertibility condition is sometimes called a fundamental MA representation. In Chapter 17 we discuss nonfundamental MA representations that have roots inside the complex unit circle and, hence, do not satisfy the invertibility condition for the MA operator.

in terms of both innovation series. According to Wold's decomposition theorem, it also has an MA representation in terms of a scalar white noise process, v_t :

$$y_{1t} = \mu_1 + \sum_{i=0}^{\infty} \psi_i v_{t-i}.$$

This MA represents the marginal process of y_{1t} that is obtained by integrating out the second variable. If $\phi_{12}(L) \neq 0$, then $v_t \neq u_{1t}$ and $\psi_i \neq \phi_{11,i}$ in general. These facts are important to keep in mind for the analysis of impulse responses in Chapter 4. The point to remember is that, in general, dropping some variables from a multivariate time series process results in a lower-dimensional process with possibly quite different MA coefficients than the process for the original set of variables.

More generally, any transformation of the variables implies changes in the MA coefficients. Consider, for example, a nonsingular transformation matrix F and a transformed process

$$z_t = Fy_t = F\mu + \sum_{i=0}^{\infty} F\Phi_i F^{-1} Fu_{t-i} = \mu_z + \sum_{i=0}^{\infty} \Psi_i v_{t-i}, \quad (2.2.13)$$

where $\Psi_i = F\Phi_i F^{-1}$ and $v_t = Fu_t$. Obviously, such transformations change the MA coefficient matrices and white noise error term, and therefore also affect the lag order of the approximating autoregressive process. The same result also holds when F is not a square matrix. Suppose F is an $M \times K$ matrix of rank M . Then one may add $K - M$ rows to the matrix such that it becomes nonsingular, consider the resulting nonsingular transformation, and finally omit the last $K - M$ components of the transformed vector.

In short, linear transformations of a VAR process have MA representations quite different from that of the original process, but both representations are equally valid. For example, a researcher may be working with a VAR for the interest rate, inflation rate, and real GDP growth. These variables could alternatively be represented as autoregressive-moving average (ARMA) processes for each series separately, which in turn can be approximated by finite-order AR models. Linear transformations are also quite common when aggregating data across households and industries to form macroeconomic aggregates. Likewise, problems of temporal aggregation fall within this framework (see Lütkepohl 2005, chapter 11). For example, it is common to aggregate monthly inflation to quarterly inflation data, which involves taking a linear combination of monthly inflation rates.

In this context it is important to stress that different types of variables require different temporal aggregation methods. For flow variables such as GDP or industrial production, temporal aggregation to a lower frequency

involves accumulating the high-frequency observations over time. For example, quarterly industrial production is obtained by summing the monthly industrial production that has taken place within each quarter. In contrast, stock variables such as the number of unemployed workers or the population of a region are aggregated from monthly data to quarterly frequency by using, for example, the last monthly value of each quarter as the quarterly value and dropping the other monthly observations. In other words, temporal aggregation is performed by what is known as skip-sampling or systematic sampling. Alternatively, one may use the average of the monthly values as a quarterly value, depending on the economic context. The important point to note here is that different temporal aggregation schemes imply different changes in the DGP and hence in the MA representation of the variables. There is an extensive literature discussing these issues, in particular in the forecasting context. Early contributions include Tiao (1972), Amemiya and Wu (1972), Brewer (1973), Abraham (1982), and Wei (1981). A more recent systematic account of this literature is provided in Lütkepohl (1987). An alternative approach has been to combine time series observed at different frequencies within the same econometric model (see Foroni, Ghysels, and Marcellino 2013). Related issues in the context of structural modeling are taken up in Chapter 15.

Another example of a linear aggregation problem is additive measurement error in the data. Suppose that the variable of interest, say y_t^* , is measured with error and denote the measurement error by m_t such that the observed process is $y_t = y_t^* + m_t$. In other words, y_t is a linear transformation of the joint process

$$\begin{pmatrix} y_t^* \\ m_t \end{pmatrix}.$$

Then, assuming that the joint process is stationary, the previous discussion implies that the MA representation of y_t differs from that of y_t^* .

These considerations demonstrate that even linear transformations can have a substantial impact on the MA representation of a stationary process. Although these issues do not invalidate the reduced-form representation of VAR models, they may affect the structural interpretation and identification of VAR models, as discussed in later chapters. Our discussion in this section has been based on the MA representation and, hence, applies to stationary processes more generally. We now return to finite-order VAR processes and discuss parameter estimation within that model class.

2.3 Estimation of VAR Models

VAR models can be estimated by standard methods. Unrestricted least-squares (LS), generalized least-squares (GLS), bias-corrected least-squares, and maximum likelihood (ML) methods are discussed in Sections 2.3.1–2.3.4. Our main focus in this chapter is on stationary VAR processes. The properties of LS and

ML methods when there are integrated variables in the VAR model in levels are briefly discussed in Section 2.3.5. A more detailed discussion of the estimation of integrated and cointegrated VAR processes can be found in Chapter 3. Bayesian estimation methods for VAR models are reviewed in Chapter 5.

2.3.1 Least-Squares Estimation

Consider the $\text{VAR}(p)$ model (2.2.4) written in the more compact form

$$y_t = [v, A_1, \dots, A_p]Z_{t-1} + u_t, \quad (2.3.1)$$

where $Z_{t-1} \equiv (1, y'_{t-1}, \dots, y'_{t-p})'$ and u_t is assumed to be iid white noise with nonsingular covariance matrix, Σ_u , such that $u_t \stackrel{iid}{\sim} (0, \Sigma_u)$. Further deterministic terms may be handled analogously. Given a sample of size T , y_1, \dots, y_T , and p presample vectors, y_{-p+1}, \dots, y_0 , ordinary LS for each equation separately results in efficient estimators. The LS estimator is

$$\hat{A} = [\hat{v}, \hat{A}_1, \dots, \hat{A}_p] = \left(\sum_{t=1}^T y_t Z'_{t-1} \right) \left(\sum_{t=1}^T Z_{t-1} Z'_{t-1} \right)^{-1} = YZ'(ZZ')^{-1}, \quad (2.3.2)$$

where $Y \equiv [y_1, \dots, y_T]$ is $K \times T$ and $Z \equiv [Z_0, \dots, Z_{T-1}]$ is $(Kp + 1) \times T$.

More precisely, stacking the columns of $A = [v, A_1, \dots, A_p]$ in the $(pK^2 + K) \times 1$ vector $\alpha = \text{vec}(A)$,

$$\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\hat{\alpha}}), \quad (2.3.3)$$

where $\Sigma_{\hat{\alpha}} = \text{plim}\left(\frac{1}{T}ZZ'\right)^{-1} \otimes \Sigma_u$, if the process is stable. Under fairly general assumptions, the LS estimator has an asymptotic normal distribution (see Mann and Wald 1943). A sufficient condition for the consistency and asymptotic normality of \hat{A} would be that u_t is a continuous iid random variable with four finite moments (see, e.g., Lütkepohl 2005, chapter 3). These assumptions may be relaxed to allow for conditional or unconditional heteroskedasticity, for example.

A consistent estimator of the innovation covariance matrix Σ_u under the assumption of iid innovations is

$$\widehat{\Sigma}_u = \frac{\widehat{U}\widehat{U}'}{T - Kp - 1}, \quad (2.3.4)$$

where $\widehat{U} = Y - \hat{A}Z$ are the LS residuals.

Thus, in large samples,

$$\text{vec}(\hat{A}) \xrightarrow{a} \mathcal{N}\left(\text{vec}(A), (ZZ')^{-1} \otimes \widehat{\Sigma}_u\right), \quad (2.3.5)$$

where \sim^a denotes the approximate large-sample distribution. In other words, asymptotically the usual t -statistics can be used for testing restrictions on individual coefficients and for setting up confidence intervals.

Moreover, if multiple restrictions are of interest, Wald tests can be used. Suppose, for example, that we want to test the pair of linear hypotheses

$$\mathbb{H}_0 : R\alpha = r \quad \text{versus} \quad \mathbb{H}_1 : R\alpha \neq r,$$

where R is a given $N \times (pK^2 + K)$ matrix of rank N and r is a given N -dimensional vector. Then, if \mathbb{H}_0 is true, the result (2.3.3) implies that the statistic

$$W = T(R\hat{\alpha} - r)'(R\hat{\Sigma}_{\hat{\alpha}}R')^{-1}(R\hat{\alpha} - r),$$

where under our assumptions $\hat{\Sigma}_{\hat{\alpha}} = (\frac{1}{T}ZZ')^{-1} \otimes \hat{\Sigma}_u$ is a consistent estimator of $\Sigma_{\hat{\alpha}}$, has an asymptotic χ^2 distribution with N degrees of freedom that can be used for testing \mathbb{H}_0 . This test is known as a Wald test.

In VAR analysis, nonlinear functions of the parameters are often of interest. Examples are the structural impulse responses and forecast error variance decompositions considered in Chapter 4. Suppose that interest focuses on the nonlinear function $\phi : \mathbb{R}^{pK^2+K} \rightarrow \mathbb{R}^N$ that maps A on the N -dimensional vector $\phi(A)$. Estimation of ϕ may be based on the LS estimator of A , denoted by $\hat{\phi} = \phi(\hat{A})$. Then, using result (2.3.3) and the delta method, it follows that

$$\sqrt{T}(\hat{\phi} - \phi) \xrightarrow{d} \mathcal{N}\left(0, \frac{\partial \phi}{\partial \alpha'} \Sigma_{\hat{\alpha}} \frac{\partial \phi'}{\partial \alpha}\right), \quad (2.3.6)$$

where $\partial \phi / \partial \alpha'$ is the $N \times (pK^2 + K)$ matrix of partial derivatives of the nonlinear function ϕ with respect to the elements of $\alpha \equiv \text{vec}(A)$. It is assumed that the matrix of partial derivatives is nonzero, when evaluated at the true parameter vector (see Serfling 1980). If the covariance matrix

$$\Sigma_{\hat{\phi}} = \frac{\partial \phi}{\partial \alpha'} \Sigma_{\hat{\alpha}} \frac{\partial \phi'}{\partial \alpha}$$

is nonsingular, result (2.3.6) can be used in the usual way for inference regarding ϕ . In other words, t -ratios may be used for significance tests of individual components of ϕ and Wald tests can be constructed for hypotheses related to more than one component of ϕ . For example, the null hypothesis $\mathbb{H}_0 : \phi(A) = \phi_0$ can be tested by using the statistic

$$W = T[\phi(\hat{A}) - \phi_0]' \widehat{\Sigma}_{\hat{\phi}}^{-1} [\phi(\hat{A}) - \phi_0]$$

that has an asymptotic χ^2 distribution if \mathbb{H}_0 is true. If $\Sigma_{\hat{\phi}}$ is singular, in contrast, Wald tests may not have the usual asymptotic χ^2 distribution anymore, whereas t -tests remain valid asymptotically, as long as all diagonal elements of $\Sigma_{\hat{\phi}}$ are nonzero. General results for the case of a singular covariance matrix

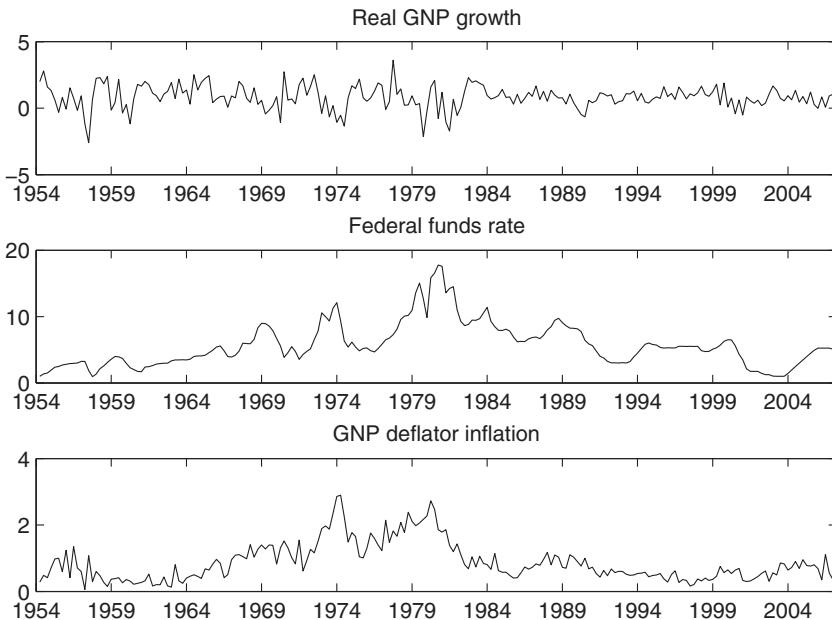


Figure 2.1. Quarterly U.S. data for real GNP growth, the federal funds rate, and the GNP deflator inflation for 1954q4–2007q4.

can be found in Andrews (1987). The issue of singularities is discussed in more detail in Chapter 12 in the context of the example of inference about structural impulse responses.

As an empirical illustration, consider a VAR(4) for $y_t = (\Delta gnp_t, i_t, \Delta p_t)'$, where gnp_t denotes the log of U.S. real GNP, p_t the corresponding GNP deflator in logs, and i_t the federal funds rate, averaged by quarter. The estimation period is restricted to 1954q4–2007q4. The data are shown in Figure 2.1. In this section we treat these data as $I(0)$.

The unrestricted LS estimates of the VAR(4) model with intercept are

$$\hat{\nu} = \begin{bmatrix} 0.7240 \\ -0.3925 \\ 0.0067 \end{bmatrix},$$

$$\hat{A}_1 = \begin{bmatrix} 0.2230 & 0.0097 & 0.3969 \\ 0.3147 & 1.0969 & 0.5979 \\ 0.0012 & 0.0636 & 0.4096 \end{bmatrix},$$

$$\hat{A}_2 = \begin{bmatrix} 0.2143 & -0.3862 & 0.1360 \\ 0.1867 & -0.4860 & 0.5037 \\ -0.0174 & -0.0510 & 0.2350 \end{bmatrix},$$

$$\widehat{A}_3 = \begin{bmatrix} -0.0053 & 0.3407 & -0.5354 \\ 0.0275 & 0.4832 & -0.3212 \\ 0.0115 & -0.0052 & 0.0815 \end{bmatrix},$$

$$\widehat{A}_4 = \begin{bmatrix} -0.0411 & 0.0013 & -0.0268 \\ -0.0226 & -0.1642 & -0.3320 \\ 0.0667 & -0.0137 & 0.2463 \end{bmatrix},$$

and

$$\widehat{\Sigma}_u = \begin{bmatrix} 0.6031 & 0.0795 & -0.0214 \\ 0.0795 & 0.6565 & 0.0375 \\ -0.0214 & 0.0375 & 0.0684 \end{bmatrix}.$$

2.3.2 Restricted Generalized Least Squares

The LS estimator $\hat{\alpha}$ in expression (2.3.2) is identical to the GLS estimator if no restrictions are imposed on the parameters. If there are parameter restrictions, LS estimation will be asymptotically inefficient and GLS estimation may be preferable. Suppose that there are linear restrictions on the parameters. For example, some of the lagged variables may be excluded from some of the equations. Such restrictions can be expressed by defining a suitable $(K^2 p + K) \times M$ restriction matrix R of rank M such that

$$\alpha = R\gamma, \quad (2.3.7)$$

where γ is the $M \times 1$ vector of unrestricted parameters. The GLS estimator for γ then is

$$\widehat{\gamma} = [R' (ZZ' \otimes \Sigma_u^{-1}) R]^{-1} R' \text{vec}(\Sigma_u^{-1} Y Z'), \quad (2.3.8)$$

where, as before, $Y \equiv [y_1, \dots, y_T]$ is $K \times T$ and $Z \equiv [Z_0, \dots, Z_{T-1}]$ is $(Kp + 1) \times T$. The GLS estimator has standard asymptotic properties under general conditions. In particular, under common assumptions it is consistent and asymptotically normally distributed. More precisely,

$$\sqrt{T}(\widehat{\gamma} - \gamma) \xrightarrow{d} \mathcal{N}(0, (R' \Sigma_u^{-1} R)^{-1}). \quad (2.3.9)$$

In practice, the white noise covariance matrix is unknown and has to be replaced by an estimator that may be based on unrestricted LS estimation of the model. Replacing Σ_u by a consistent estimator $\widehat{\Sigma}_u$ results in a feasible GLS estimator, denoted $\widehat{\gamma}$, with the same asymptotic properties as the GLS estimator (e.g., Lütkepohl 2005, chapter 5). The corresponding feasible GLS estimator of α , $\widehat{\alpha} = R\widehat{\gamma}$, is also consistent and asymptotically normal such that

$$\sqrt{T}(\widehat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, R(R' \Sigma_u^{-1} R)^{-1} R'). \quad (2.3.10)$$

Alternatively, an iterated version of the feasible GLS estimator involves reestimating the white noise covariance matrix from the first round feasible GLS residuals and using that estimator in the next round. This procedure may be continued until convergence. The asymptotic properties of the resulting estimators for γ and α are the same as without iteration.

The overall conclusion from this analysis is that standard estimation procedures can be used for restricted VAR models. They are valid asymptotically under the usual assumptions. Likewise, the asymptotic distribution of nonlinear functions of γ or α may be obtained by the delta method as in (2.3.6).

As an illustration, consider restricting the lagged feedback from GNP deflator inflation to real GNP growth at all lags in the empirical example already considered in the previous section. In that case $M = 35$ and R is a 39×35 matrix with rows 10, 19, 28, and 37 being zero and all other rows having one unit element and consisting of zeros otherwise. The feasible GLS estimates are

$$\begin{aligned}\hat{\vec{v}} &= \begin{bmatrix} 0.7517 \\ -0.3888 \\ 0.0058 \end{bmatrix}, \\ \hat{A}_1 &= \begin{bmatrix} 0.2109 & 0.0219 & 0 \\ 0.3131 & 1.0985 & 0.5455 \\ 0.0017 & 0.0631 & 0.4237 \end{bmatrix}, \\ \hat{A}_2 &= \begin{bmatrix} 0.1903 & -0.3771 & 0 \\ 0.1835 & -0.4848 & 0.4857 \\ -0.0166 & -0.0513 & 0.2398 \end{bmatrix}, \\ \hat{A}_3 &= \begin{bmatrix} 0.0108 & 0.3243 & 0 \\ 0.0297 & 0.4810 & -0.2506 \\ 0.0109 & -0.0047 & 0.0625 \end{bmatrix}, \\ \hat{A}_4 &= \begin{bmatrix} -0.0230 & -0.0127 & 0 \\ -0.0202 & -0.1661 & -0.3285 \\ 0.0660 & -0.0132 & 0.2454 \end{bmatrix}.\end{aligned}$$

2.3.3 Bias-Corrected LS

The LS estimator of the VAR slope parameters, α , may be substantially biased in small samples. For given T , the small-sample bias depends on the specification of the deterministic regressors. If no deterministic regressors are needed for an adequate representation of the DGP, the bias tends to be negligible for realistic sample sizes. The inclusion of an intercept in the VAR model (or equivalently demeaning the data prior to the analysis) substantially increases

the bias for given T . Including an additional deterministic time trend further exacerbates the bias, again holding T fixed.

This LS bias may be estimated and corrected for. There is little interest in applied work in VAR models with a known mean of zero. Closed-form solutions for the asymptotic first-order mean bias in stationary VAR models with an intercept (but no other deterministic regressors) have been derived by Nicholls and Pope (1988) under the assumption of Gaussian iid innovations and by Pope (1990) without assuming Gaussianity. There are no closed-form solutions for the bias in VAR models including deterministic trends. For the latter case, Kilian (1998c) proposes a bootstrap estimator that can be easily adapted to any set of deterministic regressors in the VAR model. This bootstrap estimator is asymptotically as accurate as the closed-form solution when the latter exists, but can accommodate situations for which no closed-form bias estimates are currently available. Unless the VAR model is very large or it includes a deterministic time trend, the closed-form solution is preferred because of its lower computational cost.

Our derivation of the formula for the first-order mean bias for stationary VAR models follows Pope (1990) who postulates a $\text{VAR}(p)$ model without a constant term for mean-adjusted data, which is equivalent to fitting a $\text{VAR}(p)$ model with constant to the unadjusted data. Without loss of generality the underlying DGP can be written in $\text{VAR}(1)$ form similar to (2.2.6),

$$Y_t = \mathbf{A}Y_{t-1} + U_t.$$

The bias of the LS estimator $\widehat{\mathbf{A}}$ for \mathbf{A} is

$$-\mathbb{B}_{\mathbf{A}}/T + O(T^{-3/2}), \quad (2.3.11)$$

where

$$\begin{aligned} \mathbb{B}_{\mathbf{A}} = \Sigma_U & \left[(I_{Kp} - \mathbf{A}')^{-1} + \mathbf{A}'(I_{Kp} - \mathbf{A}'^2)^{-1} \right. \\ & \left. + \sum_{\lambda} \lambda(I_{Kp} - \lambda\mathbf{A}')^{-1} \right] \Gamma_Y(0)^{-1}, \end{aligned} \quad (2.3.12)$$

$\Gamma_Y(0) = \mathbb{E}(Y_t Y_t')$, $\Sigma_U = \mathbb{E}(U_t U_t')$ and the sum is over the eigenvalues λ of \mathbf{A} , weighted by their multiplicities (see Pope 1990).

The direction of the bias in general depends on the autoregressive roots of the VAR process. For persistent processes, the bias tends to be downward. Obviously, this bias vanishes for $T \rightarrow \infty$. Although the bias of the LS estimator does not affect its asymptotic distribution, it may have a substantial effect on the LS estimates in samples of typical size. Using the closed-form solution (2.3.12), a bias-corrected LS estimator for \mathbf{A} may be obtained by

substituting estimators for Σ_U , \mathbf{A} , its eigenvalues and $\Gamma_Y(0)$ in the formula for $\widehat{\mathbb{B}}_{\mathbf{A}}$ in expression (2.3.12) to obtain $\widehat{\mathbb{B}}_{\mathbf{A}}$ and adding $\widehat{\mathbb{B}}_{\mathbf{A}}/T$ to the LS estimator of \mathbf{A} . As estimators for Σ_U and $\Gamma_Y(0)$ the usual quantities $\widehat{\Sigma}_U = T^{-1} \sum_{t=1}^T \widehat{U} \widehat{U}'$ and $\widehat{\Gamma}_Y(0) = T^{-1} \sum_{t=1}^T Y_t Y_t'$ can be used. The first K rows of the bias-corrected estimator of \mathbf{A} are the bias-corrected estimators of A_1, \dots, A_p .³

A practical concern is that implementing bias adjustments may push the bias-corrected LS estimator into the explosive region (i.e., some of the eigenvalues of the bias-corrected estimator of \mathbf{A} may be outside the complex unit circle), when the original estimate of \mathbf{A} has eigenvalues close to the unit circle. Kilian (1998c) therefore proposes to shrink the bias adjustment such that the bias-corrected estimate remains stationary. If the original estimate of \mathbf{A} is already unstable or explosive, no bias adjustment is carried out. This modification helps reduce the variance of the bias-corrected estimator while preserving the asymptotic normality of the LS estimator.

In practice, users of VAR models are typically not interested in the slope parameters themselves, but in smooth nonlinear functions of the VAR model parameters such as impulse responses. As demonstrated in Kilian (1998c), bias in the slope parameters tends to be associated with severe bias in the impulse response estimator, which often can be ameliorated by using bias-adjusted LS estimators in constructing the impulse responses. It is important to emphasize, however, that unbiasedness is not preserved under nonlinear transformation, so there is no reason for this alternative impulse response estimator to be unbiased in finite samples. Moreover, reductions in bias tend to be associated with increases in the variance, so it is not possible to prove that in general bias corrections will reduce the mean-squared error (MSE) of the impulse response estimator. Indeed, there is no consensus in the literature that impulse responses should be estimated based on bias-adjusted slope parameters rather than the original LS estimates. Simulation experiments, however, show that bias adjustments of the VAR slope parameters systematically and often substantially improve the coverage accuracy of bootstrap confidence intervals for impulse responses and related statistics (see, e.g., Kilian 1999b, 1998b). It is in the latter context that such bias adjustments have become an important tool in applied work. While bias adjustments almost invariably tend to improve the accuracy of bootstrap confidence intervals when the data are persistent, it is important to be aware that for samples of typical size bias adjustments may not be enough to ensure accurate inference in stationary VAR models, when the estimated model includes a deterministic time trend. A more extensive discussion of these issues is provided in Chapter 12.

³ Iterating this bias correction tends to produce only small improvements in accuracy. Typically, the first-order estimate is quite close to the iterated bias estimate. Likewise, the effect of bias corrections on $\widehat{\Sigma}_U$ is of second order and can be ignored (see Kilian 1998c).

Continuing with the empirical example used in the two preceding subsections, we obtain

$$\begin{aligned}\widehat{A}_1^{BC} &= \begin{bmatrix} 0.2358 & 0.0099 & 0.3947 \\ 0.3145 & 1.1119 & 0.5872 \\ 0.0013 & 0.0634 & 0.4210 \end{bmatrix}, \\ \widehat{A}_2^{BC} &= \begin{bmatrix} 0.2292 & -0.3841 & 0.1307 \\ 0.1820 & -0.4851 & 0.4846 \\ -0.0165 & -0.0517 & 0.2456 \end{bmatrix}, \\ \widehat{A}_3^{BC} &= \begin{bmatrix} 0.0016 & 0.3457 & -0.5544 \\ 0.0204 & 0.4875 & -0.3498 \\ 0.0121 & -0.0059 & 0.0824 \end{bmatrix}, \\ \widehat{A}_4^{BC} &= \begin{bmatrix} -0.0291 & 0.0085 & -0.0458 \\ -0.0333 & -0.1659 & -0.3633 \\ 0.0664 & -0.0121 & 0.2544 \end{bmatrix},\end{aligned}$$

where the dominant root, as measured by the maximum of the modulus of the eigenvalues of \mathbf{A} , increases from 0.9502 for the LS estimate in Section 2.3.1 to 0.9689 after the first-order mean bias adjustment.

2.3.4 Maximum Likelihood Estimation

If the sample distribution is known to have probability density function $f(y_1, \dots, y_T)$, ML estimation is possible. Denoting the vector of all parameters including the innovation covariance parameters by θ and using the decomposition,

$$f(y_1, \dots, y_T | \theta) = f_1(y_1) \times f_2(y_2 | y_1) \times \dots \times f_T(y_T | y_{T-1}, \dots, y_1),$$

the log-likelihood is

$$\log l(\theta | y_1, \dots, y_T) = \sum_{t=1}^T \log f_t(y_t | y_{t-1}, \dots, y_1). \quad (2.3.13)$$

The maximizing vector $\tilde{\theta}$ is the ML estimator. Under conditions mirroring those required for the LS estimator, this conditional ML estimator has an asymptotic normal distribution,

$$\sqrt{T}(\tilde{\theta} - \theta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_a(\theta)^{-1}),$$

where $\mathcal{I}_a(\theta)$ is the asymptotic information matrix. Recall that the information matrix is defined as minus the expectation of the Hessian of the log-likelihood,

$$\mathcal{I}(\theta) = -\mathbb{E} \left[\frac{\partial^2 \log l}{\partial \theta \partial \theta'} \right].$$

The asymptotic information matrix is the limit of this matrix divided by the sample size,

$$\mathcal{I}_a(\theta) = \lim_{T \rightarrow \infty} \mathcal{I}(\theta)/T.$$

The asymptotic normality of the ML estimator permits the use of Wald tests as discussed earlier. It also facilitates the use of likelihood ratio (LR) and Lagrange multiplier (LM) tests. Consider testing the hypotheses

$$\mathbb{H}_0 : \varphi(\theta) = 0 \quad \text{versus} \quad \mathbb{H}_1 : \varphi(\theta) \neq 0, \quad (2.3.14)$$

where $\varphi : \mathbb{R}^M \rightarrow \mathbb{R}^N$ is a continuously differentiable function, $\varphi(\theta)$ is of dimension $N \times 1$, and the $N \times M$ matrix of first-order partial derivatives $\partial\varphi/\partial\theta'$ is assumed to have rank N when evaluated at the true parameter vector.

Then the LR test statistic is

$$LR = 2[\log l(\tilde{\theta}) - \log l(\tilde{\theta}_r)],$$

where $\tilde{\theta}$ and $\tilde{\theta}_r$ denote the unrestricted and restricted ML estimators, respectively. Under the usual conditions, it has a $\chi^2(N)$ distribution under \mathbb{H}_0 .

The corresponding LM statistic for testing (2.3.14) is based on the score vector

$$s(\theta) = \frac{\partial \log l(\theta)}{\partial \theta},$$

which is zero when evaluated at the unrestricted ML estimator. The LM statistic measures the distance of this score vector from zero, when the score vector is evaluated at the restricted estimator. It is defined as

$$LM = s(\tilde{\theta}_r)' \mathcal{I}(\tilde{\theta}_r)^{-1} s(\tilde{\theta}_r) \quad (2.3.15)$$

and has an asymptotic $\chi^2(N)$ distribution under \mathbb{H}_0 . It can be expressed equivalently as

$$LM = \tilde{\lambda}' \left[\frac{\partial \varphi}{\partial \theta'} \Big|_{\tilde{\theta}_r} \right] \mathcal{I}(\tilde{\theta}_r)^{-1} \left[\frac{\partial \varphi'}{\partial \theta} \Big|_{\tilde{\theta}_r} \right] \tilde{\lambda}, \quad (2.3.16)$$

where $\tilde{\lambda}$ is the vector of Lagrange multipliers for which the Lagrange function has derivative zero when evaluated at the constrained estimator (see Lütkepohl 2005, appendix C.7).

Wald, LR, and LM test statistics have the same asymptotic distributions under the null hypothesis. They are based on the unrestricted estimator only, both the unrestricted and the restricted estimators, and the restricted estimator only, respectively. The Wald test has the disadvantage that it is not invariant under nonlinear transformations of the restrictions. Its small-sample power may be low, as shown by Gregory and Veall (1985) and Breusch and Schmidt (1988), for example.

ML theory is valid for large classes of distributions. In VAR analysis, it is common to postulate that the innovations, u_t , are iid $\mathcal{N}(0, \Sigma_u)$ random variables. This assumption implies that the y_t ($= v + A_1y_{t-1} + A_2y_{t-2} + \dots + A_py_{t-p} + u_t$) are also jointly normal and, for given initial values y_{-p+1}, \dots, y_0 ,

$$f_t(y_t | y_{t-1}, \dots, y_{-p+1}) = \left(\frac{1}{2\pi} \right)^{K/2} \det(\Sigma_u)^{-1/2} \exp \left(-\frac{1}{2} u_t' \Sigma_u^{-1} u_t \right). \quad (2.3.17)$$

Hence, the log-likelihood becomes

$$\begin{aligned} \log l = & -\frac{KT}{2} \log 2\pi - \frac{T}{2} \log(\det(\Sigma_u)) \\ & - \frac{1}{2} \sum_{t=1}^T (y_t - v - A_1y_{t-1} - \dots - A_py_{t-p})' \Sigma_u^{-1} \\ & \times (y_t - v - A_1y_{t-1} - \dots - A_py_{t-p}). \end{aligned} \quad (2.3.18)$$

Maximizing this function with respect to the unknown parameters yields the Gaussian ML estimator. If there are no restrictions on the parameters, this ML estimator for $\alpha = \text{vec}[v, A_1, \dots, A_p]$ is identical to the LS estimator and thus has the same asymptotic distribution as the LS estimator. Hence, in practice, one can rely on expression (2.3.2) when constructing the ML estimator of α . The corresponding ML estimator $\tilde{\Sigma}_u = T^{-1}\hat{U}\hat{U}'$ may be obtained as $\tilde{\Sigma}_u = \hat{\Sigma}_u(T - Kp - 1)/T$.

The numerical equivalence of the LS and the Gaussian ML estimator of α holds even when the true distribution of y_t is not Gaussian. In the latter case, the estimator is referred to as a quasi-ML or pseudo-ML estimator. It should be understood, however, that if the true error distribution is different from the normal distribution, it may be possible to obtain more precise ML estimators by imposing the true distribution in constructing the likelihood. In that case, maximizing the log-likelihood involves a nonlinear optimization problem and requires suitable iterative algorithms.

If there are additional restrictions on the parameters as in Section 2.3.2 and these restrictions are taken into account in the estimation, the Gaussian ML estimator may differ from the restricted GLS estimator, but again the same asymptotic properties are obtained. In this case, implementing the restricted ML estimator also requires numerical optimization methods.

Returning to the empirical example, the Gaussian ML estimator of the slope parameters and of the intercept are identical to the LS estimator. The corresponding ML estimator of Σ_u is

$$\tilde{\Sigma}_u = \begin{bmatrix} 0.5656 & 0.0746 & -0.0201 \\ 0.0746 & 0.6157 & 0.0351 \\ -0.0201 & 0.0351 & 0.0642 \end{bmatrix}.$$

2.3.5 VAR Processes in Levels with Integrated Variables

One may also estimate a VAR process with integrated variables using the same techniques. The LS/ML estimator is consistent and asymptotically normal under general conditions and the same estimator may be used as in the stationary case (see Park and Phillips 1988, 1989; Sims, Stock, and Watson 1990; Lütkepohl 2005, chapter 7). More precisely, in VAR(p) models with $p > 1$, standard Gaussian inference on individual VAR slope parameters remains asymptotically valid even in the presence of $I(1)$ variables. The intuition is that the slope parameter matrices A_i in levels VARs can be written as linear combinations of estimators of other parameter matrices that are asymptotically Gaussian because they relate to stationary regressors.

Note that the VAR(p) model

$$y_t = v + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + u_t$$

is algebraically equivalent to the reparametrized specification

$$y_t = v + C y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t,$$

where $C = \sum_{i=1}^p A_i$ and $\Gamma_i = -(A_{i+1} + \cdots + A_p)$, $i = 1, \dots, p - 1$. The latter representation may be viewed as the multivariate analogue of the augmented Dickey-Fuller (ADF) representation of the univariate AR model and is derived in the same manner. Estimating this reparameterized model by LS and computing estimates

$$\begin{aligned}\widehat{A}_1 &= \widehat{C} + \widehat{\Gamma}_1, \\ \widehat{A}_i &= \widehat{\Gamma}_i - \widehat{\Gamma}_{i-1}, \quad i = 2, \dots, p - 1, \\ \widehat{A}_p &= -\widehat{\Gamma}_{p-1},\end{aligned}$$

is equivalent to estimating the levels VAR representation directly by LS. Assuming that all variables in y_t are at most $I(1)$, the estimators $\widehat{\Gamma}_i$, $i = 1, \dots, p - 1$, converge at rate \sqrt{T} and have asymptotically normal marginal distributions because the lagged differences are $I(0)$ regressors. Since the asymptotic distribution of \widehat{A}_i , $i = 2, \dots, p$, consists of linear combinations of asymptotically normal estimators, by construction they must be asymptotically normal as well. Likewise, \widehat{A}_1 can be shown to be asymptotically normal. Note that y_{t-1} is an $I(1)$ regressor, so the LS estimator \widehat{C} may converge at rate T (see Chapter 3 for details). If so, it does not enter the limiting distribution of \widehat{A}_1 after \sqrt{T} -scaling. This fact allows us to ignore that \widehat{C} may have a non-Gaussian limiting distribution after T -scaling. In other words, the nature of the asymptotic distribution of \widehat{C} is irrelevant for the derivation of the asymptotic distribution of \widehat{A}_1 . The latter distribution is dominated by the distribution of $\widehat{\Gamma}_1$, which in turn is known to be asymptotically normal. To summarize, provided $p > 1$ in the VAR(p) model and an intercept is included in estimation, the LS estimator

of \widehat{A}_i , $i = 1, \dots, p$, remains consistent and the marginal asymptotic distributions remain asymptotically normal even in the possible presence of a unit root or, for that matter, a near-unit root (see Chapter 3).

There is a difference in the joint asymptotic properties of the estimated VAR coefficients, however, that is worth emphasizing. If there are integrated variables, the covariance matrix $\Sigma_{\hat{\alpha}}$ of the asymptotic distribution is singular because some components of the estimator (or their linear combinations) converge at rate T rather than \sqrt{T} . As a result, standard tests of hypotheses involving several VAR parameters jointly may be invalid asymptotically (see Toda and Phillips 1993). Hence, caution is called for in conducting inference.

Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996) show that a reparametrization of the model allows us to overcome the problems due to a nonsingular covariance matrix of the estimator. They show that if y_t consists of $I(0)$ and $I(1)$ variables only, it suffices to add an extra lag to the VAR process fitted to the data to obtain a nonsingular covariance matrix for the parameters associated with the first p lags. In other words, if the true DGP is a VAR(p) process and a lag-augmented VAR($p + 1$),

$$y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + A_{p+1} y_{t-p-1} + u_t,$$

is fitted by LS, then the estimator of $[A_1, \dots, A_p]$ has a nonsingular joint asymptotic distribution. Hence, testing hypotheses on linear combinations of these parameters with a Wald test with the usual χ^2 distribution is asymptotically valid. More generally, if $I(d)$ variables are present, the singularity problem of the covariance matrix can be resolved by augmenting the VAR model by d extra lags for estimation. Extensions of this result to unit root, local-to-unity, and long memory VAR processes of infinite order have been provided in Bauer and Maynard (2012) (see Chapter 3).

Of course, lag augmentation may involve a loss of efficiency in estimation, reducing the power of tests and inflating the width of confidence intervals, especially when the parameters of interest would have been estimated superefficiently in the absence of the redundant lag. Thus, there is a trade-off between robustness and efficiency.⁴

If there are parameter restrictions and GLS estimation is applied to $I(1)$ processes, a more detailed analysis of the integration and cointegration properties of the left-hand and right-hand side variables is called for to determine the asymptotic properties of the estimators and the associated inference procedures. It may then be preferable to estimate the process in vector error correction form. Suitable procedures are discussed in Chapter 3.

⁴ An alternative estimation approach that is robust to moderate deviations from unit roots in either direction and that does not require lag augmentation is discussed in Magdalinos and Phillips (2009). It is not clear, however, what their asymptotic results imply for inference on structural impulse responses and related statistics.

Returning to our empirical example, one could alternatively interpret i_t and Δp_t as $I(1)$ variables (such that p_t is $I(2)$ rather than $I(1)$) with a cointegrating relationship between i_t and Δp_t , such that the real interest rate $i_t - \Delta p_t$ is $I(0)$. The results of this section imply that using the LS estimates obtained earlier under the assumption of stationary data would remain justified in this case.

Alternatively, one could fit a lag-augmented VAR model. When augmenting the VAR(4) model already used as an empirical example earlier in this chapter by an additional lag and fitting a VAR(5) model to the data, we obtain the LS estimates

$$\begin{aligned}\widehat{A}_1 &= \begin{bmatrix} 0.2114 & 0.0465 & 0.3993 \\ 0.3220 & 1.0879 & 0.4908 \\ 0.0044 & 0.0553 & 0.4234 \end{bmatrix}, \\ \widehat{A}_2 &= \begin{bmatrix} 0.1694 & -0.4387 & 0.1244 \\ 0.2091 & -0.4762 & 0.5008 \\ -0.0154 & -0.0458 & 0.2429 \end{bmatrix}, \\ \widehat{A}_3 &= \begin{bmatrix} 0.0199 & 0.3888 & -0.5773 \\ 0.0080 & 0.4586 & -0.2666 \\ 0.0083 & -0.0062 & 0.1125 \end{bmatrix}, \\ \widehat{A}_4 &= \begin{bmatrix} 0.0191 & -0.1489 & -0.2000 \\ -0.0706 & -0.1259 & -0.2355 \\ 0.0606 & -0.0019 & 0.3142 \end{bmatrix},\end{aligned}$$

and

$$\widehat{\Sigma}_u = \begin{bmatrix} 0.5852 & 0.1034 & -0.0147 \\ 0.1034 & 0.6346 & 0.0339 \\ -0.0147 & 0.0339 & 0.0675 \end{bmatrix}.$$

The estimate of the augmented lag term, which under the maintained assumption of a VAR(4) DGP is known to be $0_{3 \times 3}$ in population, is

$$\widehat{A}_5 = \begin{bmatrix} -0.0888 & 0.1173 & 0.2491 \\ 0.2161 & -0.0208 & 0.0171 \\ 0.0179 & -0.0053 & -0.1335 \end{bmatrix}.$$

The latter estimate is ignored in conducting further analysis.

2.3.6 Sieve Autoregressions

It is easy to construct plausible autoregressive DGPs that are not of finite lag order. For example, the autoregressive representation of an invertible VARMA process may be a VAR(∞) process. MA components arise naturally when data are aggregated over time, across households or across sectors, or when interest

centers on a subset of the variables in a higher-dimensional VAR process (see the discussion in Section 2.2.4). They also typically arise when the data are generated by a dynamic stochastic general equilibrium (DSGE) model (see Chapter 6).

Once we allow for the possibility that the DGP is a linear $\text{VAR}(\infty)$ process, we can reinterpret the fitted VAR model with lag order p_T as an approximation to the infinite-order DGP. The subscript is a reminder that the approximating lag order depends on T . The thought experiment is that the researcher fits a sequence of finite-order autoregressions, the lag order of which is assumed to increase at a suitable rate with the sample size. The fitted VAR model is viewed as an approximation to the possibly infinite-order autoregression, the quality of which improves with the sample size. In the limit, the approximation error becomes negligible. Hence, claims that VAR models suffer from omitted-variable bias when the DGP is a VARMA model are not correct in general (see, e.g., Braun and Mittnik 1993; Cooley and Dwyer 1998; Yao, Kam, and Vahid 2017).

Whereas the traditional finite-order VAR model has been developed as a parametric time series model, the $\text{VAR}(p_T)$ model is a semiparametric model designed to approximate general linear time series models. Such methods are also commonly referred to as sieve autoregressions in the literature. Lewis and Reinsel (1985) establish that the LS estimator $\hat{\alpha}_T$ remains consistent and asymptotically normal when fitting a $\text{VAR}(p_T)$ model to a stationary $\text{VAR}(\infty)$ process, provided $p_T \rightarrow \infty$ at a rate that is not too fast and not too slow. In particular, we require the lower bound on the lag order to increase with T such that

$$\sqrt{T} \sum_{i=p_T+1}^{\infty} ||A_i|| \rightarrow 0$$

and the upper bound to increase such that $p_T^3/T \rightarrow 0$. Analogous results also apply to the estimation of the innovation variance Σ_u (Lütkepohl and Poskitt 1991; Lütkepohl 2005, section 15.2). Gonçalves and Kilian (2007) generalize this result to infinite-order processes with conditionally heteroskedastic martingale difference errors under the stronger assumption that $p_T^4/T \rightarrow 0$. The sieve approach has also proved useful in studying cointegrated processes, building on the framework developed by Saikkonen (1992) and Saikkonen and Lütkepohl (1996), who consider approximating cointegrated linear systems with iid innovations via autoregressive sieves. Thus, allowing for approximation error does not affect how the VAR parameters are estimated. The only difference is the interpretation of the fitted model and the fact that achieving a good approximation may require a larger lag order than commonly considered in applied work.

There are some subtle differences, however. For example, Onatski and Uhlig (2012) show that using the sieve approach may result in estimated processes

with roots near the unit circle, even if the actual underlying DGP is a stationary process with autoregressive roots well away from the unit circle. In fact, even if the DGP is white noise and the sieve approach is used for estimation, estimated roots near the unit circle will eventually be obtained. Thus, inferring from estimated roots near the unit circle that the underlying DGP is very persistent can be misleading.

Another important caveat is that the validity of the autoregressive sieve approximation to a VARMA process has not been investigated for VARMA DGPs, in which some roots of the autoregressive lag polynomial and of the moving-average lag polynomial nearly cancel. As noted earlier, these roots are functions of the model coefficients. For expository purposes consider the limiting case of a VARMA(1, 1) process, in which the values of the MA and AR coefficient matrices (and hence the roots) are exactly equal. In that case,

$$y_t = A_1 y_{t-1} + u_t - M_1 u_{t-1}$$

with white noise errors, u_t , where $A_1 = M_1$. This process is equivalent to the white noise process

$$y_t = u_t,$$

rendering the slope parameters A_1 and M_1 unidentified. A perhaps practically more relevant case is a VARMA(1, 1) model, in which some AR and MA roots are nearly equal in the sense that their difference is local to zero (see Andrews and Mikusheva 2015). This situation gives rise to a near-identification problem in the VARMA model that is likely to complicate estimation and inference for autoregressive sieve approximations. The theoretical properties of estimators of sieve autoregressions in this setting are not known. In practice, users of the sieve VAR approach effectively assume that there are no near-identification problems.

Extensions of the asymptotic properties of sieve estimators to smooth and differentiable functions of the slope parameters are straightforward, following the same line of reasoning as in Section 2.3.1. For example, Lütkepohl (1988) derives the asymptotic normal distribution of the estimated dynamic multipliers for the VAR(∞) model. Under standard assumptions estimators of impulse responses and related statistics will also be \sqrt{T} -consistent and asymptotically normal. There is one important difference, however. Notwithstanding the invariance of the distribution of the LS estimator of the VAR parameters, the asymptotic variance of the dynamic multipliers will differ, depending on whether the underlying DGP is a finite-order VAR model or a VAR(∞) model, forcing the user to choose between alternative asymptotic approximations. This choice may be avoided by the use of bootstrap methods of inference. For example, Paparoditis (1996) showed that an asymptotically valid bootstrap approximation of the distribution of dynamic multipliers can be constructed in exactly the same way whether the underlying DGP is a finite-order VAR

model or a VAR(∞) model. Further results for structural VAR models can be found in Inoue and Kilian (2002b). We will return to this point in Chapter 12 when discussing inference on structural impulse responses.

2.4 Prediction

It is useful to review some of the basic facts about VAR prediction. Reduced-form VAR models represent the conditional mean of a stochastic process given past observations. Hence, they are natural tools for prediction. Before we discuss how to predict future realizations of the data from estimated VAR processes, it is useful to treat the true process as known, allowing us to neglect estimation uncertainty.

2.4.1 Predicting from Known VAR Processes

Suppose y_t is generated by the K -dimensional VAR(p) process (2.2.4),

$$y_t = \nu + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t.$$

If the white noise process u_t is a martingale difference such that $\mathbb{E}(u_t|y_{t-1}, y_{t-2}, \dots) = \mathbb{E}(u_t|u_{t-1}, u_{t-2}, \dots) = 0$, then

$$\begin{aligned} y_{T+h|T} &\equiv \mathbb{E}(y_{T+h}|y_T, y_{T-1}, \dots) \\ &= \nu + A_1 y_{T+h-1|T} + \cdots + A_p y_{T+h-p|T}, \end{aligned} \quad (2.4.1)$$

where $y_{T+j|T} = y_{T+j}$ for $j \leq 0$, is the optimal, minimum mean squared error (MSE) h -step ahead prediction, given y_t , $t \leq T$. The martingale difference property of u_t is, for instance, satisfied if the u_t are a zero mean iid sequence (see Section 2.2.1). By iterating forward equation (2.4.1), the predictions can easily be computed recursively for $h = 1, 2, \dots$.

The prediction error associated with an h -step ahead prediction is

$$y_{T+h} - y_{T+h|T} = u_{T+h} + \Phi_1 u_{T+h-1} + \cdots + \Phi_{h-1} u_{T+1}, \quad (2.4.2)$$

where $\Phi_i = JA^i J'$, $i = 0, 1, \dots$, as defined in Section 2.2. Equation (2.4.2) implies that the VAR innovation u_t is the prediction error for the 1-step ahead prediction as of period $t - 1$. The errors have mean zero and, hence, the predictions are unbiased. The prediction error covariance or mean-squared prediction error (MSPE) matrix is

$$\Sigma_y(h) \equiv \mathbb{E}[(y_{T+h} - y_{T+h|T})(y_{T+h} - y_{T+h|T})'] = \sum_{j=0}^{h-1} \Phi_j \Sigma_u \Phi_j'. \quad (2.4.3)$$

In short, $y_{T+h} - y_{T+h|T} \sim (0, \Sigma_y(h))$. It is worth emphasizing that these expressions hold not only for stationary but also for integrated processes for any finite h .

If y_t is $I(0)$, then, as $h \rightarrow \infty$,

$$\Sigma_y(h) \rightarrow \Sigma_y \equiv \Gamma_y(0)$$

(see expression (2.2.10)). In other words, the prediction error covariance matrix converges to the unconditional covariance matrix of y_t when the prediction horizon goes to infinity and, hence, the prediction uncertainty remains bounded. In contrast, for integrated processes, the Φ_i do not converge to zero for $i \rightarrow \infty$ and the unconditional covariance matrix of y_t does not exist. Thus, for integrated processes, the prediction error covariance matrix is unbounded in the limit.

If the u_t are just uncorrelated and not martingale differences (i.e., the conditional mean, given past innovations, is not zero), the predictions obtained recursively from equation (2.4.1) are still the best linear predictions, but may not be minimum MSPE in a larger class of possibly nonlinear predictors.

Note that the prediction error does not depend on the deterministic term v . In fact, all deterministic terms such as deterministic trend polynomials cancel in constructing the prediction error. Hence, they do not contribute to the prediction uncertainty. Many researchers find it implausible that trending behavior is not reflected in the uncertainty about long-term predictions. In the present setup, this is an implication of our assumption that the parameters of the true process are known. If the deterministic terms are estimated, they contribute to the variance of the predictions, as discussed in the next subsection.

2.4.2 Predicting from Estimated VAR Processes

In practice, the VAR process is unknown and estimated parameters are used in computing predictions. Denoting an h -step prediction based on the estimated process by $\hat{y}_{T+h|T}$, the prediction error is

$$y_{T+h} - \hat{y}_{T+h|T} = (y_{T+h} - y_{T+h|T}) + (y_{T+h|T} - \hat{y}_{T+h|T}). \quad (2.4.4)$$

Expression (2.4.2) shows that the first term on the right-hand side includes innovations u_t for $t > T$ only. In contrast, the second term on the right-hand side of expression (2.4.4) involves only observations y_t up to time T . Assuming that the model has been estimated on data up to time T , these two terms are independent (or at least uncorrelated), and the MSE matrix is

$$\begin{aligned} \Sigma_{\hat{y}}(h) &= \mathbb{E}[(y_{T+h} - \hat{y}_{T+h|T})(y_{T+h} - \hat{y}_{T+h|T})'] \\ &= \Sigma_y(h) + \text{MSE}(y_{T+h|T} - \hat{y}_{T+h|T}), \end{aligned} \quad (2.4.5)$$

where the first term on the right-hand side denotes the population MSPE and the second term captures the estimation uncertainty. For a properly specified VAR model, the last term in this relation approaches zero, as the sample size grows, because, under standard assumptions, the difference $y_{T+h|T} - \hat{y}_{T+h|T}$ vanishes asymptotically in probability as $T \rightarrow \infty$. Thus, estimation uncertainty is negligible asymptotically if the fitted model is the DGP. In practice, only finite samples are available, and the precision of the predictions depends on the precision of the parameter estimators. Finite-sample correction factors for MSPEs and prediction intervals for stationary processes are provided in Baillie (1979), Reinsel (1980), Samaranayake and Hasza (1988), and Lütkepohl (2005, chapter 3).

There are several obvious extensions of this framework. First, one may consider explicitly the situation where the DGP is only approximated by the finite-order VAR model used for prediction. Second, if h -step ahead predictions are of interest, one may instead fit models specifically targeted at this horizon. Such extensions are discussed in Lütkepohl (2009).

If $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_u)$, the prediction errors are also multivariate normal, $y_{T+h} - y_{T+h|T} \sim \mathcal{N}(0, \Sigma_y(h))$, and prediction intervals may be set up in the usual way based on the MSPE matrices (2.4.5) (see Lütkepohl 2005).

Constructing prediction intervals for non-Gaussian VAR processes subject to estimation uncertainty requires the use of bootstrap methods, as discussed in Chapter 12.

2.5 Granger Causality Analysis

The VAR model describes the joint DGP of the variables under consideration. A proposal for assessing the dynamic relationship between economic variables based on the VAR model was made by Granger (1969). His definition of what has become known as Granger causality is based on the notion of linear predictability. Granger calls a variable y_{2t} causal for a variable y_{1t} if the information in past and present values of y_{2t} helps reduce in expectation the squared prediction error for y_{1t} . Many researchers find it problematic to interpret a predictive relationship as a causal relationship. We provide a more extensive discussion of this issue in Chapter 7. Here we simply present a formal definition of what has become known as Granger causality.

Let Ω_t be the information set containing all information relevant to predicting y_1 available up to and including period t . Denote the optimal (minimum MSE) h -step prediction of y_{1t} at origin t , based on the information in Ω_t , by $y_{1,t+h|\Omega_t}$ and the corresponding MSPE by $\sigma_{y_1}^2(h|\Omega_t)$. Then the process y_{2t} is said to Granger cause y_{1t} if

$$\sigma_{y_1}^2(h|\Omega_t) < \sigma_{y_1}^2(h|\Omega_t \setminus \{y_{2s}|s \leq t\}) \quad \text{for at least one } h \in \{1, 2, \dots\}. \quad (2.5.1)$$

Here $\Omega_t \setminus \{y_{2s} | s \leq t\}$ denotes the set of all relevant information in the universe apart from the past and present information about the y_{2t} process. In other words, y_{2t} is Granger causal for y_{1t} if the latter variable can be predicted with lower mean squared error by taking into account the information in y_{2s} , $s \leq t$, in addition to all other relevant information.

This concept is easy to implement within a VAR framework if Ω_t is limited to past and present values of y_{1t} and y_{2t} . Suppose these two variables are generated by a bivariate VAR(p) process,

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \sum_{i=1}^p \begin{bmatrix} a_{11,i} & a_{12,i} \\ a_{21,i} & a_{22,i} \end{bmatrix} \begin{pmatrix} y_{1,t-i} \\ y_{2,t-i} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

Then y_{2t} is not Granger causal for y_{1t} if and only if $a_{12,i} = 0$, $i = 1, 2, \dots, p$. In other words, no lags of y_{2t} appear in the y_{1t} equation of the model. In contrast, y_{2t} is Granger causal for y_{1t} if lags of the former variable appear in the y_{1t} equation with nonzero coefficients.

Given that Granger-noncausality is characterized by zero restrictions on the VAR coefficients, it can be tested using a standard Wald test if the process is stationary. If the process contains integrated variables, the testing problem becomes more complicated because the Wald test statistics will not have their usual asymptotic distributions anymore (see Toda and Phillips 1993). If the variables are known to be integrated but not cointegrated, one can appeal to standard asymptotics for the Granger causality test after differencing the data. If the model variables are known to be integrated and cointegrated, Granger causality may be assessed within the vector error correction framework considered in Chapter 3, but the distribution theory may be nonstandard and the asymptotic distributions depend on nuisance parameters.

If we are unsure about the integration and cointegration status of the model variables, an easy cure for this problem is to add a further lag to the VAR process and perform the tests on the first p lags of the lag-augmented VAR, as discussed in Section 2.3.5. More precisely, if the true process is a VAR(p), then a VAR($p+1$) is also a true process with $A_{p+1} = 0$. Since the last lag has coefficient matrix zero, the test for Granger causality can be performed on the first p lags. If the variables are at most $I(1)$, one extra lag is sufficient to ensure that the standard tests have standard asymptotic properties under the null hypothesis. More lags have to be added if the order of integration is higher than 1 (see Toda and Yamamoto 1995; Dolado and Lütkepohl 1996). Extensions to testing Granger causality in infinite-order VAR processes are discussed in Lütkepohl and Poskitt (1996) for stationary processes, in Saikkonen and Lütkepohl (1996) for integrated processes, and in Bauer and Maynard (2012) for local-to-unity processes and fractionally integrated processes (see Chapter 3). As noted in Section 2.3.5, the use of lag augmentation involves a loss of power. There is evidence, however, that the magnitude of this efficiency loss is small

in practice. For example, simulation results in Dolado and Lütkepohl (1996) for lag-augmented Wald tests of Granger causality suggest that the power loss is negligible, provided a reasonably large number of lags is used in the first place.⁵

The introduction of the concept of Granger causality has stimulated a heated debate about causality in econometrics and economics. Some of this discussion has focused on technical problems such as how to extend the information set. While it is straightforward to extend the concept of Granger causality to allow for vectors of variables y_{1t} and y_{2t} instead of scalars (e.g., Lütkepohl 2005, section 2.3.1), it is more difficult to assess the Granger causality for individual variables y_{2t} to y_{1t} in the presence of additional variables. In this case, the conditions for Granger causality become more complicated, even within the framework of finite-order VAR processes (see, e.g., Lütkepohl 1993; Dufour and Renault 1998). The Granger causal ordering of the variables may change if the information set changes. For example, if y_{2t} is Granger causal for y_{1t} in a bivariate model, it may not be Granger causal after adding a third variable to the VAR model, if that third variable is dynamically correlated with the two original variables. It is also possible that y_{2t} is not Granger causal for y_{1t} in a bivariate model and becomes Granger causal if the information set is extended to include other variables, as pointed out in Lütkepohl (1982a).

Another limitation of the concept of Granger causality is that it does not speak to the instantaneous relationships of the variables. This problem can be overcome by defining what has sometimes been referred to as instantaneous causality. In this context a variable y_{2t} is said to be instantaneously causal for y_{1t} if

$$\sigma_{y_1}^2(1|\Omega_t \cup \{y_{2,t+1}\}) < \sigma_{y_1}^2(1|\Omega_t). \quad (2.5.2)$$

In other words, taking into account $y_{2,t+1}$ improves the 1-step prediction of y_{1t} . It turns out that this concept is symmetric, i.e., instantaneous causality from y_{2t} to y_{1t} implies instantaneous causality in the reverse direction from y_{1t} to y_{2t} as well. In fact, in a bivariate VAR model this relation is equivalent to instantaneous correlation of the innovations u_{1t} and u_{2t} (Lütkepohl 2005, section 2.3.1). This observation highlights a fundamental problem with this concept. It fails to distinguish between correlation and causality and it does not speak to the direction of causality. A more extensive discussion of the causality issue is provided in Chapter 7.

⁵ An alternative approach to testing Granger causality, when the model variables are potentially integrated and/or cointegrated, that does not involve lag augmentation has been proposed by Toda and Phillips (1993).

2.6 Lag-Order Selection Procedures

In Section 2.3, the lag order p of the VAR model was assumed to be known when estimating the model. In practice, p often has to be chosen on the basis of the available sample of data. Sequential testing procedures and information criteria are commonly used tools for deciding on an adequate VAR order. Both types of procedures require the specification of an upper bound and a lower bound on the range of admissible lag orders. The maximum lag order that is considered reasonable is denoted by p_{\max} in the following. Typically, the corresponding p_{\min} is set to zero. For sieve autoregressions, both the lower bound p_{\min} and the upper bound p_{\max} on the lag order are required to increase with the sample size T at a suitable rate.

2.6.1 Top-Down Sequential Testing

One possible sequential procedure involves testing the sequence of null hypotheses: $\mathbb{H}_0 : A_{p_{\max}} = 0$ vs. $\mathbb{H}_1 : A_{p_{\max}} \neq 0$, $\mathbb{H}_0 : A_{p_{\max}-1} = 0$ vs. $\mathbb{H}_1 : A_{p_{\max}-1} \neq 0, \dots, \mathbb{H}_0 : A_1 = 0$ vs. $\mathbb{H}_1 : A_1 \neq 0$. The procedure continues as long as the null hypothesis is not rejected. If there is a rejection, the testing procedure terminates, and we conclude that we require as many lags as maintained under the last alternative hypothesis. If none of the null hypotheses can be rejected, we conclude that $p = 0$. Such procedures are known as top-down or general-to-specific procedures because they start with the largest model and sequentially reduce the lag order of the model. In implementing this sequential test, the usual Wald or LR tests for parameter restrictions can be employed. For example, the LR statistic for testing a VAR(m) against a VAR($m + 1$) has the form

$$LR(m) = T[\log(\det(\tilde{\Sigma}_u(m))) - \log(\det(\tilde{\Sigma}_u(m + 1)))],$$

where

$$\tilde{\Sigma}_u(m) = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}'_t$$

is the ML estimator of the residual covariance matrix for a VAR(m) model. If $\mathbb{H}_0 : A_{m+1} = 0$ is tested against $\mathbb{H}_1 : A_{m+1} \neq 0$, the $LR(m)$ statistic has an asymptotic $\chi^2(K^2)$ distribution under \mathbb{H}_0 .

In implementing this procedure it may be worth taking into account that the small-sample distributions of the test statistics can differ substantially from their asymptotic counterparts, in particular for larger models with many variables and lags (e.g., Lütkepohl 2005, section 4.3.4). A simple small-sample adjustment proposed by Sims (1980a) has the form

$$\begin{aligned} LR^{adj}(m) &= (T - K(m + 1)) \\ &\times [\log(\det(\tilde{\Sigma}_u(m))) - \log(\det(\tilde{\Sigma}_u(m + 1)))]. \end{aligned}$$

The LR statistic $LR(m)$ is suitable for testing the $\text{VAR}(m)$ model against the $\text{VAR}(m + 1)$ model. It is also possible to test each lag length against the full $\text{VAR}(p_{\max})$ model using the LR test statistic

$$LR^*(m) = T[\log(\det(\tilde{\Sigma}_u(m))) - \log(\det(\tilde{\Sigma}_u(p_{\max})))]$$

or a similar LR statistic including small-sample adjustments. An alternative approach, which may be viewed as a generalization of the sequential t -test for selecting the lag order of a univariate autoregression, is the sequential Wald test.

A common problem with all such sequential testing procedures is how to control the overall size of the test. A related point of concern is that using the same critical value at each stage of the test does not account for the fact that the tests are conditional on the outcomes of earlier tests (for further discussion, see Lütkepohl 2005, section 4.2.3).

2.6.2 Bottom-Up Sequential Testing

Alternatively, one may consider a bottom-up or specific-to-general approach by starting from the smallest model and adding lags only if residual autocorrelation tests suggest that the model does not fully capture the dynamic structure in the data. In other words, tests for residual autocorrelation are applied to the $\text{VAR}(0)$ model, $\text{VAR}(1)$ model, and so forth. The testing procedure terminates when no statistically significant autocorrelation is found. Thus, we do not have to fix the maximum VAR lag order in advance. Examples of sequential tests for residual autocorrelation in VAR models are the Portmanteau and LM tests.

Portmanteau Test for Residual Autocorrelation. The Portmanteau test for autocorrelation in the innovations evaluates the null hypothesis $\mathbb{H}_0 : \mathbb{E}(u_t u'_{t-i}) = 0, i = 1, 2, \dots$. The alternative hypothesis is that at least one autocovariance is nonzero. The test statistic is

$$Q_h = T \sum_{j=1}^h \text{tr}(\hat{C}'_j \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1}), \quad (2.6.1)$$

where $\hat{C}_j = T^{-1} \sum_{t=j+1}^T \hat{u}_t \hat{u}'_{t-j}$ and the \hat{u}_t are the LS residuals. In other words, the test statistic is based on the estimated residual autocovariances. If the $\text{VAR}(p)$ process is stationary, no parameter restrictions are imposed and the errors are serially independent, the distribution of Q_h under \mathbb{H}_0 is approximately $\chi^2(K^2(h-p))$ when both T and h are large. The approximate χ^2 distribution is obtained, specifically, as $h/T \rightarrow 0$ for $T \rightarrow \infty$.

The degrees of freedom have to be adjusted if there are parameter restrictions. In that case, they are obtained as the difference between the number

of (non-instantaneous) autocovariances included in the statistic ($K^2 h$) and the number of estimated VAR parameters (e.g., Ahn 1988; Hosking 1980, 1981a, 1981b; Li and McLeod 1981; Lütkepohl 2005, section 5.2.9). Likewise, if the VAR process contains integrated components, the degrees of freedom are affected (Brüggemann, Lütkepohl, and Saikkonen 2006). Since the degrees of freedom depend on the typically unknown cointegration properties of the process, the Portmanteau test cannot be recommended for nonstationary processes. Moreover, Francq and Raïssi (2007) show that the asymptotic distribution may be quite different from the usual χ^2 distribution if the errors are not autocorrelated but still contain some dependence structure in higher-order moments. Thus, the test has to be used with caution.

Even when the necessary conditions for deriving the asymptotic distribution hold, the approximate χ^2 distribution may be far from the actual distribution in small samples. To improve the match between actual and approximating distribution, the following modified statistic was proposed by Hosking (1980):

$$Q_h^* = T^2 \sum_{j=1}^h \frac{1}{T-j} \text{tr}(\hat{C}_j' \hat{C}_0^{-1} \hat{C}_j \hat{C}_0^{-1}).$$

The number h of autocovariance terms in the test statistic should be considerably larger than p for a good approximation to the null distribution. Choosing h too large, however, may undermine the power of the test. In practice, usually a number of different values of h is considered. The Portmanteau test should only be applied to test for a large number of nonzero autocovariances. It is not suitable for testing the absence of low-order autocorrelation. For the latter purpose, the LM test is preferred.

LM Test for Residual Autocorrelation. The LM test for autocorrelation in the innovations was proposed by Breusch (1978) and Godfrey (1978). It may be viewed as a test for zero coefficient matrices in the model

$$u_t = D_1 u_{t-1} + \cdots + D_h u_{t-h} + e_t.$$

The quantity e_t denotes a white noise error term. Thus, testing

$$\begin{aligned} \mathbb{H}_0 : D_1 &= \cdots = D_h = 0 \quad \text{versus} \\ \mathbb{H}_1 : D_i &\neq 0 \text{ for at least one } i \in \{1, \dots, h\} \end{aligned}$$

is a test for autocorrelation in u_t .

An easy way of computing the LM statistic based on the residuals of the VAR(p) model is to consider the auxiliary model

$$\hat{u}_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + D_1 \hat{u}_{t-1} + \cdots + D_h \hat{u}_{t-h} + e_t^*, \quad (2.6.2)$$

where the \hat{u}_t are the residuals from the original model, and the \hat{u}_t with $t \leq 0$ are replaced by zero. The term e_t^* is an auxiliary error term. The LM statistic may be computed as

$$Q_{LM} = T \left(K - \text{tr}(\tilde{\Sigma}_u^{-1} \tilde{\Sigma}_e) \right),$$

where $\tilde{\Sigma}_u = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}'_t$, $\tilde{\Sigma}_e = T^{-1} \sum_{t=1}^T \hat{e}_t^* \hat{e}_t^{*'}$ and \hat{e}_t^* ($t = 1, \dots, T$) are the residuals from the estimated auxiliary model. Under the null hypothesis of no autocorrelation, the LM statistic has an asymptotic $\chi^2(hK^2)$ distribution.

In a Monte Carlo simulation study, Edgerton and Shukur (1999) find that this statistic may also have a small-sample distribution that differs substantially from its asymptotic χ^2 distribution and propose an F version with better small-sample properties. It is noteworthy that the asymptotic distribution remains valid under the null hypothesis even if there are integrated variables, as shown by Brüggemann, Lütkepohl, and Saikkonen (2006). As in the case of top-down sequential testing, the usual caveats about sequential testing apply.

2.6.3 Information Criteria

An alternative to sequential testing procedures is the use of information criteria for lag-order selection. Information criteria used for VAR lag-order selection have the general form

$$C(m) = \log(\det(\tilde{\Sigma}_u(m))) + c_T \varphi(m), \quad (2.6.3)$$

where $\tilde{\Sigma}_u(m) = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}'_t$ is the residual covariance matrix estimator for a reduced-form VAR model of order m based on LS residuals \hat{u}_t , m is the candidate lag order at which the criterion function is evaluated, $\varphi(m)$ is a function of the order m that penalizes large lag orders, and c_T is a sequence of weights that may depend on the sample size.

The function $\varphi(m)$ corresponds to the total number of regressors in the system of VAR equations. Since there are mK lagged regressors in each equation and K equations in the VAR model, in the absence of any deterministic regressors, $\varphi(m) = mK^2$. More typically, when including an intercept, $\varphi(m) = mK^2 + K$.

Information criteria are based on the premise that there is a trade-off between the improved fit of the VAR model, as m increases, and the parsimony of the model. Given T , the fit of the model by construction improves with larger m , indicated by a reduction in $\log(\det(\tilde{\Sigma}_u(m)))$. At the same time the penalty term, $c_T \varphi(m)$, unambiguously increases with m . We are searching for the value of m that balances the objectives of model fit and parsimony. The choice of the penalty term determines the trade-off between these conflicting objectives.

The three most commonly used information criteria for VAR models are known as the AIC, HQC, and SIC:

Akaike Information Criterion (AIC)

$$\text{AIC}(m) = \log(\det(\tilde{\Sigma}_u(m))) + \frac{2}{T}(mK^2 + K),$$

where $c_T = 2/T$. This criterion was proposed by Akaike (1973, 1974).

Hannan-Quinn Criterion (HQC)

$$\text{HQC}(m) = \log(\det(\tilde{\Sigma}_u(m))) + \frac{2 \log(\log(T))}{T}(mK^2 + K),$$

where $c_T = 2 \log(\log(T))/T$. This criterion is suggested by the work of Hannan and Quinn (1979) and Quinn (1980).

Schwarz Information Criterion (SIC)

$$\text{SIC}(m) = \log(\det(\tilde{\Sigma}_u(m))) + \frac{\log(T)}{T}(mK^2 + K),$$

where $c_T = \log(T)/T$. This criterion is due to Schwarz (1978) and Rissanen (1978). Because Schwarz (1978) derives this criterion in a Bayesian setting, it is also known as the Bayesian Information Criterion (BIC). It should be noted, however, that most Bayesian users of VAR models would object to the use of information criteria for lag-order selection (see Chapter 5).

The VAR order is chosen such that the respective criterion is minimized over the possible orders $m = 0, \dots, p_{\max}$. Alternative information criteria have been proposed by Hurvich and Tsai (1993), Phillips and Ploberger (1996), Sin and White (1996), and Inoue and Kilian (2006), among others.

One concern with the use of sequential testing procedures is that the sequence of the tests matters. Moreover, as noted earlier, it can be challenging to control the size of sequential testing procedures. It can be shown that the use of information criteria is equivalent to simultaneously testing each candidate model against each other model, with the critical value being determined implicitly by the penalty function. Unlike in sequential testing, no one model is favored because it is chosen as the null hypothesis, and the order in which the criterion function is evaluated does not affect the lag-order choice. This advantage comes at the cost of the user no longer being able to control the effective size of the procedure in finite samples. By choosing an appropriate penalty term in the information criterion, one can, however, make sure that the size approaches zero asymptotically.

A key issue in implementing information criteria is the choice of the upper and lower bounds p_{\max} and p_{\min} . In the context of a model of unknown finite lag order, the default is to set $p_{\min} = 0$ or sometimes $p_{\min} = 1$, reducing the problem to one of choosing a suitable upper bound. The value of p_{\max} must be chosen long enough to allow for delays in the response of the model variables

to the shocks. In practice, common choices would be 12 or 24 lags for monthly data and 4 or 8 lags for quarterly data.

Occasionally users of VAR models implement information criteria for lag-order selection incorrectly. One common mistake relates to the choice of the evaluation period. When evaluating the fit of the model for a given lag order m , it is essential that we compute $\tilde{\Sigma}_u(m)$ on exactly the same evaluation period, $t = p_{\max} + 1, \dots, T$, for all m . If the evaluation period used in computing $\tilde{\Sigma}_u(m)$ differs across $m \in \{1, \dots, p_{\max}\}$, we risk that differences in the fit of different models are driven by the inclusion of additional observations rather than the inclusion of additional lags. In other words, we end up comparing apples and oranges, rendering the resulting ranking meaningless. For example, many canned software packages produce AIC values as a by-product of the regression output. It may seem that these values could be used to rank alternative VAR models according to their AIC values. This is not the case. The reason is that the relevant evaluation period for computing $\tilde{\Sigma}_u(m)$ depends on p_{\max} . Without the user specifying p_{\max} , canned software packages cannot possibly compute the correct estimate $\tilde{\Sigma}_u(m)$ or the corresponding AIC values. Instead, they display AIC values based on the evaluation period $t = m + 1, \dots, T$, where $p_{\min} \leq m \leq p_{\max}$. It may seem that this mistake could not have important consequences, but Ng and Perron (2005) demonstrate by simulation that this mistake can result in severe distortions in the estimated order \hat{p} .

Another potential mistake relates to the estimation of $\Sigma_u(m)$. It is essential that the criterion function be evaluated at the ML estimator $\widehat{\Sigma}_u(m) = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}'_t$ rather than the LS estimator $\widetilde{\Sigma}_u(m) = (T - Km - \kappa)^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}'_t$, where κ is the number of deterministic regressors in each equation. The use of information criteria is predicated on the trade-off between improved fit (measured by decline in $\det(\tilde{\Sigma}_u(m))$) and the reduction in parsimony (measured by an increase in the penalty term), as more lags are added. There is no such trade-off when we estimate $\Sigma_u(m)$ by

$$\widehat{\Sigma}_u(m) = \frac{T}{T - Km - \kappa} \widetilde{\Sigma}_u(m).$$

Adding more lags lowers $\det(\widetilde{\Sigma}_u(m))$ by construction, but it also increases $T/(T - Km - \kappa)$, invalidating the model rankings.

Table 2.1 illustrates the implementation of the SIC, HQC, and AIC criteria in the context of the three-variable VAR model example already used earlier in this chapter to illustrate alternative estimation procedures. Let $m \in \{0, 1, \dots, 4\}$. The lag-order values that minimize a given criterion function are shown in bold. Table 2.1 shows that the SIC favors a lag order of $p = 1$, whereas the HQC chooses a lag order of $p = 3$, and the AIC chooses $p = 4$.

Table 2.1. Alternative Lag-Order Selection Criteria for VAR Models

m	SIC(m)	HQC(m)	AIC(m)
0	0.4594	0.4308	0.4114
1	-2.9750	-3.0893	-3.1669
2	-2.9394	-3.1394	-3.2752
3	-2.9570	-3.2428	-3.4367
4	-2.8697	-3.2412	-3.4933

Note: Based on a VAR(m) model with intercept for $y_t = (\Delta gnp_t, i_t, \Delta p_t)'$, as defined in the earlier empirical example.

2.6.4 Recursive Mean-Squared Prediction Error Rankings

Yet another approach to selecting the lag order of a VAR model is to rank the candidate models based on their recursive mean-squared prediction error (MSPE) in simulated out-of-sample forecasts. The recursive MSPE is obtained as follows. Suppose we have a sample consisting of T observations and candidate VAR models of order $m \in \{p_{\min}, \dots, p_{\max}\}$. Further suppose that we are interested in forecasting one of the VAR model variables. We initially estimate the models under consideration on the first R observations of the sample, $R < T$, and compare their predictions for the variable of interest in period $R + h$ to the realized value. We then expand the initial sample by estimating the models under consideration on the first $R + 1$ observations and construct the prediction error for $R + 1 + h$. We continue this procedure until the estimation window extends to $T - h$. This procedure results in a sequence of $T - R - h$ recursive prediction errors. The recursive MSPE of a given model is the average of its squared recursive prediction errors. We choose the lag order corresponding to the model with the lowest recursive MSPE. This procedure may be adapted to allow for forecasts of more than one VAR variable by evaluating the trace of the MSPE matrix.

This so-called predictive least-squares approach was originally introduced by Rissanen (1986) in the context of one-step ahead prediction. A complete asymptotic analysis is provided in Wei (1992). There is a close relationship between the predictive least-squares approach to lag-order selection and the use of information criteria. Rissanen studied the asymptotic properties of the simulated out-of-sample method under the assumption that the length of the initial recursive sample, R , is fixed with respect to T , where $R < T$ is the length of the initial window of data used for recursive estimation. Under these assumptions, Rissanen's predictive least-squares method can be shown to be asymptotically equivalent to the SIC, provided the true VAR model is included among the VAR models compared. In contrast, Inoue and Kilian (2006)

establish that under the alternative (and more conventional) assumption that R is a fixed fraction of T , the predictive least-squares method is asymptotically equivalent to the AIC.

Which of these asymptotic thought experiments provides a better small-sample approximation is not clear a priori, nor is it clear whether any of these asymptotic results are useful in practice. Simulation evidence for univariate autoregressions in Inoue and Kilian (2006) suggests that the AIC method tends to generate lower one-step ahead MSPEs than predictive least squares when both methods are feasible, and that both methods are less accurate than the SIC. There are no comparable simulation results for multivariate autoregressions nor are there simulation results for multi-step ahead predictions.

Given that information criteria are constructed on the basis of one-step ahead MSPEs rather than multi-step ahead MSPEs, one would expect the predictive least-squares method applied to the h -step ahead MSPE to be particularly useful for selecting the lag order of VAR models used to construct iterated forecasts. This line of inquiry recently has been explored by Ing (2003, 2004) for stationary autoregressive processes. Ing, Lin, and Yu (2009) provide an extension to integrated autoregressive processes. An empirical example of the use of predictive least squares for selecting the lag order of iterated VAR forecasting models is Baumeister and Kilian (2012).

2.6.5 The Relative Merits of Alternative Lag-Order Selection Tools

There are different ways of evaluating the relative merits of alternative lag-order selection criteria.

Consistency for the True VAR Lag Order. If the DGP is a $\text{VAR}(p_0)$ model, one criterion is whether the lag-order estimator is consistent for the true lag order p_0 , provided $p_{\min} \leq p_0 \leq p_{\max}$. Clearly, sequential tests will not be consistent for p_0 because of the positive probability of committing a type I error in statistical testing. In contrast, information criteria will be consistent for p_0 under suitable conditions. The reason is that the probability of committing a type I error vanishes asymptotically if we are careful about the choice of penalty function. As is easy to verify, the AIC is not consistent for p_0 , whereas the SIC and HQC are (see Lütkepohl 2005, section 4.3.2). The HQC was designed to be the least parsimonious yet still consistent information criterion for VAR models. Although standard proofs of the consistency of the SIC and HQC and of the inconsistency of the AIC require the VAR model to be stationary, these results can be extended to VAR models with unit roots (see Paulsen 1984; Tsay 1984).

The price of obtaining a consistent lag-order estimator is greater parsimony in model selection. Parsimony here means that the criterion will favor VAR models with fewer lags. The larger the penalty term for given T , the more

parsimonious the lag-order estimate. It can be shown that for $T \geq 16$,

$$\hat{p}^{\text{SIC}} \leq \hat{p}^{\text{HQC}} \leq \hat{p}^{\text{AIC}},$$

indicating that the SIC tends to be more parsimonious than the HQC, which in turn tends to be more parsimonious than the AIC. Of course, in a given application, all three criteria may suggest the same lag order.

Finite-Sample Properties of the Lag-Order Estimator. Even if we grant the premise that $p_{\min} \leq p_0 \leq p_{\max}$, one may not want to overrate the importance of the lag-order estimator being consistent. The convergence of \hat{p} toward p_0 can be very slow in practice, and in small samples consistent lag-order selection criteria tend to be strongly downbiased toward p_{\min} . For example, Kilian (2001) examines a stylized bivariate AR(4) data generating process and shows that for a sample size of $T = 80$, the SIC selects a lag order of 1 among $1 \leq p \leq 8$ with probability 92%, but the true lag order of 4 only with probability 2%. Even for $T = 160$, the probability of selecting $p = 1$ remains at 61% with a probability of only 28% of selecting the true lag order. Similar, if less extreme results hold for the HQC. In contrast, the AIC has a probability of selecting the true lag order of 57% for $T = 80$ and of 83% for $T = 160$. The probability of the AIC underestimating the lag order is 26% for $T = 80$ and 1% for $T = 160$, compared with 98% and 73% for the SIC. The finding that in small samples the distribution of the AIC lag-order estimates tends to be more balanced about the true lag order than for the SIC lag-order estimates is also consistent with simulation results in Nickelsburg (1985) and Lütkepohl (1985).

The high accuracy of the AIC in these simulation studies – even in large samples – may be surprising at first, but it reflects the asymptotic properties of the AIC. Although

$$\lim_{T \rightarrow \infty} \mathbb{P}(\hat{p}^{\text{AIC}} > p_0) > 0,$$

reflecting the inconsistency of the AIC, it can be shown that

$$\lim_{T \rightarrow \infty} \mathbb{P}(\hat{p}^{\text{AIC}} < p_0) = 0.$$

In other words, in the limit, the AIC will never select a lag order that is lower than p_0 , but it will have a tendency to select with positive probability a lag order in excess of p_0 . This point is important. Economists using VAR models have no inherent interest in the lag order of the process. They are interested in impulse responses, forecasts, and related statistics that can be written as smooth functions of VAR model parameters. These statistics of interest can be consistently estimated, as long as the lag order is not underestimated asymptotically, so there is little to choose between the SIC, HQC, and AIC on the grounds of consistency. The only potential concern is that in large samples the

AIC may choose an excessively large lag order and inflate the variance of, for example, the impulse response estimator.

This concern as well is largely unfounded. Paulsen and Tjøstheim (1985) establish that the probability of the AIC overfitting the VAR model is negligible asymptotically. Whereas for $K = 1$, the asymptotic probability of overfitting is about 30%, for $K = 2$ it drops to at most 12%, for $K = 3$ to 4%, and for $K = 4$ to 1%, so for most VAR applications, the efficiency loss is not a major concern. These asymptotic results are consistent with simulation evidence for large-dimensional VAR models in Gonzalo and Pitarakis (2002) who conclude that the AIC tends to be by far the most reliable estimator of p_0 compared with sequential tests as well as other information criteria.

Finite-sample considerations also favor the AIC. In related work, Kilian (2001) observes that in the context of impulse response analysis in finite samples overestimation of the lag order is costly only to the extent that the impulse response estimates are less precise, but underestimation tends to greatly distort the impulse response functions especially at longer horizons. Hence, users of VAR models ought to employ an asymmetric loss function, erring on the side of including too many lags. Kilian (2001) provides simulation evidence that VAR models based on the AIC lag-order estimate provide more accurate impulse response confidence intervals than VAR models based on more parsimonious lag-order selection criteria. He also shows that AIC-based impulse response estimates have lower MSE. This conclusion is further supported by simulation evidence in Kilian and Ivanov (2005) based on a wide range of monthly finite-order VAR models of the type used in empirical work. Kilian and Ivanov show that for monthly VAR models the AIC implies impulse response estimates with systematically lower MSE than more parsimonious criteria such as the SIC or HQC. For quarterly VAR models, which tend to imply smoother impulse responses, using the HQC generates the impulse response estimates with the lowest MSE.

There also is simulation evidence on the performance of sequential tests for lag-order selection. At least for univariate autoregressions, several studies have shown that the general-to-specific sequential-testing approach is preferred to the specific-to-general sequential testing approach. For VAR models neither approach is satisfactory for selecting the true lag order (see, e.g., Lütkepohl 1985; Nickelsburg 1985; Gonzalo and Pitarakis 2002). Moreover, the simulation evidence in Kilian and Ivanov (2005) shows that sequential testing procedures tend to produce impulse response estimators with systematically larger MSE than information criteria and cannot be recommended for applied work.

Lag-Order Selection in $\text{VAR}(\infty)$ Models. The consistency for p_0 becomes entirely irrelevant once we allow for the possibility that the DGP is a $\text{VAR}(\infty)$ process. In that case, one can reinterpret the $\text{VAR}(p)$ model as an approximation to a $\text{VAR}(\infty)$ data generating process, as discussed in Section 2.3.6.

In such a setting, consistency for p_0 is not a meaningful concept. Instead, the objective of lag-order selection is to select the best approximating model. As a practical matter, the use of information criteria in this setting is subject to the same concerns regarding the finite-sample performance of lag-order selection criteria as in the case of finite lag-order VAR models. While the AIC may have a better asymptotic rationale than more parsimonious criteria in the $\text{VAR}(\infty)$ context, even the AIC underfits in small samples.

There are few optimality results regarding the lag-order choice for sieve autoregressions. It can be shown that the AIC is asymptotically efficient for the one-step ahead MSPE, provided the lower bound and upper bound on the lag order are chosen appropriately. This asymptotic efficiency result was originally obtained under the assumption that the processes used for estimation and for prediction are independent. This assumption is inappropriate for time series analysis, because the value to be predicted depends on past realizations. More recently, Ing and Wei (2003, 2005) established the optimality of the AIC for same-realization prediction from stationary $\text{AR}(\infty)$ processes, and Ing, Sin, and Yu (2010) extended this result to integrated $\text{AR}(\infty)$ processes. Of course, even this result applies only to one-step ahead prediction and is only asymptotic. It does not provide much guidance for applied econometric work based on sieve vector autoregressions.⁶

Moreover, much depends on the choice of the lower bound and upper bound, which also depend on T . Asymptotic theory implies that p has to grow at a suitable rate with the sample size T to preserve the consistency and asymptotic normality of the LS estimator of the approximating $\text{VAR}(p)$ model. This insight does not tell us how to choose p for given T , however. In practice, this question has been addressed by simulating data from VARMA models fitted to actual data and searching for the lag order that provides the best sieve approximation for the statistic of interest (see, e.g., Berkowitz and Kilian 2000; Inoue and Kilian 2002b). These simulation studies suggest that the best approximating lag order tends to be larger than the lag-order choices commonly used in applied work and larger than the lag orders suggested by the AIC.

Implications of Data-Dependent Lag-Order Selection for Inference. An important question usually neglected in discussions of lag-order selection is that statistical inference about the VAR model parameters and inference about smooth functions of these parameters such as impulse responses is affected by the use of data-dependent VAR lag-order selection procedures. It is useful to start with the premise of a $\text{VAR}(p_0)$ data generating process. In that case, one immediate problem is that the use of inconsistent lag-order selection procedures such as the AIC or sequential testing procedures affects the asymptotic

⁶ For a comparison of the asymptotic properties of the AIC, SIC, and HQC and predictive least-squares approaches when the DGP is of infinite order, see Ing (2007).

distribution of the slope parameters conditional on the estimated lag orders. In particular, the post-model selection estimators are no longer asymptotically normal, invalidating the use of the delta method (see, e.g., Leeb and Pötscher 2005). This point is often ignored in applied work.

It may seem that this problem could be avoided by the use of consistent lag-order selection criteria such as the SIC or HQC. The traditional view in the literature for many years has been that consistent lag-order selection procedures allow one to employ the same asymptotic distributions that would be appropriate if the correct $\text{VAR}(p_0)$ model were specified instead of $\text{VAR}(\hat{p})$ (see, e.g., Lütkepohl 1990; Kilian 1998c). This view has recently been shown to be mistaken by Leeb and Pötscher (2005). The reason is that the usual starting point in the literature has been an analysis of the pointwise asymptotic distribution of the slope parameters, where pointwise refers to the thought experiment of holding the true parameter fixed while letting the sample size diverge to infinity. Indeed the pointwise asymptotic distribution coincides with the usual asymptotic distributions derived for correctly specified models. As discussed in Leeb and Pötscher (2005), however, the finite-sample distribution of the slope parameters will not always be well approximated by the pointwise asymptotic normal distribution. Leeb and Pötscher provide examples in which this finite-sample distribution is bimodal rather than Gaussian. The reason is that the convergence of the finite-sample distribution to their pointwise limits is typically not uniform with respect to the underlying parameter values. Leeb and Pötscher demonstrate for a generic linear regression model that it may require arbitrarily large T for the standard pointwise asymptotics to become a good approximation.

This problem arises when the parameter in question is close to zero, but different from zero, and cannot simply be overcome by the use of bootstrap methods either (see Chapter 12). Although bootstrap methods of inference for $\text{VAR}(\hat{p}^{\text{AIC}})$ models appear reasonably accurate in large samples in the simulation studies of Kilian (1998a, 2001), there is no proof of the uniform validity of the bootstrap in this case. In fact, Leeb and Pötscher (2006) conjecture that no bootstrap method can solve the problem of nonuniformity. Furthermore, these problems do not vanish if we drop the fiction of a $\text{VAR}(p_0)$ data generating process. Similar finite-sample problems arise even when the underlying data generating process is of infinite order. Leeb and Pötscher (2005) show that in the latter case even the pointwise asymptotic distributions are affected by the model selection procedure and no longer coincide with the usual pointwise asymptotic distributions that apply in the absence of model selection.

It is important to be clear that Leeb and Pötscher (2005, 2006) are concerned with the worst possible outcome when conducting inference, not the likely outcome. They establish that conventional inference may be unreliable in some cases. They do not establish that it will always be unreliable. In fact,

in many cases, standard inference will be perfectly adequate. It is difficult to determine how much we should be concerned with the worst possible outcome in practice. Without knowing the DGP it is impossible to know how practically relevant the situations described by Leeb and Pötscher are. Existing simulation results for DGPs based on actual data have not revealed any problems with the reliability of standard inference based on pointwise asymptotics (see, e.g., Kilian 1998a, 2001). Nevertheless, in light of the results in Leeb and Pötscher (2005, 2006), the use of all data-dependent lag-order selection procedures for VAR models must be reconsidered. An alternative approach is to estimate VAR models with fixed lag orders instead. For example, fitting a $\text{VAR}(p_{\max})$ model, where p_{\max} is a conservative bound on the range of possible lag orders, allows the use of standard asymptotic approximations for statistics such as impulse responses and circumvents the problems of inference described by Leeb and Pötscher.

Impulse Response Analysis. It is known (a) that the use of the AIC in some cases can be formally justified even when the underlying VAR is of infinite order; (b) that in a number of simulation studies the AIC performed better at selecting the true VAR order than alternative lag-order selection criteria such as sequential tests, the SIC, or the HQC; and (c) that AIC-based impulse response point estimates tend to have lower MSE than impulse response estimates based on other criteria, when working with monthly data (see Kilian and Ivanov 2005). At the same time, the use of the AIC tends to result in more accurate impulse response confidence intervals in monthly models, as demonstrated in Kilian (1998a, 2001). All this evidence does not necessarily mean that users of VAR models in practice should condition on the AIC lag-order estimate, however, because even the AIC has a tendency to underfit in small samples (see, e.g., Kilian 1998a, 2001; Gonzalo and Pitarakis 2002). Hence, in many applied situations, when estimating VAR impulse responses, a reasonable alternative approach is to impose a fixed lag order *ex ante* rather than to rely on lag-order selection criteria. In light of the asymmetric loss associated with overfitting and underfitting the model to be used for impulse response analysis, this lag order ought to be larger rather than smaller in case of doubt. The use of a reasonably large fixed lag order also avoids the problems of inference discussed in Leeb and Pötscher (2005, 2006).

It is useful to elaborate on how this fixed lag order may be chosen in practice. The reasoning underlying this choice is similar to the reasoning underlying the choice of p_{\max} in implementing lag-order selection criteria. The appropriate number of autoregressive lags has nothing to do with the persistence of the data. Indeed, even a $\text{VAR}(1)$ model is fully capable of capturing any degree of persistence in the data. Rather it has to do with the delays in the responses to shocks. One situation in which it is safer to allow for long lags

is when we model slowly building cycles in the data. A case in point are studies of the evolution of commodity prices. Inspection of commodity price data reveals cycles that may last for more than ten years. These cycles build and wane only gradually. A VAR model that truncates the lag order too early will miss these cycles, which are associated with gradual changes in the demand for commodities. This argument prompted Kilian (2009), for example, to allow for 24 monthly lags in a VAR model of the global oil market. It can be shown that the importance of global demand shocks in this model vanishes if the lag order is restricted to, say, 12 lags. The reason is that much of the response to the underlying demand shocks is small initially and accumulates only with a delay. Allowing for enough lags is crucial for detecting the role of demand shocks in this model.

This does not mean that such a large number of lags is always required. The appropriate lag-order choice depends on the economic context and on the number of variables included. The responses of typical U.S. macroeconomic aggregates, for example, can be reliably estimated with a much smaller number of lags. Many empirical VAR studies impose 12 monthly lags or 4 quarterly lags. These choices reflect the premise that much of the response occurs within the first year of a shock. Although lag-order selection criteria could be used to make the model more parsimonious, such criteria have to be viewed with caution, given the proclivity of even the least parsimonious lag order selection criteria to underfit in small samples. Edelstein and Kilian (2009), for example, in a study of the transmission of monthly energy price shocks to the U.S. economy since the 1970s, found that the AIC in some cases produces implausibly low lag-order estimates associated with counterintuitive estimates of the response functions. These problems vanished once a larger lag order of 6 was imposed throughout, and the VAR estimates remained qualitatively robust to using even higher lag orders between 6 and 12.

In practice, it is useful to assess the sensitivity of estimates to reasonable changes in the lag structure. The insistence on using a large number of lags does not mean that information criteria cannot provide useful information. For example, we may interpret the lag-order estimate implied by the AIC as a lower bound on the lag order to be imposed in estimation. The dangers of relying on lag-order selection criteria that are too parsimonious are also illustrated by Hamilton and Herrera (2004) who show that raising the lag order from 7, which is the estimate suggested by the AIC, to 12 or to 16 overturns the substantive findings in Bernanke, Gertler, and Watson (1997) regarding the effect of oil price shocks on U.S. real GDP growth. Specifically, Hamilton and Herrera find large and statistically significant direct effects of oil price shocks for $p = 12$ and $p = 16$ that were not apparent with $p = 7$. This example illustrates the importance of allowing for enough lags. Any empirical result that is reversed with higher lag orders must be considered suspect. On the other hand, results that remain intact with higher lag orders are likely to be trustworthy.

One objection to the strategy of erring on the side of including too many lags may be that allowing for so many lags may cause the impulse response estimates to become statistically insignificant. This is indeed a risk, although the inclusion of more lags need not increase the uncertainty about the impulse response estimates. In practice, the increase in uncertainty in the response estimates in many cases is smaller than one might have conjectured.

The risk of overfitting, moreover, does not mean that we are justified in imposing a lower lag order. Kilian (2001) illustrates by simulation that in this case we are indeed likely to obtain precisely estimated response functions with tight confidence intervals, except that these estimates bear little resemblance to the responses in the DGP. In other words, it is the nature of the economic problem rather than the data constraints faced by the researcher that dictates the appropriate lag order. In some cases, this may mean that a VAR model should not be estimated because the sample is too short to accommodate the required number of lags.

VAR Forecasts. One key difference between impulse response analysis and out-of-sample forecasting is that in the latter case we are not concerned with selecting the DGP, but with selecting the most accurate out-of-sample forecasting model. The usual criterion for measuring out-of-sample forecast accuracy is the MSPE. The well-known bias variance trade-off means that even if we knew the true lag order p_0 , a VAR(p) model with $p < p_0$ may have lower out-of-sample MSPE in small samples. Thus consistency for p_0 is not a valid criterion for choosing between information criteria. Inoue and Kilian (2006) provide extensive discussion of the conditions under which information criteria will be consistent for the forecasting model with the lowest one-step ahead MSPE, even when one or more candidate models are misspecified. They show that the SIC is superior to the AIC in the context of choosing between a VAR(m) model and a VAR($m + 1$) model for one-step ahead prediction. The asymptotic validity of the SIC for selecting the lag order of the VAR model does not depend on the existence of p_0 or the inclusion of p_0 in the set $\{p_{\min}, \dots, p_{\max}\}$. It would apply even if the underlying data generating process were a VAR(∞) model, for example, provided the set of candidate models does not depend on T . It is useful to keep in mind, however, that the results in Inoue and Kilian (2006) are asymptotic in nature. There is no guarantee that the SIC will select the most accurate one-step ahead VAR forecasting model in finite samples.

One of the limitations of standard lag-order selection criteria including the SIC is that they focus on the one-step ahead MSPE in assessing the fit of the model. The latter approach is appropriate if the model is to be used for one-step ahead forecasts, but in practice VAR models often are used for forecasting several steps ahead. For example, we may be concerned with forecasting the inflation rate not for the next month, but for six months from now. Even more

commonly, we will be interested in forecasting variables such as the cumulative inflation rate between today and some future date, based on a monthly VAR model. One option in that case is to employ a direct forecast of $p_{t+h} - p_t$, where h denotes the horizon over which inflation is to be forecast and p_t is the log of the price level. Unless $h = 1$, the construction of this direct forecast requires the forecaster to abandon the VAR framework we have considered so far. The other option is to recursively iterate forward the fitted VAR model, to obtain forecasts of the monthly inflation rate at horizons 1 through h , and to cumulate these forecasts to construct the implied forecast of the inflation rate between now and h periods from now.

It is immediately obvious that an iterated forecast based on a VAR model using lag orders selected by criteria designed to select the true lag order (or designed to select the VAR model with the smallest one-step ahead MSPE) will not necessarily produce optimal MSPEs at longer horizons, unless the VAR model is correctly specified (see Schorfheide 2005). Marcellino, Stock, and Watson (2006) and Pesaran, Pick, and Timmermann (2011) nevertheless show that, in practice, iterated forecasts tend to be more accurate in out-of-sample forecasting than direct forecasts, but only provided the lag order chosen is sufficiently large. Iterated forecasts based on VAR models with lag orders chosen by the SIC, for example, tend to be less accurate at longer horizons than VAR models with lag orders chosen by the AIC. This does not mean that the AIC lag order is optimal, of course.

Ing (2004) considers the problem of choosing the multi-step ahead forecast with the smallest possible MSPE for an autoregressive process of finite but unknown order. He provides several examples that illustrate that optimal multi-step forecasts are not guaranteed even when one is able to correctly identify the underlying autoregressive lag order. In response, Ing proposes an alternative lag-order selection criterion for multi-step ahead prediction that is valid for stationary autoregressions of finite lag order. A generalization of this idea to stationary VAR models was developed in Schorfheide (2005). Further generalizations to autoregressions of possibly infinite order are provided in Ing (2007), and extensions to possibly integrated autoregressions of possibly infinite order are provided in Ing, Lin, and Yu (2009) and Ing, Sin, and Yu (2010).

2.7 Model Diagnostics

There is a large set of tools for checking whether a given VAR model represents the DGP of the variables adequately. They range from informal graphical procedures such as residual plots to formal statistical specification tests of the adequacy of the assumptions underlying the model. In the following subsections selected formal specification tests are discussed. For the underlying theory and for related procedures see, e.g., Lütkepohl (2004, 2005). The presentation in this section partly follows Lütkepohl (2009).

2.7.1 Tests for Autocorrelation in the Innovations

A basic assumption for the VAR model is that the reduced-form innovation process u_t is white noise. In other words, the u_t are assumed to exhibit no serial correlation. As discussed in the previous section, the lag order of the model is typically chosen such that this condition is at least approximately satisfied. For a given VAR model it may still be desirable to check for residual autocorrelation. In that case, the Portmanteau and Breusch-Godfrey LM tests presented in Subsection 2.6.2 can be used.

2.7.2 Tests for Nonnormality

Although normality of the innovations of a VAR model and, hence, normality of the observed variables is not required for the validity of most asymptotic procedures related to VAR modeling, normality can still be a property of interest. For example, it facilitates predictive inference, as discussed earlier. Moreover, knowing that the distribution of the observations is far from normal is useful for assessing possible efficiency gains from using other estimation procedures. For example, there may be asymptotic efficiency gains if the true distribution were used instead of the Gaussian distribution in setting up the likelihood function.

In practice, such efficiency gains are not likely to be substantial in the small samples common in macroeconomic applications. If the residuals do not have a normal distribution, however, this may be a signal of other potential model defects. A rejection of normality, for example, may arise from unusually large residuals of three or even four times the size of the standard deviation, sometimes called outliers. Such outliers could be an indication that the VAR model is misspecified. Thus, a rejection of normality may suggest a closer look at the time series of residuals.

In this context, multivariate nonnormality tests may be applied to the full residual vector of the VAR model and univariate versions can be used for the errors of the individual equations. Lomnicki (1961) and Jarque and Bera (1987) have proposed nonnormality tests for univariate models that can be extended easily to multivariate models. The idea is to check whether the third and fourth moments of the residuals are conformable with those of a normal distribution. To use this approach, the residual vector of a VAR model is first transformed to make the individual components uncorrelated. Then the moments are compared with those of the normal distribution. For given residuals \hat{u}_t , $t = 1, \dots, T$, of an estimated VAR process, the residual covariance matrix $\tilde{\Sigma}_u$ is decomposed such that $\tilde{\Sigma}_u = PP'$, where P is a suitable $K \times K$ matrix. The tests for nonnormality are then based on the skewness and kurtosis of the standardized residuals $\hat{u}_t^s = P^{-1}\hat{u}_t$ (see Lütkepohl 2005).

There are many possible P matrices that decompose $\tilde{\Sigma}_u$, and the tests depend to some extent on the transformation matrix used. Doornik and Hansen

(1994) propose to use the square root matrix of $\tilde{\Sigma}_u$ whereas Lütkepohl (2005, chapter 4) considers a Cholesky decomposition of the residual covariance matrix. Following the latter convention, let

$$\hat{b}_j = (\hat{b}_{1j}, \dots, \hat{b}_{Kj})', \quad \text{where} \quad \hat{b}_{kj} = \frac{1}{T} \sum_{t=1}^T (\hat{u}_{kt}^s)^j$$

for $j = 3, 4$. It can be shown that $\lambda_3 = \hat{b}'_3 \hat{b}_3 / 6 \xrightarrow{d} \chi^2(K)$ can be used to test the symmetry of the error distribution and $\lambda_4 = (\hat{b}_4 - 3_K)'(\hat{b}_4 - 3_K) / 24 \xrightarrow{d} \chi^2(K)$, where $3_K = (3, \dots, 3)'$ is a $K \times 1$ vector, can be used to test for excess kurtosis relative the kurtosis of the Gaussian distribution. Moreover, the sum of the two test statistics is asymptotically χ^2 distributed with $2K$ degrees of freedom and can be viewed as an omnibus test of the null of Gaussianity. In small samples this test may be severely oversized. Small-sample corrections of the critical values based on bootstrap approximations are proposed in Kilian and Demiroglu (2000).

The literature on testing for nonnormality is extensive and many other tests are available. In the context of VAR modeling the tests based on the third and fourth moments are nevertheless the most popular tests. Nonnormality in the unconditional error distribution also arises if the innovations are conditionally heteroskedastic. Arguably more powerful tests for this feature are discussed in the next subsection.

2.7.3 Residual ARCH Tests

Unmodeled conditional heteroskedasticity in the VAR innovations does not invalidate the consistency of standard estimators of the VAR slope parameters, as long as the unconditional error variances remain finite. It undermines the efficiency of the estimator, however, and affects how we conduct inference about the parameters of interest (see Chapter 12). The consistent estimation of the unconditional error covariance matrix requires additional moment conditions, given that the assumption of iid errors is violated in the presence of autoregressive conditionally heteroskedastic (ARCH) errors, for example. Moreover, ARCH dynamics in the innovations may invalidate the assumption of a finite fourth moment required for asymptotic and bootstrap inference on structural impulse responses and related statistics, depending on the parameters of the model of conditional heteroskedasticity (see He and Teräsvirta 1999). Hence, knowing whether the innovations are conditionally heteroskedastic or not is important.

A multivariate ARCH model of order q for the VAR innovations u_t has the form

$$\text{vech}(\Sigma_{t|t-1}) = \delta_0 + D_1 \text{vech}(u_{t-1} u'_{t-1}) + \dots + D_q \text{vech}(u_{t-q} u'_{t-q}),$$

where vech is the column-stacking operator for symmetric matrices that stacks the columns from the main diagonal downward, and $\Sigma_{t|t-1}$ is the conditional covariance matrix of u_t given u_{t-1}, u_{t-2}, \dots . Moreover, δ_0 is a $\frac{1}{2}K(K+1)$ -dimensional parameter vector and the D_j , $j = 1, \dots, q$, are $\frac{1}{2}K(K+1) \times \frac{1}{2}K(K+1)$ coefficient matrices.

A test for ARCH or GARCH dynamics can be based on similar ideas as the LM test for residual autocorrelation by considering the pair of hypotheses

$$\begin{aligned}\mathbb{H}_0 : D_1 &= \dots = D_q = 0 \quad \text{versus} \\ \mathbb{H}_1 : D_i &\neq 0 \text{ for at least one } i \in \{1, \dots, q\}.\end{aligned}$$

Clearly, there is no ARCH in the innovations if \mathbb{H}_0 is true. The relevant LM statistic can be obtained by estimating the auxiliary model

$$\text{vech}(\hat{u}_t \hat{u}'_t) = \delta_0 + D_1 \text{vech}(\hat{u}_{t-1} \hat{u}'_{t-1}) + \dots + D_q \text{vech}(\hat{u}_{t-q} \hat{u}'_{t-q}) + e_t \quad (2.7.1)$$

and computing the statistic

$$LM_{ARCH}(q) = \frac{1}{2}TK(K+1) \left(1 - \frac{2}{K(K+1)} \text{tr}(\widehat{\Omega} \widehat{\Omega}_0^{-1}) \right),$$

where $\widehat{\Omega}$ is the residual covariance matrix of the $\frac{1}{2}K(K+1)$ -dimensional error term e_t of the regression model (2.7.1) with $q > 0$ and $\widehat{\Omega}_0$ is the corresponding matrix for the case $q = 0$. The statistic is similar to the one described by Doornik and Hendry (1997, section 10.9.2.4). It may be used with critical values from a $\chi^2(qK^2(K+1)^2/4)$ distribution, or finite sample approximations to critical values may be obtained by bootstrapping the statistic under the null of iid innovations.

One concern with multivariate tests for conditional heteroskedasticity is their low power in finite samples. An alternative approach is to test for GARCH in each VAR equation individually. The trade-off between these approaches, as measured by size and power, depends on whether there is conditional heteroskedasticity in all or only some of the VAR equations.

2.7.4 Time Invariance

An important assumption underlying standard VAR analysis is the time-invariance of the model. As defined earlier, stationarity requires time-invariant first and second unconditional moments. That assumption is violated not only if the stability condition is not satisfied; but it may also be violated if the parameters change over time. A wide range of procedures for checking the stability

or time-invariance of a given model exists (e.g., Doornik and Hendry 1997; Lütkepohl 2004, 2005, chapter 17). These procedures may be used to detect potential structural breaks during the sample period. The leading example of the class of tests of the null hypothesis of parameter stability is the Chow test, of which different versions have been proposed in the literature.

The central idea is to compare the null hypothesis of time-invariant parameters throughout the sample period against the possibility of a change in the parameter values at some date T_B . One version of the Chow test involves estimating the model on the full sample of T observations as well as on the first T_1 and the last T_2 observations, where $T_1 < T_B$ and $T_2 \leq T - T_B$. A formal LR test may be constructed from the Gaussian likelihood by comparing the maximum of the likelihood in the constant-parameter model to the maximum obtained after allowing for different parameter values before and after period T_B , possibly after dropping some observations around the break date.

Denoting the conditional log-density of the t^{th} observation vector by $l_t \equiv \log f_t(y_t | y_{t-1}, \dots, y_{-p+1})$, the test statistic can be written as

$$\begin{aligned} LR_{\text{Chow}} = 2 \left[& \sup \left(\sum_{t=1}^{T_1} l_t \right) + \sup \left(\sum_{t=T-T_2+1}^T l_t \right) \\ & - \sup \left(\sum_{t=1}^{T_1} l_t + \sum_{t=T-T_2+1}^T l_t \right) \right]. \end{aligned} \quad (2.7.2)$$

Under the null hypothesis of time-invariant parameters, the statistic has an asymptotic χ^2 distribution with degrees of freedom given by the number of restrictions implied by the assumption of a constant-coefficient model for the full sample period. In other words, the degrees of freedom are the difference between the sum of the number of free coefficients estimated in the first and last subperiods and the number of free coefficients in the full-sample model. It must be emphasized, however, that the distribution may be different if the process contains integrated components, even if the process is time invariant (see Lütkepohl 2005, section 8.4.3; Hansen 2003). For an investigation of the properties of Chow tests for VAR models, the reader is referred to Candelon and Lütkepohl (2001).

One practical question is how many observations to drop in between the two subsamples. Asymptotic arguments allow one to choose $T_1 = T_B - 1$ and $T_2 = T - T_B + 1$ without dropping any observations. If the parameter change is smooth or its exact timing is unknown, however, leaving out some observations may improve the small-sample power of the test.

Another version of a Chow test for multivariate time series models considers predictions from a model fitted to the first $T_B - 1$ observations and evaluates whether they are in line with the observed data (see Doornik and Hendry

1997). The latter authors also propose versions of these tests based on the F test statistic with the aim of improving the size of the test in small samples.

Candelon and Lütkepohl (2001) point out that, in particular for multivariate time series models, the asymptotic χ^2 distribution may be a poor approximation in small samples. Even F critical values may fail to correct the small-sample size distortions. Candelon and Lütkepohl therefore propose the use of bootstrap versions of Chow tests in order to improve their small-sample properties.

Chow-type tests can be generalized in various directions. For example, one can test for more than one break or one can assess the constancy of a subset of parameters keeping the remaining parameters fixed. It is also common to construct critical values for the null of a break at T_B under the premise that the test was performed repeatedly for a range of potential break points T_B . In this case, the critical values are derived for the test statistic $\sup_{T_B \in \mathcal{T}} LR_{\text{Chow}}$, where $\mathcal{T} \subset \{1, \dots, T\}$ is the set of periods for which the test statistic is computed. In practice, it is necessary to trim some observations at the beginning and the end of the sample when defining the range of possible break dates to ensure that the test has power. The sup-test statistic does not have an asymptotic χ^2 distribution (see Andrews 1993; Andrews and Ploberger 1994; Hansen 1997a). Its asymptotic distribution depends on the fraction of the sample over which a search is performed. Critical values are given by Andrews (1993) for a range of situations. They can be derived analytically in some cases, but more commonly they are derived by bootstrap methods under the null hypothesis of no break (see Christiano 1992; Diebold and Chen 1996). The critical values obtained from this distribution guard against the danger of data mining across possible break points by considering only break dates that appear favorable to the rejection of the null of no break based on preliminary inspection of the data. Using conventional critical values in this case would result in spurious evidence of a structural break. Only if the break date can be pinned down uniquely based on non-sample information will conventional critical values for the Chow test apply.

Other tests for structural change are based on recursive residuals obtained by fitting models to samples of increasing length. In other words, starting from some sample size T_0 , a recursive residual $\hat{u}_\tau^{(r)}$ is the last residual obtained by fitting a model to a sample for $t = 1, \dots, \tau$ for $\tau \in \{T_0, \dots, T\}$. Cumulated recursive residuals or squared residuals are then used for testing for structural change. These tests are known as CUSUM and CUSUM-of-squares tests and were originally proposed by Brown, Durbin, and Evans (1975) for testing structural change in linear regression models. Krämer, Ploberger, and Alt (1988) establish their validity in dynamic models. These tests are designed to test the structural stability of a model if the specific change point is not known. An application of this class of tests in the VAR context can be found in Lütkepohl (2004).

It should be noted that tests for structural change are prone to rejecting the null of no break in small samples, even when the null is true, whenever there are large transitory dynamics in the DGP. In the latter case, large transitory dynamics and permanent breaks tend to be observationally equivalent, and caution is called for in interpreting the results of these tests.

In this chapter the point of departure has been a VAR model with time-invariant parameters. Given this premise, it makes sense to treat the constant-parameter model as the null hypothesis. An alternative approach would be to start the specification search from a more general model class that allows for time-varying slope coefficients. Such models are considered in Chapter 18. For example, one may model the evolution of the model coefficients as a random walk process. Under the null hypothesis that the innovation variance of this random walk process is zero, the model coefficients are time-invariant. Suitable tests of this hypothesis have been discussed in Nyblom (1989) and Teräsvirta, Tjøstheim, and Granger (2010, section 6.4.4), among others.

2.8 Subset VAR Models, AVAR Models, and VARX Models

VAR models allow for unrestricted feedback between the model variables for a given number of lags. The motivation for this approach is that we are typically unable to derive exclusion restrictions on the reduced-form lag structure from economic theory. Such restrictions simply are not economically credible (see Sims 1980a). The advantage of unrestricted VAR model estimates is that they are more robust to misspecification than estimates from dynamic simultaneous equation models imposing more exclusion restrictions. The disadvantage of unrestricted VAR model estimates is that they tend to be less precise and hence may not be informative about the questions of interest. Almost from the inception of the VAR approach, therefore, researchers have tried to explore the use of statistical model selection criteria to reduce the perceived parameter profligacy of $\text{VAR}(p)$ models and to improve the efficiency of VAR model estimates.

2.8.1 *Subset VAR Models*

The most common proposal is to allow for different lag structures in different equations of the model, rather than imposing the same lag order on all VAR model equations (see, e.g., Hsiao 1979, 1981). The resulting subset VAR models differ from the standard VAR models considered so far in that the regressors differ across equations. Subset VAR models may be estimated using equation-by-equation LS, but that estimator will not be efficient. Alternatives include (possibly iterated) feasible GLS estimation methods or constrained full information ML estimation methods, as discussed in Section 2.3.

One situation in which imposing different lag orders on the VAR equations may be plausible is the following example. Consider a bivariate VAR(1) model for real GDP growth and stock returns. Whereas real GDP growth exhibits some sluggishness, requiring the inclusion of at least one lag, stock returns are well approximated by a white noise process, suggesting that zero lags of the model variables in the second VAR equation may be adequate in practice. This argument is not based on a priori economic theory, but on the time series properties of the data. Rather than imposing this restriction *ex ante*, one may use statistical tests to determine whether to include the lagged VAR coefficients in the second equation. One testing procedure for determining subset VAR models is based on the t -ratios of individual model parameters. We sequentially eliminate the variables with the lowest t -ratios until all remaining variables have t -ratios greater than some threshold value, say 1.96. A more common approach is to evaluate information criteria for all plausible combinations of lag orders in the VAR equations, which allows us to compare all models at the same time. Compared with conventional lag-order selection in VAR(p) models, subset VAR model selection is more computationally demanding because there is no natural ordering between alternative subset models and the number of permutations to be considered may be as high as $2^{K^2 p}$. Moreover, the asymptotic theory motivating the use of information criteria for lag-order selection is derived under the premise that the number of candidate models is small relative to T (see, e.g., Inoue and Kilian 2006). One would not expect these criteria to be reliable when the number of subset models considered is large.

Brüggemann and Lütkepohl (2001) compare alternative strategies of selecting subset VAR models in a simulation study. Although none of these strategies was found to be reliable when it comes to detecting the DGP, using subset VAR models in some cases may result in models with improved forecast precision or with tighter impulse response confidence intervals than in the unrestricted VAR model. Nevertheless, subset VAR models have not played an important role in empirical research. One concern, already alluded to, is that pretests of this type may undermine subsequent inference about impulse responses or related statistics of interest. In addition, examples in which one would expect the appropriate lag order to differ a lot across equations are not common in empirical macroeconomics.

2.8.2 Asymmetric VAR Models

Keating (1993) introduces the closely related idea of asymmetric VAR (AVAR) models, in which the lag length differs across variables such that the lag length is the same for each variable in all equations, but may differ across variables.⁷

⁷ The notion of asymmetry here is not related to the notion of asymmetric responses in Chapter 18, but merely to the lag structure.

This contrasts with subset VAR models in which the lag lengths of a given variable may differ across equations. There is no systematic evidence that asymmetric VAR models improve the accuracy of impulse response estimates or forecasts.

2.8.3 *VARX Models*

A third example of a restricted VAR model with different lag structures across equations is the VARX model (e.g., Lütkepohl 2005, chapter 10). VARX models are VAR models in which one or more variables are exogenous with respect to the remaining variables (see Chapter 7). This implies that there is no feedback from lagged endogenous variables to the exogenous variables in the system of reduced-form equations, allowing us to restrict these lag coefficients to zero. The equations for the exogenous variables include the same number of lags as the other equations but only lags of the exogenous variables. The reduced-form representation of the VARX model effectively imposes Granger noncausality from endogenous to exogenous variables. A VARX model may be estimated as a system of equations by restricted LS or ML methods (see Sections 2.3.3 and 2.3.4). Alternatively, the model may be estimated without separate equations for the exogenous variables.

The use of exogenous variables is not common in the VAR literature, but there are exceptions. For example, it has been argued that ocean temperatures are exogenous with respect to the global business cycle at least over the horizons considered in business cycle analysis. Similarly, one could make the case that the hours of sunlight per day in New York are exogenous with respect to the Dow Jones stock price index. Finding examples of exogenous economic variables is even harder. Until a few years ago, the price of crude oil was considered exogenous with respect to the U.S. economy. A number of recent studies, however, has shown that there is an important endogenous element in this price series (see Kilian 2008a). A better example is a small open economy that faces exogenous variation in world interest rates or in its terms of trade (see, e.g., Cushman and Zha 1997). Another application of VARX models involves models that include extraneous estimates of exogenous monetary and fiscal policy shocks. These shocks may be treated as exogenous variables that are subject to further exclusion restrictions on their own dynamics (see Chapters 6, 13, and 15).

3 Vector Error Correction Models

Many economic variables exhibit persistent upward or downward movement. This feature can be generated by stochastic trends in integrated variables. If the same stochastic trend is driving a set of integrated variables jointly, they are called cointegrated. In this case, certain linear combinations of integrated variables are stationary. Such linear combinations that link the variables to a common trend path are called cointegrating relationships. They sometimes may be interpreted as equilibrium relationships in economic models.

Cointegrating relationships can be imposed by reparameterizing the VAR model as a vector error correction model (VECM).¹ In Section 3.1 cointegrated variables are introduced and VECMs are set up. Sections 3.2 and 3.3 consider the estimation as well as the specification of VECMs. Diagnostic tools are presented in Section 3.4, and the implications of cointegrated variables in VAR models for forecasting and Granger causality analysis are discussed in Section 3.5. Our focus in this chapter is on reduced-form models. We leave extensions to structural VECMs to later chapters.

The concept of cointegration was introduced in the econometrics literature by Granger (1981) and Engle and Granger (1987). Early work on error correction models goes back to Sargan (1964), Davidson, Hendry, Srba, and Yeo (1978), Hendry and von Ungern-Sternberg (1981), and Salmon (1982). Lütkepohl (1982b) discusses the cointegration feature without using the cointegration terminology. A full analysis of the VECM is presented in Johansen (1995), among others. Parts of the present chapter follow closely Lütkepohl (2005, part II; 2006, 2009).

3.1 Cointegrated Variables and Vector Error Correction Models

3.1.1 Common Trends and Cointegration

Cointegrated processes were introduced by Granger (1981) and Engle and Granger (1987). If two integrated variables share a common stochastic trend

¹ Some researchers also refer to these models as vector equilibrium correction models.

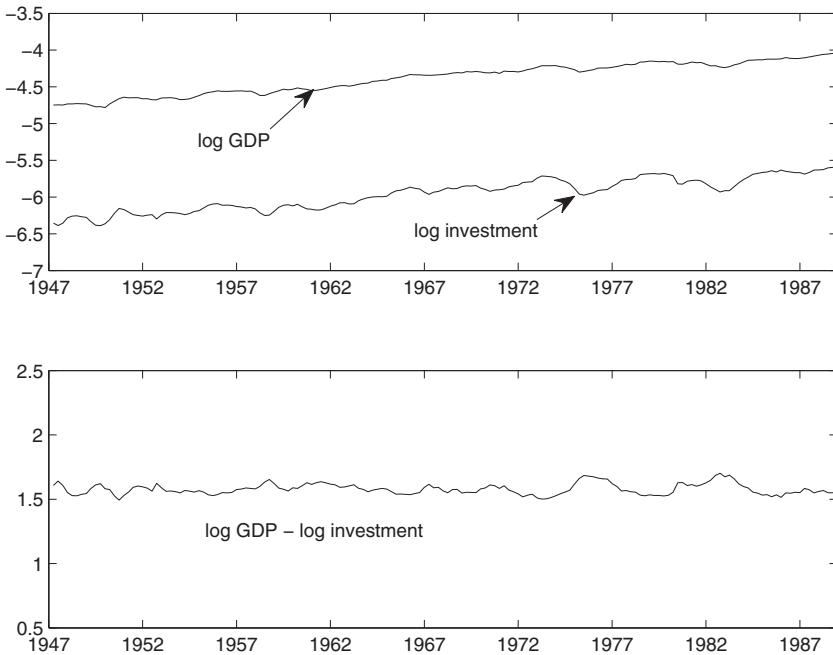


Figure 3.1. Logs of quarterly U.S. real GDP and real investment for 1947q1–1988q4.

such that a linear combination of these variables is stationary, they are called cointegrated. For example, the plots of quarterly U.S. log output and investment in the upper panel of Figure 3.1 both exhibit an upward trend. Because both series are driven by the same trend, the log of the GDP-investment ratio is fluctuating about a constant mean. As a result, the difference between the log series in the lower panel of Figure 3.1 has no obvious trend anymore. It is mean reverting and appears stationary.

The concept of cointegration may also be applied to linear combinations of more than two $I(1)$ variables. More formally, we say that a set of $I(1)$ time series variables is cointegrated if there exists a linear combination of these variables that is $I(0)$. Generalizing this concept to higher orders of integration, the variables in a K -dimensional process y_t are cointegrated if the components are $I(d)$ and there exists a linear combination $z_t = \beta'y_t$ with $\beta = (\beta_1, \dots, \beta_K)' \neq 0$ such that z_t is $I(d^*)$ with $d^* < d$. The vector β is called a cointegrating vector or a cointegration vector. For example, let $z_t = p_t - p_t^* - e_t \sim I(0)$, where $p_t \sim I(1)$ denotes the log of the domestic consumer price index, $p_t^* \sim I(1)$ denotes the log of the foreign consumer price index, and $e_t \sim I(1)$ denotes the log of the nominal exchange rate expressed in domestic currency values per unit of foreign currency. Then $\beta = (1, -1, -1)'$ is a cointegrating vector. This

relationship embodies the view that under standard arbitrage conditions the real exchange rate must be $I(0)$, even when its components are not.

A cointegrating vector is not unique. In the real exchange rate example, multiplying β by any nonzero constant would result in another equally valid cointegrating vector. As a matter of convention, we typically normalize the coefficients of β such that one of the elements of β is 1. Note that the remaining values of the cointegrating vector in general do not have to be restricted to integers.

More generally, it is convenient to call a K -dimensional $I(d)$ process y_t cointegrated if there is a linear combination $\beta'y_t$ with $\beta \neq 0$ that is integrated of order less than d . Notice that this definition differs slightly from that given by Engle and Granger (1987) in that it also covers the case when the components of y_t have no common trend. For instance, if $y_t = (y_{1t}, y_{2t})'$ is a bivariate process such that $y_{1t} \sim I(1)$ and $y_{2t} \sim I(0)$, then the bivariate process y_t is $I(1)$ and has to be differenced once to make it stationary. However, the linear combination

$$(0, 1) \begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = y_{2t}$$

is $I(0)$ and, hence, the process is cointegrated according to our slightly more general definition, although there is no common trend and, hence, there is no genuine cointegration in the original sense.

The VEC Model. In a system of variables, there may be several linearly independent cointegrating vectors. In that case linear combinations of these vectors are also cointegrating vectors because linear combinations of stationary variables are stationary. To embed the concept of cointegration in the VAR framework, suppose for the moment that all individual variables are $I(1)$ or $I(0)$ and the DGP is a K -dimensional $\text{VAR}(p)$ process,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad (3.1.1)$$

without deterministic terms. Subtracting y_{t-1} on both sides of the equation and rearranging terms yields the VECM

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \quad (3.1.2)$$

where

$$\Pi = -(I_K - A_1 - \cdots - A_p)$$

and

$$\Gamma_i = -(A_{i+1} + \cdots + A_p), \quad i = 1, \dots, p-1.$$

Among the regressors in (3.1.2) the only nonstationary variable is y_{t-1} . Since the left-hand side of equation (3.1.2) is $I(0)$, so must be the right-hand side, which requires Πy_{t-1} to be $I(0)$.

Because the variables have unit roots individually,

$$\det(I_K - A_1 z - \cdots - A_p z^p) = 0$$

for $z = 1$, and, thus, the matrix Π is singular. Suppose this matrix has rank r . Then there are r linearly independent cointegrating relationships. Hence, the rank of Π is called the cointegration rank or cointegrating rank of the process y_t . Observe that any $K \times K$ matrix of rank r can be decomposed as a product of two $K \times r$ matrices of full column rank. Let α and β be two $K \times r$ matrices of rank r such that $\Pi = \alpha\beta'$. Since (1) any linear transformation of stationary variables is stationary, since (2) Πy_t is stationary, and since (3) $(\alpha'\alpha)^{-1}\alpha'\Pi y_t$ is a linear transformation of Πy_t , the latter linear transformation, which equals $\beta'y_t$ after substituting $\Pi = \alpha\beta'$, is also stationary, and, hence, the rows of β' are cointegration vectors. The matrix β' is therefore called the cointegrating matrix, and the matrix α is sometimes referred to as the loading matrix. Substituting the matrix $\alpha\beta'$ for Π in (3.1.2) yields

$$\Delta y_t = \alpha\beta'y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \quad (3.1.3)$$

which is called a VECM because it explicitly includes the lagged error correction (EC) term $\alpha\beta'y_{t-1}$.

Two special cases of this model merit further discussion. One special case arises when $r = K$, in which case the process is already stable in levels. All variables are $I(0)$ in levels, and there is no need to consider a VECM. The other special case arises when $r = 0$. In that case, the EC term is zero and Δy_t has a stable VAR($p - 1$) representation in differences.

If the point of departure is the VECM (3.1.2), the corresponding levels VAR representation can be recovered easily by noting that

$$\begin{aligned} A_1 &= \Pi + I_K + \Gamma_1, \\ A_i &= \Gamma_i - \Gamma_{i-1}, \quad i = 2, \dots, p-1, \\ A_p &= -\Gamma_{p-1}. \end{aligned} \quad (3.1.4)$$

As mentioned earlier, cointegration relationships are not unique. This fact is reflected in the nonuniqueness of the decomposition of the $K \times K$ matrix $\Pi = \alpha\beta'$. Any nonsingular $r \times r$ matrix Q gives rise to a decomposition $\Pi = \alpha^*\beta^{*\prime}$, where $\alpha^* = \alpha Q'$ and $\beta^* = \beta Q^{-1}$. A convenient normalisation relies on the fact that β can always be chosen as

$$\beta = \begin{bmatrix} I_r \\ \beta_{(K-r)} \end{bmatrix}, \quad (3.1.5)$$

where $\beta_{(K-r)}$ is $(K-r) \times r$, possibly after the variables have been rearranged suitably. For example, if the cointegrating rank is 1, all cointegrating relationships are multiples of a single relationship, and the normalization in (3.1.5) implies that this relationship can be written as a single equation with one variable on the left-hand side and the others on the right-hand side. Using $\beta = (1, \beta'_{(K-1)})'$ with $\beta'_{(K-1)} = (\beta_{(K-1),1}, \dots, \beta_{(K-1),K-1})'$, yields

$$y_{1t} = -\beta_{(K-r),1}y_{2t} - \dots - \beta_{(K-r),K-1}y_{Kt} + z_t,$$

where $z_t \sim I(0)$. Usually we think of cointegration relationships as linear combinations of $I(1)$ variables. If some of the elements of y_t are $I(0)$ in levels, there is an additional cointegrating relationship for each stationary component of y_t . Because these cointegrating vectors are linearly independent columns of β , the cointegrating rank must be at least as large as the number of $I(0)$ variables in the system.

The Granger Representation of the VECM. Granger's representation theorem, as stated by Johansen (1995, Theorem 4.2), is another useful representation of cointegrated processes. It may be stated using the orthogonal complements of matrices. For $m \geq n$, an orthogonal complement of the $m \times n$ matrix M with $\text{rank}(M) = n$ is denoted by M_\perp . Put differently, M_\perp is any $m \times (m-n)$ matrix with $\text{rank}(M_\perp) = m-n$ and $M'M_\perp = 0$. Note that the $m \times m$ matrix $[M, M_\perp]$ is nonsingular. If M is a nonsingular square matrix ($m = n$), then $M_\perp = 0$, and, if $M = 0$, then $M_\perp = I_m$.

Let y_t be a K -dimensional cointegrated $I(1)$ process as in (3.1.3) with cointegrating rank r , $0 \leq r < K$. Then it can be shown that

$$y_t = \Xi \sum_{i=1}^t u_i + \Xi^*(L)u_t + y_0^*, \quad (3.1.6)$$

where

$$\Xi = \beta_\perp \left[\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right]^{-1} \alpha'_\perp, \quad (3.1.7)$$

$\Xi^*(L)u_t = \sum_{j=0}^\infty \Xi_j^* u_{t-j}$ is an $I(0)$ process and y_0^* contains the initial values.

The representation in (3.1.6) is a multivariate version of the Beveridge-Nelson decomposition of y_t discussed in Chapter 2 and is known as the Granger representation of the process. It is also known as the common-trends representation. Equation (3.1.6) decomposes the process y_t into $I(1)$ and $I(0)$ components. The term $\sum_{i=1}^t u_i$ is a K -dimensional random walk. However, the matrix Ξ has rank $K-r$. For Ξ to be properly defined in (3.1.7), the

$(K - r) \times (K - r)$ matrix

$$\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp$$

must be invertible. Hence, $\text{rank}(\boldsymbol{\Xi}) = K - r$ and, thus, the term $\boldsymbol{\Xi} \sum_{i=1}^t u_i$ on the right-hand side of (3.1.6) effectively consists of a $(K - r)$ -dimensional random walk or common trends component. Consequently, y_t is driven by $K - r$ common trends.

3.1.2 Deterministic Terms in Cointegrated Processes

Deterministic terms complicate the specification of integrated processes. For example, a constant term in a random walk process generates a linear trend in the mean, as seen in Chapter 2. This linear trend is distinct from the stochastic trend implied by the random walk component. Therefore, it is necessary to pay special attention to the implications of deterministic terms in cointegrated processes. Suppose that

$$y_t = \mu_t + x_t, \quad (3.1.8)$$

where x_t is a K -dimensional zero mean $\text{VAR}(p)$ process with possibly cointegrated variables and μ_t is a $K \times 1$ deterministic term. In practice, μ_t often includes a constant and possibly an additional deterministic time trend. In other words, $\mu_t = \mu_0$ or $\mu_t = \mu_0 + \mu_1 t$, where μ_0 and μ_1 are fixed K -dimensional parameter vectors.

The VECM with Intercept. The DGP of x_t is assumed to be

$$\begin{aligned} \Delta x_t &= \alpha \beta' x_{t-1} + \Gamma_1 \Delta x_{t-1} + \cdots + \Gamma_{p-1} \Delta x_{t-p+1} + u_t \\ &= \Pi x_{t-1} + \Gamma_1 \Delta x_{t-1} + \cdots + \Gamma_{p-1} \Delta x_{t-p+1} + u_t. \end{aligned} \quad (3.1.9)$$

For $\mu_t = \mu_0$, we have $x_t = y_t - \mu_0$ such that $\Delta y_t = \Delta x_t$ and, hence,

$$\begin{aligned} \Delta y_t &= \alpha \beta' (y_{t-1} - \mu_0) + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \\ &= \alpha \beta'^o \begin{pmatrix} y_{t-1} \\ 1 \end{pmatrix} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t \\ &= \Pi^o y_{t-1}^o + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \end{aligned} \quad (3.1.10)$$

where $\beta'^o = [\beta', \delta']$ with $\delta' = -\beta' \mu_0$ being an $r \times 1$ vector,

$$y_{t-1}^o = \begin{pmatrix} y_{t-1} \\ 1 \end{pmatrix}$$

and $\Pi^o = [\Pi, \vartheta]$ is a $K \times (K + 1)$ matrix with $\vartheta = -\Pi \mu_0 = \alpha \delta'$. Thus, a constant mean term in the additive representation (3.1.8) becomes an intercept

term in the cointegration relationship. In the implied VECM for Δy_t ,

$$\begin{aligned}\Delta y_t &= v_0 + \alpha\beta'y_{t-1} + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t \\ &= v_0 + \Pi y_{t-1} + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t,\end{aligned}\quad (3.1.11)$$

the intercept v_0 has to satisfy the restrictions $v_0 = \alpha\delta'$. As a result, the intercept can be absorbed into the cointegration relationship. This fact ensures that the intercept does not generate a linear time trend in the mean of the y_t variables, consistent with our assumption that none of the model variables exhibits a linear time trend.

The VECM with Intercept and Trend. If $\mu_t = \mu_0 + \mu_1 t$ is a linear trend, we have $x_t = y_t - \mu_0 - \mu_1 t$ and $\Delta x_t = \Delta y_t - \mu_1$. Thus, substituting in (3.1.9) yields

$$\begin{aligned}\Delta y_t - \mu_1 &= \alpha\beta'(y_{t-1} - \mu_0 - \mu_1(t-1)) + \Gamma_1(\Delta y_{t-1} - \mu_1) + \cdots \\ &\quad + \Gamma_{p-1}(\Delta y_{t-p+1} - \mu_1) + u_t.\end{aligned}\quad (3.1.12)$$

Rearranging the deterministic terms we have

$$\begin{aligned}\Delta y_t &= v + \alpha[\beta', \eta'] \binom{y_{t-1}}{t-1} + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t \\ &= v + \Pi^+ y_{t-1}^+ + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t,\end{aligned}\quad (3.1.13)$$

with $v = -\Pi\mu_0 + (I_K - \Gamma_1 - \cdots - \Gamma_{p-1})\mu_1$, $\eta' = -\beta'\mu_1$, $\Pi^+ = \alpha[\beta', \eta']$ is a $K \times (K+1)$ matrix and

$$y_{t-1}^+ = \binom{y_{t-1}}{t-1}.$$

In this representation the intercept term v is left unrestricted, whereas the linear trend term can be absorbed into the cointegration relationships.

Finally, it can be shown that in the model

$$\Delta y_t = v_0 + \nu_1 t + \Pi y_{t-1} + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t,$$

with unrestricted linear trend term, $I(1)$ variables actually generate quadratic deterministic trends in y_t .

It is also possible that the trend slope parameter μ_1 is orthogonal to the cointegration matrix such that $\beta'\mu_1 = 0$. In that case, $\eta = 0$, and there is no linear trend term in the cointegrating relationships, although the individual variables have linear trends. The model

$$\begin{aligned}\Delta y_t &= v + \alpha\beta'y_{t-1} + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t \\ &= v + \Pi y_{t-1} + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t,\end{aligned}\quad (3.1.14)$$

with unrestricted intercept term ν is popular in applied work because a trend in the cointegrating relationships is sometimes regarded as implausible. Note that cointegrating relationships occasionally may be interpreted as equilibrium relationships. In that case, it is particularly implausible that the variables are driven apart by a deterministic trend. It is worth emphasizing, however, that, when there is no trend in the cointegrating relationships, the cointegration rank must be smaller than K in order to obtain a linear trend in the variables. If $r = K$, the process is stationary and, hence, a constant term in the model does not generate a linear trend.

Finally it is worth noting that the additive form of the drifting process in (3.1.8) facilitates the derivation of a common trends representation of y_t processes with deterministic terms. This representation may be obtained from the Granger representation (3.1.6) of x_t by adding the deterministic term, resulting in

$$y_t = \Xi \sum_{i=1}^t u_i + \Xi^*(L)u_t + y_0^* + \mu_t, \quad (3.1.15)$$

where all symbols are defined as in expression (3.1.6) and μ_t is the deterministic term defined in (3.1.8).

3.2 Estimation of VARs with Integrated Variables

VECMs are a convenient reparameterization of VAR models in levels. Clearly, from the estimator of the VECM parameters one can derive an estimator of the parameters of the corresponding VAR in levels. Alternatively, one can estimate the VAR model in levels directly, as already mentioned in the previous chapter. In this subsection, we discuss alternative estimation approaches for VECMs and compare their properties.

3.2.1 The VAR(1) Case

It is instructive to start with a simple VAR(1) model without any deterministic terms. Detailed derivations for this case are provided in Lütkepohl (2005, section 7.1). Here we just summarize and discuss the results. Consider the model

$$y_t = A_1 y_{t-1} + u_t, \quad (3.2.1)$$

where u_t is iid white noise with nonsingular covariance matrix Σ_u , i.e., $u_t \stackrel{iid}{\sim} (0, \Sigma_u)$. The corresponding VECM is

$$\Delta y_t = \Pi y_{t-1} + u_t = \alpha \beta' y_{t-1} + u_t, \quad (3.2.2)$$

where $\Pi = A_1 - I_K$ and $\text{rank}(\Pi) = \text{rank}(\alpha) = \text{rank}(\beta) = r$.

LS Estimation. If the cointegrating rank r is unknown and the A_1 matrix, or equivalently Π , is estimated by LS, we have

$$\widehat{A}_1 - A_1 = \widehat{\Pi} - \Pi = \left(\sum_{t=1}^T u_t y'_{t-1} \right) \left(\sum_{t=1}^T y_{t-1} y'_{t-1} \right)^{-1}.$$

The asymptotic distribution of the estimators depends on the cointegrating rank. For $r = 0$, the process consists of K random walks, and the asymptotic distribution is

$$T(\widehat{A}_1 - A_1) \xrightarrow{d} \Sigma_u^{1/2} \left\{ \int_0^1 \mathbf{W}_K d\mathbf{W}'_K \right\}' \left\{ \int_0^1 \mathbf{W}_K \mathbf{W}'_K ds \right\}^{-1} \Sigma_u^{-1/2},$$

where \mathbf{W}_K denotes a K -dimensional standard Brownian motion and $\Sigma_u^{1/2}$ is the square-root matrix of Σ_u .² Note that we have written the asymptotic distribution in matrix form, which is different from our previous notation which expressed multivariate distributions in vector form. This change of notation is of no consequence. The important point is that in this model a well-defined asymptotic distribution is obtained upon standardizing the estimator by T rather than \sqrt{T} . In other words, the estimator converges at a faster rate than in the stationary case. Moreover, its asymptotic distribution is not Gaussian. For univariate AR processes this situation is well known from the literature on Dickey-Fuller tests for unit roots. Our analysis in this subsection generalizes this result to the multivariate case.

If the cointegrating rank $r > 0$ and the cointegrating matrix β is normalized such that

$$\beta = \begin{bmatrix} I_r \\ \beta_{(K-r)} \end{bmatrix},$$

only the elements of the $(K - r) \times r$ matrix $\beta_{(K-r)}$ are unknown. These elements can be estimated consistently. Given the normalization,

$$\Pi = [\alpha, \alpha \beta'_{(K-r)}],$$

where the matrix Π is constructed by concatenating the matrices α and $\alpha \beta'_{(K-r)}$, one can express model (3.2.2) as

$$\Delta y_t - \alpha y_{t-1}^{(1)} = \alpha \beta'_{(K-r)} y_{t-1}^{(2)} + u_t = (y_{t-1}^{(2)'} \otimes \alpha) \text{vec}(\beta'_{(K-r)}) + u_t, \quad (3.2.3)$$

where $y_{t-1}^{(1)}$ and $y_{t-1}^{(2)}$ consist of the first r and the last $K - r$ elements of y_{t-1} , respectively.

² A Brownian motion is a continuous-time stochastic process with independent Gaussian increments that is commonly used to characterize limiting distributions of statistics that depend on integrated processes. For a formal definition of Brownian motions see Hamilton (1994).

GLS Estimation. Model (3.2.3) is a multivariate regression model with potentially different regressors in different equations. Therefore, GLS estimation may be more efficient than LS estimation. The GLS estimator of $\beta'_{(K-r)}$ can be shown to be

$$\begin{aligned}\widehat{\beta}'_{(K-r)} &= (\alpha' \Sigma_u^{-1} \alpha)^{-1} \alpha' \Sigma_u^{-1} \\ &\times \left(\sum_{t=1}^T (\Delta y_t - \alpha y_{t-1}^{(1)}) y_{t-1}^{(2)\prime} \right) \left(\sum_{t=1}^T y_{t-1}^{(2)} y_{t-1}^{(2)\prime} \right)^{-1}. \quad (3.2.4)\end{aligned}$$

Note that this GLS estimator does not reduce to the LS estimator because the regressors differ across equations. Since α and Σ_u are unknown in practice, these quantities have to be replaced by estimates. Consistent estimators of α and Σ_u can be obtained from LS estimation of model (3.2.2). Given our normalization of β , the first r columns of the LS estimator $\widehat{\Pi}$ are a consistent estimator of α , and Σ_u can be estimated in the usual way from the LS residuals.

Using these estimators in the expression for $\widehat{\beta}'_{(K-r)}$ in (3.2.4), we obtain a feasible GLS estimator that converges at rate T . In fact,

$$T(\widehat{\beta}'_{(K-r)} - \beta'_{(K-r)})$$

has a mixed normal asymptotic distribution that can be used for constructing valid asymptotic tests for the coefficients.

Using these estimators of α and β , we can estimate Π as $\widehat{\Pi} = [\widehat{\alpha}, \widehat{\alpha} \widehat{\beta}'_{(K-r)}]$, which allows the construction of the estimator $\widehat{A}_1 = \widehat{\Pi} + I_K$ of A_1 . The convergence rate of this estimator is \sqrt{T} , and its asymptotic distribution is Gaussian. It can be shown that

$$\sqrt{T} \text{vec}(\widehat{A}_1 - A_1) = \sqrt{T} \text{vec}(\widehat{\Pi} - \Pi) \xrightarrow{d} \mathcal{N}(0, \beta \Gamma^{-1} \beta' \otimes \Sigma_u), \quad (3.2.5)$$

where $\Gamma \equiv \text{plim } T^{-1} \sum_{t=1}^T \beta' y_{t-1} y_{t-1}' \beta$ is a nonsingular $r \times r$ matrix.

LS Estimation with Known Cointegrating Matrix. In fact, the same asymptotic distribution is obtained when the cointegrating matrix β is known and only α is estimated by LS from the model

$$\Delta y_t = \alpha \beta' y_{t-1} + u_t.$$

Denoting the resulting estimator by $\widehat{\alpha}$, where

$$\widehat{\alpha} = \left(\sum_{t=1}^T \Delta y_t y_{t-1}' \beta \right) \left(\sum_{t=1}^T \beta' y_{t-1} y_{t-1}' \beta \right)^{-1},$$

the corresponding estimator $\widehat{A}_1 = \widehat{\alpha}\beta' + I_K$ has precisely the asymptotic distribution given in expression (3.2.5). Thus, knowing the cointegrating rank r or the cointegrating matrix β does not improve the asymptotic efficiency of the LS estimator of A_1 or Π . This result is a direct consequence of the faster convergence rate of the estimator of the cointegrating parameters. In fact, the same asymptotic distribution is obtained if the cointegrating rank is unknown and A_1 or Π are estimated by LS without accounting for the cointegrating rank. Thus, in a model with $r > 0$, knowing the true cointegrating rank is no advantage in estimation, as long as only asymptotic results are of interest. This purely asymptotic argument, however, ignores that not using all available information about the cointegrating structure of the model variables reduces the accuracy of the estimator in small samples.

The possible presence of cointegrated variables in the VAR model also may complicate statistical inference. An important feature of the asymptotic distribution in expression (3.2.5) is the singularity of the covariance matrix if $r < K$. This feature implies that standard inference will not be valid in general. For example, confidence intervals and t -tests for the coefficients may be misleading when they are based on the usual normal asymptotics. Clearly, the matrix $\beta\Gamma^{-1}\beta'$ depends on the cointegrating structure of the variables because it depends on the cointegrating matrix β . Suppose, for example, that the cointegrating rank is $r = 1$ and that y_t consists of the two components y_{1t} and y_{2t} such that $\beta' = (\beta_1, \beta_2)$. The null hypothesis for testing Granger causality from y_{1t} to y_{2t} , for example, is

$$\mathbb{H}_0 : a_{21,1} = 0.$$

This hypothesis can be tested with the t -ratio of $a_{21,1}$ obtained from (3.2.5) if the corresponding variance in the asymptotic covariance matrix is nonzero. The latter condition, however, requires that $\beta_1 \neq 0$. If both components of y_t are $I(1)$, there is no problem, and the t -ratio can be safely used for testing \mathbb{H}_0 because the cointegration relationship must necessarily involve both variables, and, thus, β_1 and β_2 are both nonzero. This fact was emphasized in Lütkepohl and Reimers (1992b). If, however, y_{2t} happens to be $I(0)$, then $\beta_1 = 0$, and a t -test based on standard normal critical values is not valid. This example shows that knowing the true cointegration structure (or at least some aspects of this structure such as the order of integration of both variables) can be important for valid inference in cointegrated VAR models. If the cointegration structure is not known, one option is to determine the cointegration properties of the data by statistical procedures. The other option is to estimate the model in levels, as discussed in Section 3.2.3.

FIML Estimation of Gaussian Processes with Known Cointegrating Rank. For a given rank r , it is also possible to estimate the VECM (3.2.2) without

the standardization of the cointegration matrix β in (3.1.5). In that case, one estimates β by the reduced-rank (RR) regression or canonical correlation procedure of Johansen (1988). Johansen proposes minimizing the determinant

$$\det \left(T^{-1} \sum_{t=1}^T (\Delta y_t - \Pi y_{t-1})(\Delta y_t - \Pi y_{t-1})' \right)$$

subject to the rank restriction $\text{rank}(\Pi) = r$ or, equivalently, to minimize the determinant

$$\det \left(T^{-1} \sum_{t=1}^T (\Delta y_t - \alpha \beta' y_{t-1})(\Delta y_t - \alpha \beta' y_{t-1})' \right)$$

with respect to the $K \times r$ matrices α and β . The solution to this problem is based on the ordered eigenvalues $\lambda_1 \geq \dots \geq \lambda_K$ and associated orthonormal eigenvectors η_1, \dots, η_K of the matrix

$$\begin{aligned} & \left(\sum_{t=1}^T y_{t-1} y_{t-1}' \right)^{-1/2} \left(\sum_{t=1}^T y_{t-1} \Delta y_t' \right) \left(\sum_{t=1}^T \Delta y_t \Delta y_t' \right)^{-1} \\ & \times \left(\sum_{t=1}^T \Delta y_t y_{t-1}' \right) \left(\sum_{t=1}^T y_{t-1} y_{t-1}' \right)^{-1/2}. \end{aligned}$$

The resulting estimators are

$$\tilde{\beta}' = [\eta_1, \dots, \eta_r]' \left(\sum_{t=1}^T y_{t-1} y_{t-1}' \right)^{-1/2}$$

and

$$\tilde{\alpha} = \left(\sum_{t=1}^T \Delta y_t y_{t-1}' \tilde{\beta} \right) \left(\sum_{t=1}^T \tilde{\beta}' y_{t-1} y_{t-1}' \tilde{\beta} \right)^{-1}.$$

The procedure is equivalent to maximizing the log-likelihood function for a model with Gaussian residuals, $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_u)$. The estimators $\tilde{\alpha}$ and $\tilde{\beta}$ are not consistent because they are not separately identified. However, the corresponding estimator $\tilde{\Pi} = \tilde{\alpha} \tilde{\beta}'$ for Π is consistent and has the same asymptotic distribution as the LS estimator in expression (3.2.5).

3.2.2 Estimation of VECMs

Now suppose that the cointegrating rank r is greater than zero and consider a more general VECM without deterministic terms, but of lag order $p > 1$ such

that

$$\begin{aligned}\Delta y_t &= \alpha\beta'y_{t-1} + \Gamma_1\Delta y_{t-1} + \cdots + \Gamma_{p-1}\Delta y_{t-p+1} + u_t \\ &= \alpha\beta'y_{t-1} + \Gamma X_{t-1} + u_t,\end{aligned}\quad (3.2.6)$$

where

$$\Gamma = [\Gamma_1, \dots, \Gamma_{p-1}], \quad \text{and} \quad X_{t-1} = \begin{pmatrix} \Delta y_{t-1} \\ \vdots \\ \Delta y_{t-p+1} \end{pmatrix}.$$

The estimation of this model involves three steps. It is useful to begin by concentrating out Γ , which allows us to focus on the problem of estimating $\alpha\beta'$. Given $\alpha\beta'$, the LS estimator of Γ is known to be

$$\widehat{\Gamma}(\alpha\beta') = \left(\sum_{t=1}^T (\Delta y_t - \alpha\beta'y_{t-1}) X'_{t-1} \right) \left(\sum_{t=1}^T X_{t-1} X'_{t-1} \right)^{-1}. \quad (3.2.7)$$

Replacing Γ in (3.2.6) by this estimator yields the concentrated model

$$R_{0t} = \alpha\beta'R_{1t} + u_t^*, \quad (3.2.8)$$

where

$$R_{0t} = \Delta y_t - \left(\sum_{t=1}^T \Delta y_t X'_{t-1} \right) \left(\sum_{t=1}^T X_{t-1} X'_{t-1} \right)^{-1} X_{t-1}$$

and

$$R_{1t} = y_{t-1} - \left(\sum_{t=1}^T y_{t-1} X'_{t-1} \right) \left(\sum_{t=1}^T X_{t-1} X'_{t-1} \right)^{-1} X_{t-1}.$$

It is not difficult to see that R_{0t} and R_{1t} are just the residuals from regressing Δy_t and y_{t-1} , respectively, on X_{t-1} . In the second step, ML and GLS methods are used to estimate α and β from the concentrated model (3.2.8). In the last step, the estimator of Γ is obtained by substituting the estimators of α and β into expression (3.2.7).

ML Estimation for Gaussian Processes. The parameters α and β can be estimated by RR regression as proposed by Johansen (1988) (see also Anderson 1951). As in the VAR(1) case considered earlier, this estimation method is

equivalent to ML estimation if the process is Gaussian. The concentrated log-likelihood function is

$$\begin{aligned}\log l = & -\frac{KT}{2} \log(2\pi) - \frac{T}{2} \log(\det(\Sigma_u)) \\ & - \frac{1}{2} \text{tr} \left[\sum_{t=1}^T (R_{0t} - \alpha\beta'R_{1t})' \Sigma_u^{-1} (R_{0t} - \alpha\beta'R_{1t}) \right]\end{aligned}\quad (3.2.9)$$

and the RR or ML estimators are

$$\tilde{\beta}' = [\eta_1, \dots, \eta_r]' S_{11}^{-1/2} \quad \text{and} \quad \tilde{\alpha} = S_{01} \tilde{\beta} (\tilde{\beta}' S_{11} \tilde{\beta})^{-1}, \quad (3.2.10)$$

where tr denotes the trace of a matrix, $S_{ij} = \sum_{t=1}^T R_{it} R_{jt}' / T$, $i = 0, 1$, and η_1, \dots, η_K are the orthonormal eigenvectors of the matrix $S_{11}^{-1/2} S_{10} S_{00}^{-1} S_{01} S_{11}^{-1/2}$ corresponding to its eigenvalues in nonincreasing order, $\lambda_1 \geq \dots \geq \lambda_K$.

Implementation of the Johansen procedure does not require consistent estimation of the cointegrating matrix and hence does not require any normalization of the cointegrating matrix. Such normalizations are not required for many applications of VECMs such as impulse response analysis. They are only required if we are specifically interested in the cointegrating vectors. If β is normalized such that $\beta' = [I_r, \beta'_{(K-r)}]$ as in (3.1.5) and the RR/ML estimator is normalized accordingly, then, under general conditions, $T \text{vec}(\tilde{\beta}'_{(K-r)} - \beta'_{(K-r)})$ converges in distribution to a Gaussian mixture distribution (Johansen 1995 or Lütkepohl 2005, chapter 7). Asymptotic mixed normality of the cointegration parameter estimator means that inference can be conducted similarly to asymptotically normal estimators.

The ML estimator of α given in (3.2.10) is the LS estimator obtained from the multivariate regression model

$$R_{0t} = \alpha \tilde{\beta}' R_{1t} + u_t^*$$

with regressor matrix $\tilde{\beta}' R_{1t}$. If $\tilde{\beta}$ is replaced with a superconsistent estimator, the asymptotic properties of the corresponding $\tilde{\alpha}$ estimator are standard and so are those of

$$\tilde{\Gamma}(\tilde{\alpha} \tilde{\beta}') = \left(\sum_{t=1}^T (\Delta y_t - \tilde{\alpha} \tilde{\beta}' y_{t-1}) X'_{t-1} \right) \left(\sum_{t=1}^T X_{t-1} X'_{t-1} \right)^{-1}. \quad (3.2.11)$$

Under general conditions, these estimators converge at the usual \sqrt{T} rate to an asymptotic normal distribution. It can be shown that

$$\sqrt{T} \text{vec}([\tilde{\alpha}, \tilde{\Gamma}] - [\alpha, \Gamma]) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\alpha, \Gamma}),$$

where

$$\Sigma_{\alpha, \Gamma} = \text{plim } T \begin{bmatrix} \beta' \sum_{t=1}^T y_{t-1} y_{t-1}' \beta & \beta' \sum_{t=1}^T y_{t-1} X_{t-1}' \\ \sum_{t=1}^T X_{t-1} y_{t-1}' \beta & \sum_{t=1}^T X_{t-1} X_{t-1}' \end{bmatrix}^{-1} \otimes \Sigma_u.$$

We illustrate the Johansen approach to estimating VECMs based on the same trivariate example already used in the context of the estimation of stationary VAR models. Recall that $y_t = (\Delta gnp_t, i_t, \Delta p_t)'$. We treat i_t and Δp_t as individually $I(1)$, but cointegrated. For expository purposes we also assume that the log of U.S. real GNP is not cointegrated with any of the other model variables. This implies that the cointegrating rank is 2 with $\Delta gnp_t \sim I(0)$ and, hence, trivially cointegrated with itself. The model includes an intercept as in (3.1.14). As before, we impose a lag order of $p = 4$, implying the existence of three augmented lags in the VECM representation. The ML estimates are

$$\begin{aligned} \tilde{\nu} &= \begin{bmatrix} 0.7520 \\ -0.3416 \\ -0.0564 \end{bmatrix}, \\ \tilde{\alpha} &= \begin{bmatrix} -0.2617 & 0.1086 \\ 0.2470 & 0.1187 \\ 0.0353 & 0.0153 \end{bmatrix}, \quad \tilde{\beta}' = \begin{bmatrix} 2.1932 & -0.0594 & 0.8530 \\ -0.3637 & -0.4489 & 1.4588 \end{bmatrix}, \\ \tilde{\Gamma}_1 &= \begin{bmatrix} -0.1637 & 0.0450 & 0.4505 \\ -0.1841 & 0.1687 & 0.1938 \\ -0.0701 & 0.0678 & -0.6178 \end{bmatrix}, \\ \tilde{\Gamma}_2 &= \begin{bmatrix} 0.0492 & -0.3409 & 0.5782 \\ 0.0000 & -0.3168 & 0.6822 \\ -0.0844 & 0.0162 & -0.3638 \end{bmatrix}, \\ \tilde{\Gamma}_3 &= \begin{bmatrix} 0.0421 & 0.0011 & 0.0352 \\ 0.0244 & 0.1686 & 0.3473 \\ -0.0690 & 0.0082 & -0.2652 \end{bmatrix}, \end{aligned}$$

and

$$\tilde{\Sigma}_u = \begin{bmatrix} 0.5659 & 0.0751 & -0.0207 \\ 0.0751 & 0.6165 & 0.0341 \\ -0.0207 & 0.0341 & 0.0654 \end{bmatrix}.$$

Since the three model variables do not have a deterministic linear trend component, one may alternatively absorb the intercept into the error correction

term, as in model (3.1.10). Estimating the model with this restriction of the intercept term, we obtain the ML estimates

$$\begin{aligned}\tilde{\alpha} &= \begin{bmatrix} 0.2617 & 0.1089 \\ -0.2471 & 0.1192 \\ -0.0352 & 0.0149 \end{bmatrix}, \\ \tilde{\beta}' &= \begin{bmatrix} -2.1930 & 0.0595 & -0.8534 & 2.1854 \\ -0.3628 & -0.4490 & 1.4726 & 1.6272 \end{bmatrix}, \\ \tilde{\Gamma}_1 &= \begin{bmatrix} -0.1639 & 0.0450 & 0.4491 \\ -0.1844 & 0.1688 & 0.1918 \\ -0.0701 & 0.0676 & -0.6171 \end{bmatrix}, \\ \tilde{\Gamma}_2 &= \begin{bmatrix} 0.0490 & -0.3409 & 0.5772 \\ -0.0002 & -0.3168 & 0.6807 \\ -0.0843 & 0.0161 & -0.3632 \end{bmatrix}, \\ \tilde{\Gamma}_3 &= \begin{bmatrix} 0.0420 & 0.0010 & 0.0347 \\ 0.0242 & 0.1686 & 0.3466 \\ -0.0688 & 0.0081 & -0.2650 \end{bmatrix},\end{aligned}$$

and

$$\tilde{\Sigma}_u = \begin{bmatrix} 0.5658 & 0.0750 & -0.0206 \\ 0.0750 & 0.6165 & 0.0341 \\ -0.0206 & 0.0341 & 0.0655 \end{bmatrix}.$$

Note that these estimates are very similar to those obtained with an unrestricted intercept.

ML Estimation of Gaussian Processes with Known Cointegrating Vectors. A common situation in applied work is that the cointegrating vectors are known. In that case, it makes sense to impose these cointegrating vectors in estimation. For example, the expectations theory of the term structure implies that the spread of interest rates for bonds of different maturities is stationary if the risk premium is stationary. Hence, in a system of two interest rates r_{1t} and r_{2t} , for example, there is a known cointegrating relationship of the form $r_{1t} - r_{2t} = (1, -1)(r_{1t}, r_{2t})'$, allowing ML estimation of α and Γ to condition on $\beta = (1, -1)'$. Similarly, if some nominal variable and the inflation rate are both $I(1)$, whereas the corresponding real variable is $I(0)$, then the nominal variable and inflation are cointegrated with a known cointegration vector. It must be kept in mind, however, that cointegration relations that are implied by economic theory may not be present in the data. Therefore, it is not advisable to blindly impose cointegration vectors without confirming that they are consistent with the data.

Returning to the earlier empirical example, one could postulate that GNP growth is stationary and, hence, constitutes a trivial cointegration relationship, and that the real interest rate is stationary such that $i_t - 4\Delta p_t$ is another cointegration relation, where the factor of 4 arises because the interest rate series is annualized, whereas the inflation rate series is not. In that case, the known cointegration matrix is

$$\beta' = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & -4 \end{bmatrix}$$

which can be used in the ML estimation procedure for ν , α , and Γ_i , $i = 1, 2, 3$. The estimates are

$$\tilde{\nu} = \begin{bmatrix} 0.5156 \\ -0.1828 \\ -0.0593 \end{bmatrix},$$

$$\tilde{\alpha} = \begin{bmatrix} -0.5304 & -0.0355 \\ 0.4271 & -0.0690 \\ 0.0870 & -0.0067 \end{bmatrix},$$

$$\tilde{\Gamma}_1 = \begin{bmatrix} -0.2310 & 0.0259 & 0.3128 \\ -0.1281 & 0.1853 & 0.2636 \\ -0.0808 & 0.0641 & -0.5988 \end{bmatrix},$$

$$\tilde{\Gamma}_2 = \begin{bmatrix} 0.0053 & -0.3587 & 0.4930 \\ 0.0364 & -0.3023 & 0.7228 \\ -0.0912 & 0.0136 & -0.3498 \end{bmatrix},$$

$$\tilde{\Gamma}_3 = \begin{bmatrix} 0.0221 & -0.0214 & -0.0082 \\ 0.0417 & 0.1844 & 0.3672 \\ -0.0727 & 0.0073 & -0.2574 \end{bmatrix},$$

and

$$\tilde{\Sigma}_u = \begin{bmatrix} 0.5722 & 0.0680 & -0.0180 \\ 0.0680 & 0.6223 & 0.0330 \\ -0.0180 & 0.0330 & 0.0648 \end{bmatrix}.$$

Note that the Johansen approach uses a purely statistical standardization of the estimates for α and β . Therefore, the ML estimate of β differs substantially from the cointegration matrix we conditioned upon earlier, and so does the estimate of α obtained under the assumption of known and unknown β .

Feasible GLS Estimation. In small samples, the ML estimator occasionally produces estimates far from the true parameter values, as demonstrated in Brüggemann and Lütkepohl (2005), for example. This problem may be alleviated by considering a more robust GLS estimator. Like in the VAR(1) case,

using the normalization $\beta' = [I_r, \beta'_{(K-r)}]$ given in (3.1.5), equation (3.2.8) can be rewritten as

$$R_{0t} - \alpha R_{1t}^{(1)} = \alpha \beta'_{(K-r)} R_{1t}^{(2)} + u_t^*, \quad (3.2.12)$$

where $R_{1t}^{(1)}$ and $R_{1t}^{(2)}$ denote the first r and last $K - r$ components of R_{1t} , respectively. For a given α , the GLS estimator of $\beta'_{(K-r)}$ based on this specification is

$$\widehat{\beta}'_{(K-r)} = (\alpha' \Sigma_u^{-1} \alpha)^{-1} \alpha' \Sigma_u^{-1} \sum_{t=1}^T (R_{0t} - \alpha R_{1t}^{(1)}) R_{1t}^{(2)\prime} \left(\sum_{t=1}^T R_{1t}^{(2)} R_{1t}^{(2)\prime} \right)^{-1} \quad (3.2.13)$$

(see Lütkepohl 2005, chapter 7). A feasible GLS estimator can be obtained by estimating the matrix $\boldsymbol{\Pi}$ from $R_{0t} = \boldsymbol{\Pi} R_{1t} + u_t^*$ with unrestricted equation-by-equation LS, where $\boldsymbol{\Pi} = [\alpha : \alpha \beta'_{(K-r)}]$. Thus, the first r columns of the estimator for $\boldsymbol{\Pi}$ can be used as an estimator for α , say $\widehat{\alpha}$, as in the VAR(1) case considered in Section 3.2.1. Substituting this estimator and the corresponding estimator of the white noise covariance matrix, $\widetilde{\Sigma}_u = T^{-1} \sum_{t=1}^T \widehat{u}_t^* \widehat{u}_t^{*\prime}$, in expression (3.2.13) yields the feasible GLS estimator

$$\widehat{\beta}'_{(K-r)} = (\widehat{\alpha}' \widetilde{\Sigma}_u^{-1} \widehat{\alpha})^{-1} \widehat{\alpha}' \widetilde{\Sigma}_u^{-1} \sum_{t=1}^T (R_{0t} - \widehat{\alpha} R_{1t}^{(1)}) R_{1t}^{(2)\prime} \left(\sum_{t=1}^T R_{1t}^{(2)} R_{1t}^{(2)\prime} \right)^{-1}. \quad (3.2.14)$$

This estimator was proposed earlier by Ahn and Reinsel (1990) and Saikkonen (1992) (see also Reinsel 1993, chapter 6). It has the same asymptotic distribution as the ML estimator. Likewise, the asymptotic properties of the associated estimators of α and $\boldsymbol{\Gamma}$ are the same as in the previous section.

The GLS estimator has two advantages that make it worthwhile to consider this estimator in macroeconomic applications. First, it tends to be much more reliable than the ML estimator in small samples, as illustrated in Brüggemann and Lütkepohl (2005). Second, the GLS estimator can also be adjusted easily to account for conditional heteroskedasticity (see Herwartz and Lütkepohl 2011). Simulation evidence shows reductions in the mean squared error and the mean absolute error of the estimator by more than 30% after allowing for GARCH in the VECM errors.

An alternative approach to estimating cointegration relations was proposed by Engle and Granger (1987) in the early literature on cointegration. Engle and Granger (1987) observed that if there is a single cointegrating relationship between the $I(1)$ variables y_{1t}, \dots, y_{Kt} , the cointegrating relationship can be estimated consistently by regressing y_{1t} on y_{2t}, \dots, y_{Kt} . The GLS method just outlined may be viewed as a generalization of this procedure which allows for serial correlation and possible conditional heteroskedasticity in estimating

the cointegrating parameters. Because the GLS procedure takes into account the full system of equations, it can also be used when there is more than one cointegrating relationship.

Returning to the previously used empirical example, we now illustrate how to estimate the VECM by the feasible GLS estimation method. We first estimate α and Σ_u from $R_{0t} = \Pi R_{1t} + u_t^*$ using equation-by-equation LS. The estimates are

$$\hat{\alpha} = \begin{bmatrix} -0.6092 & -0.0344 \\ 0.5064 & -0.0701 \\ 0.0620 & -0.0063 \end{bmatrix},$$

and

$$\tilde{\Sigma}_u = \begin{bmatrix} 0.5656 & 0.0746 & -0.0201 \\ 0.0746 & 0.6157 & 0.0351 \\ -0.0201 & 0.0351 & 0.0642 \end{bmatrix}.$$

Then we use expression (3.2.14) to compute the estimate of $\beta'_{(K-r)}$,

$$\hat{\beta}'_{(K-r)} = \begin{bmatrix} 0.2970 \\ -3.8369 \end{bmatrix}, \quad \text{so that } \hat{\beta}' = \begin{bmatrix} 1 & 0 & 0.2970 \\ 0 & 1 & -3.8369 \end{bmatrix}.$$

Finally, we compute the estimate of $\Gamma = [\nu, \Gamma_1, \dots, \Gamma_{p-1}]$ as $\hat{\Gamma} = \Gamma(\hat{\alpha}\hat{\beta}')$ such that

$$\hat{\nu} = \begin{bmatrix} 0.7413 \\ -0.3670 \\ -0.0549 \end{bmatrix},$$

$$\hat{\Gamma}_1 = \begin{bmatrix} -0.1674 & 0.0445 & 0.4387 \\ -0.1911 & 0.1676 & 0.1679 \\ -0.0621 & 0.0685 & -0.6076 \end{bmatrix},$$

$$\hat{\Gamma}_2 = \begin{bmatrix} 0.0465 & -0.3414 & 0.5701 \\ -0.0048 & -0.3180 & 0.6647 \\ -0.0785 & 0.0165 & -0.3558 \end{bmatrix},$$

$$\hat{\Gamma}_3 = \begin{bmatrix} 0.0408 & 0.0000 & 0.0310 \\ 0.0221 & 0.1663 & 0.3381 \\ -0.0656 & 0.0087 & -0.2611 \end{bmatrix},$$

and

$$\hat{\Sigma}_u = \begin{bmatrix} 0.5657 & 0.0748 & -0.0205 \\ 0.0748 & 0.6159 & 0.0345 \\ -0.0205 & 0.0345 & 0.0656 \end{bmatrix},$$

where the latter estimate is computed as

$$\widehat{\Sigma}_u = \frac{1}{T} \sum_{t=1}^T \widehat{u}_t \widehat{u}_t'$$

based on the GLS residuals, \widehat{u}_t . This error covariance matrix estimate differs slightly from the first-stage estimate $\widehat{\Sigma}_u$.

Estimation with Additional Linear Restrictions on the VECM. If restrictions are imposed on the parameters of the VECM representation, the previously discussed estimation methods may be asymptotically inefficient. An alternative is the use of restricted GLS methods which are easy to implement as long as no overidentifying restrictions are imposed on the cointegration matrix. Zero restrictions are the most common constraints for the parameters of these models. Consider, for example, a bivariate system $y_t = (y_{1t}, y_{2t})'$ of two $I(1)$ variables that are cointegrated such that $\beta' y_t \sim I(0)$. If the cointegrating relation appears only in the first equation of the VECM, the loading vector $\alpha = (\alpha_1, 0)'$ has a zero element. Such a restriction can be taken into account in estimation.

If there are no restrictions on the cointegration matrix, the cointegration parameters may be estimated in the first stage by the previously discussed ML or GLS procedures, ignoring exclusion restrictions on the other parameters. Denote this estimator by $\widehat{\beta}$. In the second stage, the remaining parameters may then be estimated from

$$\Delta y_t = \alpha \widehat{\beta}' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t. \quad (3.2.15)$$

Conditional on $\widehat{\beta}$, this is a linear system, and feasible GLS estimation may be used just like for a restricted VAR in levels. For example, if there are linear restrictions of the form

$$\text{vec}[\alpha, \Gamma] = R\gamma, \quad (3.2.16)$$

then we can rewrite (3.2.15) as

$$\Delta y_t = \text{vec} \left[[\alpha, \Gamma] \begin{pmatrix} \widehat{\beta}' y_{t-1} \\ X_{t-1}' \end{pmatrix} \right] + u_t = [(y_{t-1}' \widehat{\beta}, X_{t-1}') \otimes I_K] R\gamma + u_t.$$

The parameter vector γ can be estimated by feasible GLS. The restricted estimator of α and Γ then is easily obtained from relationship (3.2.16). If a super-consistent estimator $\widehat{\beta}$ is used for the cointegration matrix, the asymptotic properties of the estimators are the same as if β were known.

There are also many situations in which economic theory suggests specific cointegration parameters. As mentioned earlier, the spread of two interest rates of different maturities may be stationary, even if the interest rates are $I(1)$. Knowledge of cointegrating parameters allows us to restrict elements of the cointegrating matrix β . If there are restrictions on β , one may use nonlinear

optimization algorithms to obtain ML estimators of all parameters simultaneously, provided the β parameters are identified. It is also possible to use a two-step procedure that estimates the restricted β matrix first and in a second step conditions on that estimator. When the entire matrix β is known, that matrix can be used directly in the second stage, of course. Technical estimation problems from restrictions on β obviously arise only if some of the elements of β remain unrestricted. Restricted estimation of β is analyzed, for example, in Johansen (1995), Boswijk and Doornik (2004), and Lütkepohl (2005, chapter 7).

Finally, it is worth noting that including deterministic terms in the ML or GLS estimation procedures is straightforward. All that is required is to include these terms in the list of regressors or the cointegration term as appropriate.

3.2.3 Estimation of Levels VAR Models with Integrated Variables

Recovering the VAR Model in Levels. Using the mapping (3.1.4), the parameters of the levels VAR model corresponding to a VECM can be estimated by substituting any of the VECM estimators discussed in the previous subsection. Denoting the resulting estimator of the levels parameters $\alpha = \text{vec}[A_1, \dots, A_p]$ by $\hat{\alpha}_{\text{VECM}}$, it can be shown that

$$\sqrt{T}(\hat{\alpha}_{\text{VECM}} - \alpha) \xrightarrow{d} \mathcal{N}(0, \Sigma_\alpha), \quad (3.2.17)$$

if the cointegrating rank $r > 0$ or the lag order $p > 1$. Although this asymptotic result looks like a standard convergence result, there is one important caveat. In this case, the covariance matrix Σ_α is in general singular. In fact, it is the same covariance matrix that one would obtain if the cointegration matrix β were given and the α and Γ were estimated conditional on β . Since $\hat{\alpha}$ is an estimator that satisfies all restrictions implied by the VECM, the singularity of the asymptotic distribution is no surprise. As already discussed in Section 3.2.1, one implication of this result is that some of the conventional statistics for inference on the parameters do not have standard asymptotic distributions. In particular, Wald statistics for linear parameter restrictions may not have their usual asymptotic χ^2 distributions under the null hypothesis. In fact, if the VAR order is $p = 1$, even t -statistics may not be asymptotically standard normal anymore, as shown in Section 3.2.1.

Estimation of the VAR Model in Levels. In practice, the cointegration structure is often unknown. At best it can be estimated and is subject to estimation uncertainty. In that case, an alternative approach is to estimate the VAR in levels without imposing cointegration restrictions. As in the VAR(1) model considered in Section 3.2.1, the LS estimator of this model, denoted by $\hat{\alpha}_{\text{LS}}$, has

exactly the same asymptotic distribution as the $\widehat{\alpha}_{\text{VECM}}$ estimator in (3.2.17),

$$\sqrt{T}(\widehat{\alpha}_{\text{LS}} - \boldsymbol{\alpha}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\boldsymbol{\alpha}}), \quad (3.2.18)$$

even if the cointegration restrictions are not imposed in estimation. As before, the reason is that the cointegration parameters and, hence, the cointegrating relationships are estimated superconsistently. Also, as before, the common trends in some components of the process for y_t induce a singularity in the asymptotic covariance matrix. Not knowing the precise structure of $\Sigma_{\boldsymbol{\alpha}}$ is a problem in general because it implies that the distributions of some test statistics are unknown.

As discussed in Chapter 2, a possible remedy for this problem is to add a redundant lag to the VAR in levels and to estimate a VAR($p + 1$) model instead of a VAR(p) model. If the highest order of integration of the variables is $I(1)$, then this lag augmentation approach ensures that the parameter matrices associated with the first p lags have a nonsingular asymptotic distribution. Thus, tests for hypotheses involving only parameters from the first p slope parameter matrices retain their standard asymptotic properties. Because we know that A_{p+1} is zero by construction, there is no need to test hypotheses about that term.

More generally, if some of the components are $I(d)$ and none of the component series has a higher order of integration, fitting a VAR($p + d$) solves the singularity problem for the LS estimator of the parameters associated with the first p lags (Toda and Yamamoto 1995; Dolado and Lütkepohl 1996). Hence, lag augmentation or overfitting can be used more generally to overcome problems with asymptotic distributions. Of course, using this device has a cost in terms of lower estimation precision.

Sieve Estimation. As discussed in Chapter 2, fitting a finite-order VAR model can be justified as an approximation to an infinite-order VAR process. The asymptotic theory justifying such sieve approximations assumes that the lag order increases with the sample size at a suitable rate. The same device can be used for VECMs. Using the framework of Saikkonen (1992), Saikkonen and Lütkepohl (1996) state conditions that ensure the same asymptotic properties of the GLS estimator for the cointegrating matrix β as in the finite-order case. They also show that under their conditions, the VECM parameters, after suitable rescaling, converge to a Gaussian marginal limiting distribution, as in the case of a true finite-order VECM. Thus, even if the finite-order VECM used for a particular system of variables is just an approximation, standard methods of estimation and inference for VECM parameters remain valid in this case.

It should be noted that Saikkonen and Lütkepohl's results are derived for processes without deterministic terms. These results can be extended to the

case of an intercept term capturing a nonzero mean of the process. Extensions for other deterministic terms appear to be nontrivial.

Moreover, all these results presume that the cointegrating vectors have been correctly specified and that the variables are either $I(0)$ or $I(1)$. If the model variables are not well approximated by either the $I(0)$ or $I(1)$ assumption, yet another option is to appeal to local-to-unity asymptotics (also known as near unit root asymptotics) or to treat the variables as fractionally integrated.

Local-to-Unity Asymptotics. In general, small-sample distributions of estimators of VAR models in levels may not be well approximated by their asymptotic counterparts. For VAR(1) models, the small-sample distortions were found to be particularly troublesome. In response to this problem, Stock (1996), Phillips (1998), and Pesavento and Rossi (2006), among others, have considered alternative asymptotic approximations based on models in which some roots of the autoregressive polynomial are close to 1 but not exactly equal to 1. We discuss that approach next.

A possibly integrated or cointegrated $\text{VAR}(p)$ model without deterministic terms can alternatively be written as

$$y_t = Cy_{t-1} + \boldsymbol{\Gamma}_1\Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1}\Delta y_{t-p+1} + u_t, \quad (3.2.19)$$

where

$$C = I_K + \alpha\beta' = I_K + \boldsymbol{\Pi} = \sum_{i=1}^p A_i.$$

Thus, if the cointegrating rank is zero, $C = I_K$ and, more generally, if $\boldsymbol{\Pi}$ is close to zero, the $K \times K$ matrix C is close to an identity matrix.

Local-to-unity asymptotics are motivated by the observation that in small samples one cannot reliably discriminate between the hypothesis that C equals I_K and the hypothesis that it does not. Following Stock (1996), Phillips (1998), and Pesavento and Rossi (2006), among others, this situation may be modeled by postulating that $C = I_K + \Lambda/T$ in population where Λ is a $K \times K$ diagonal matrix with fixed negative elements $\lambda_1, \lambda_2, \dots, \lambda_K$ along the main diagonal. For any finite T , the diagonal elements of C are smaller than 1. This setup differs from the standard asymptotic thought experiment, in which the parameter matrix C is fixed as $T \rightarrow \infty$. Instead, C is treated as local-to-unity or near unity in the sense that C becomes arbitrarily close to I_K , as $T \rightarrow \infty$.

Modeling one or more roots in the vector autoregressive lag polynomial as local to unity is only a statistical device for obtaining better approximations to the finite-sample distribution of the estimator of interest. It does not mean that we believe that the autoregressive roots of the DGP actually depend on the observed sample size. The fact that we allow the coefficient matrix C to depend on the sample size ensures that it does not become easier with increasing T to

tell the difference between the diagonal elements of C being exactly unity or below unity, because the target to be estimated shifts closer to I_K at the same rate as the estimator of C approaches the target. In other words, we describe a situation in which it is not possible to tell whether the roots are unity or not, even asymptotically.³

A key difference from conventional asymptotics is that, in the local-to-unity framework, C cannot be estimated consistently and that the asymptotic distribution of many estimators of interest in applied work is no longer normal, but nonstandard. Methods of constructing confidence intervals based on such nonstandard asymptotic approximations are discussed in Chapter 12. Whether local-to-unity asymptotics generate more accurate approximations to the finite-sample distribution of an estimator than do conventional asymptotic thought experiments is an empirical question. The fact that an observed time series is highly persistent does not automatically imply that it must be modeled using the local-to-unity framework. As we have seen, inference based on the level VAR(p) model (or based on the first p coefficients of a VAR($p + 1$) model) may also be considered and indeed may be more robust to possible (near) cointegration of unknown form among the model variables. Ultimately, the choice between these asymptotic approximations is determined by their finite-sample accuracy.

Fractional Integration. An alternative approach to modeling persistence in the variables that may not be captured well by unit roots, is to treat the variables as fractionally integrated. A general representation of a system of fractionally integrated and possibly cointegrated (or cofractional) variables is

$$\Delta^d y_t = \Delta^{d-b} L_b \alpha \beta' y_t + \sum_{i=1}^p \Gamma_i \Delta^d L_b^i y_t + u_t, \quad (3.2.20)$$

where for real numbers d , Δ^d signifies the fractional differencing operator defined in Chapter 2 as

$$\Delta^d = (1 - L)^d = \sum_{i=0}^{\infty} (-1)^i \binom{d}{i} L^i$$

and L_b is the fractional lag operator defined as

$$L_b = 1 - \Delta^b$$

(see Johansen 2008; Johansen and Nielsen 2012). The model (3.2.20) assumes variables with fractional integration order d and cofractional order $d - b \geq 0$. The $K \times r$ matrices α and β are assumed to have rank r and $\beta' y_t$ is

³ Setting the off-diagonal elements of Λ to zero rules out that any of the variables is $I(2)$ (see Elliott 1998; Phillips 1998).

fractionally integrated of order $d - b$. For $b > 0$, variables with this property are called cofractional of order $d - b$. For given r , under the assumption that $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_u)$, the parameters of the model, including the fractional parameters b and d , can be estimated by ML. The asymptotic theory is developed by Johansen and Nielsen (2012).

The use of fractionally integrated and/or fractionally cointegrated models in structural VAR analysis has been very limited thus far. In practice, it is common to explore the fractional integration orders of the model variables first, before estimating the model conditional on these orders (see, e.g., Tschernig, Weber, and Weigand 2013). How well this approach works in small samples is not known. Nor is it known whether this approach leads to a better understanding of the DGP than, for example, using a local-to-unity approach. Which of these approaches is preferable in practice is likely to depend on the specific DGP. In either case, departing from the standard $I(1)$ setup complicates the analysis. Tschernig, Weber, and Weigand (2013) present simulation evidence that suggests that such complications may be worth contemplating when the deviations from the $I(1)$ model are sufficiently large and large samples are available.

3.3 Model Specification

Specifying VECMs involves choosing the lag order and determining the cointegrating rank. These two issues are discussed next.

3.3.1 Choosing the Lag Order

As in the stationary case, the VAR order can be chosen by sequential tests or model selection criteria if there are $I(1)$ variables among the components of the VAR process. If some of the variables are $I(1)$, the usual LR or Wald tests for the lag order have standard asymptotic χ^2 distributions under the null hypothesis as long as one does not test the hypothesis $\mathbb{H}_0 : A_1 = 0$.

Tests for residual autocorrelation can again be used in a bottom-up sequential testing strategy for lag-order selection as in Chapter 2. However, if there are integrated variables, the approximate distribution of the Portmanteau tests must be adjusted, as pointed out by Brüggemann, Lütkepohl, and Saikkonen (2006). For cointegrated processes the degrees of freedom also depend on the cointegrating rank. The approximate distribution for a VECM with cointegrating rank r and $p - 1$ lagged differences on the right-hand side is $\chi^2(K^2h - K^2(p - 1) - Kr)$. Since the cointegrating rank is typically unknown at the time of lag-order selection, the Portmanteau test is not useful for lag-order selection. In contrast, the Breusch-Godfrey LM test for residual autocorrelation can be applied to levels VAR processes with unknown cointegrating rank. Its asymptotic $\chi^2(hK^2)$ distribution under the null hypothesis is valid for

both $I(0)$ and $I(1)$ systems, as shown in Brüggemann, Lütkepohl, and Saikkonen (2006). Moreover, the information criteria discussed in Section 2.4 of the previous chapter maintain their asymptotic properties and, hence, can be used for cointegrated processes as well with the same justification as for stationary processes (see Paulsen 1984).

3.3.2 Specifying the Cointegrating Rank

Several proposals have been made for determining the cointegrating rank of a VAR process. Many of them are reviewed and compared in Hubrich, Lütkepohl, and Saikkonen (2001). Generally, a good case can be made for using the Johansen (1995) likelihood ratio approach (and its modifications) of testing for the cointegrating rank under the maintained assumption of Gaussian errors. Even if the DGP is not Gaussian, the resulting pseudo-LR tests have better properties than do many competitors. These tests are also attractive from a computational point of view because, for a given cointegrating rank r , ML estimates and, hence, the maximum of the likelihood function are easy to compute (see Section 3.2.2). Of course, in some cases alternative tests may be more suitable, as discussed in Hubrich, Lütkepohl, and Saikkonen (2001).

Denoting the matrix $\alpha\beta'$ in the error correction term by $\boldsymbol{\Pi}$, as before, the following sequence of hypotheses may be considered for selecting the cointegrating rank:

$$\begin{aligned} \mathbb{H}_0(r_0) : \text{rank}(\boldsymbol{\Pi}) = r_0 \text{ versus } \mathbb{H}_1(r_0) : \text{rank}(\boldsymbol{\Pi}) > r_0, \\ r_0 = 0, \dots, K-1. \end{aligned} \quad (3.3.1)$$

The corresponding LR test statistic is

$$LR^{trace}(r_0) = -2(\log l(r_0) - \log l(K)) = -T \sum_{i=r_0+1}^K \log(1 - \lambda_i),$$

where $l(r)$ denotes the maximum of the Gaussian likelihood function (3.2.9) given the cointegration rank r and λ_i is the i^{th} eigenvalue obtained by the Johansen procedure in Section 3.2.2. This test is often referred to as the trace test. The cointegrating rank specified in the first null hypothesis that cannot be rejected is chosen as the estimate for the true cointegrating rank r . If $\mathbb{H}_0(0)$, the first null hypothesis in this sequence, cannot be rejected, we proceed with a VAR process in first differences. If all the null hypotheses are rejected including $\mathbb{H}_0(K-1)$, the process is treated as $I(0)$, and a levels VAR model is specified.

Instead of testing the sequence of hypotheses specified in (3.3.1), one may alternatively test the sequence

$$\begin{aligned} \mathbb{H}_0(r_0) : \text{rank}(\boldsymbol{\Pi}) = r_0 \text{ versus } \mathbb{H}_a(r_0 + 1) : \text{rank}(\boldsymbol{\Pi}) = r_0 + 1, \\ r_0 = 0, \dots, K-1. \end{aligned} \quad (3.3.2)$$

Table 3.1. *Trace and Maximum Eigenvalue Tests for Cointegrating Rank*

λ_i	\mathbb{H}_0	LR^{trace}	critical value	LR^{max}	critical value
unrestricted intercept					
0.2020	$r = 0$	60.6041	28.71	47.1489	18.90
0.0398	$r = 1$	13.4552	15.66	8.4972	12.91
0.0234	$r = 2$	4.9580	6.50	4.9580	6.50
restricted intercept					
0.2020	$r = 0$	60.6968	32.00	47.1550	19.77
0.0399	$r = 1$	13.5418	17.85	8.5117	13.75
0.0238	$r = 2$	5.0301	7.52	5.0301	7.52

Note: The critical values for the upper panel are taken from table 1.1* and the critical values for the lower panel from table I* of Osterwald-Lenum (1992) for a 10% level test.

In other words, a specific rank r_0 is tested against the rank $r_0 + 1$. The LR test statistic for this pair of hypotheses is

$$LR^{max}(r_0) = -T \log(1 - \lambda_{r_0+1}).$$

This test is known as the maximum eigenvalue test. As in the case of the trace test, the test sequence terminates when the null cannot be rejected.

The LR test statistics corresponding to the null hypotheses in (3.3.1) and (3.3.2) have nonstandard asymptotic distributions. Their asymptotic distributions depend on the difference $K - r_0$ and on the deterministic terms included in the DGP, but they do not depend on the short-term dynamics. More precisely, under \mathbb{H}_0 the asymptotic distribution of the LR test corresponding to (3.3.1) is the trace of a matrix functional of multivariate Brownian motions, and that corresponding to (3.3.2) is the maximum eigenvalue of the corresponding matrix functional, which explains the specific names of the tests as trace and maximum eigenvalue tests. Critical values for various possible sets of deterministic components such as constants and linear trends have been computed by simulation methods and are available in the literature (e.g., Johansen 1995, chapter 15; Osterwald-Lenum 1992).

An Empirical Illustration. To illustrate the use of cointegration rank tests, we again utilize the trivariate empirical example from Section 3.2.2. Table 3.1 shows the eigenvalues obtained for a model with unrestricted intercept and for a model with the intercept contained only in the cointegration relations. The table also displays the corresponding values of the trace and maximum eigenvalue test statistics together with critical values for a test level of 10%. The level refers to each individual test and not to the joint level of the testing sequence.

Clearly, all tests reject the cointegrating rank $r = 0$, but they do not reject $r = 1$ and $r = 2$. Thus, based on Table 3.1, one may conclude that the preferred

cointegrating rank is $r = 1$. Because these tests tend to have low power in small samples, however, using the model with $r = 2$ rather than a model with $r = 1$ may also be justified when economic considerations suggest two cointegration vectors.

Size and Power Considerations. As in the case of sequential tests for the lag order, the overall size of tests for the cointegrating rank differs from the nominal size chosen for individual tests in the sequence. It is not clear how to control the overall size of these tests. The small-sample size and power of the trace and maximum eigenvalue tests are compared in Cheung and Lai (1993) and Toda (1994, 1995), among others. The performance of these tests is often similar. There is no clear ranking between the two types of testing sequences. Generally, the power of both types of tests tends to be low in many situations of practical interest.

The power of these tests may be improved by specifying the deterministic terms as tightly as possible. For example, if there is no deterministic linear trend term, it is desirable to perform the cointegrating rank tests without such terms. On the other hand, incorrectly omitting them may cause major size distortions. The asymptotic theory for testing hypotheses regarding the deterministic terms provided by Johansen (1995) can be helpful in this respect.

If a linear trend is required, but it is unknown whether or not this trend is orthogonal to the cointegration relationships, Demetrescu, Lütkepohl, and Saikkonen (2009) propose to apply rank tests with both alternative trend terms and to reject the null hypothesis if one of these tests rejects. They provide an asymptotic justification for this procedure.

Tests for the cointegrating rank have also been developed for the case of a structural break in the deterministic term either in the form of a level shift or a break in the trend slope, or both. In this case the critical values of the LR tests also depend on the timing of the break. This feature is inconvenient if the break point is not known *a priori* and has to be estimated. In that case, a test variant proposed by Saikkonen and Lütkepohl (2000a, 2000b) may be preferable. They suggest to estimate the deterministic term first by a GLS procedure and to adjust the data before applying a modified LR-type test to the adjusted system. The advantage is that the asymptotic null distribution of the test statistic does not depend on the break point, if only a level shift is considered. This fact facilitates the development of procedures that work even when the break date is unknown (e.g., Lütkepohl, Saikkonen, and Trenkler 2004; Saikkonen, Lütkepohl, and Trenkler 2006).

Although the short-run dynamics do not matter for the asymptotic properties of the rank test, they have a substantial impact in small and moderate samples. Therefore, the choice of the lag order p is quite important in conducting rank tests. On the one hand, choosing p rather large to avoid misspecifying the short-run dynamics tends to cause a substantial loss in the power of the

cointegrating rank tests. On the other hand, choosing the lag order too small may lead to dramatic size distortions. In a small-sample simulation study, Lütkepohl and Saikkonen (1999) conclude that using the AIC criterion for lag-order selection is a good compromise when determining the cointegrating rank.

There are many other proposals for modifying and improving the Johansen approach to cointegration testing. For example, Johansen (2002) presents a Bartlett correction designed to improve the performance of the Johansen cointegration tests in small samples. For further discussion of this and other approaches the reader is referred to Hubrich, Lütkepohl, and Saikkonen (2001). At present it appears that the Johansen approach should be the default among tests for the cointegrating rank, unless there is a compelling reason for using another type of test.

Subsystem Tests. Clearly, the Johansen approach to testing for the cointegrating rank has its drawbacks, in particular when used in large-dimensional systems or when many lags are necessary to capture the short-term dynamics. In this case, the test may lack the power to detect all cointegration relationships and, as a result, may underestimate the true cointegrating rank (see Gonzalo and Pitarkakis 1999). Hence, it is recommended to apply cointegration tests to all possible subsystems as well and to verify whether the results are consistent with those for the full model. For example, in a K -dimensional system where all variables are individually $I(1)$, if all variables are cointegrated in pairs, the cointegrating rank must be $K - 1$. Cointegration for the bivariate subsystems may be easier to analyze than cointegration in the full K -dimensional system. This observation suggests that one should analyze the subsystems first and then assess whether the subsystem results are consistent with the results for the full system, taking into account that the cointegrating rank tests may have reduced power for larger systems.

As an explicit example, consider a four-dimensional system $y_t = (y_{1t}, y_{2t}, y_{3t}, y_{4t})'$ and suppose that the first three components are individually $I(1)$. Put differently, $y_{it} \sim I(1)$ for $i = 1, 2, 3$, and $y_{4t} \sim I(0)$. If there are two linearly independent cointegrating relations between the first three variables, it is easy to see that all pairs $(y_{1t}, y_{2t})'$, $(y_{2t}, y_{3t})'$, and $(y_{1t}, y_{3t})'$ are also cointegrated. Moreover, taking into account that y_{4t} is $I(0)$, the cointegrating rank of y_t is three. If unit root tests are consistent with y_{it} being $I(1)$ for $i = 1, 2, 3$, and $y_{4t} \sim I(0)$, then we can proceed to testing cointegration in the three pairs $(y_{1t}, y_{2t})'$, $(y_{2t}, y_{3t})'$, and $(y_{1t}, y_{3t})'$. If the tests are consistent with cointegration in all three pairs, and if we also take into account that y_{4t} is $I(0)$, we can conclude that y_t has cointegrating rank three. Given that there are error probabilities associated with all these tests, we may want to follow up with a sequential test for the cointegrating rank of y_t applied to the full system. It is quite possible that in this case the testing sequence rejects ranks zero and 1 but

not rank two. Had we only applied the testing sequence to the full system, we might have incorrectly concluded that the cointegrating rank is two, whereas taking into account the results from the previous subsystem tests, we conclude that the rank is three and that the non-rejection of rank two for the full system is just due to a lack of power against the alternative of a larger rank.

In this example, other seemingly conflicting test results are, of course, possible. For example, the tests may suggest cointegration between y_{1t} and y_{2t} as well as y_{2t} and y_{3t} . In that case, y_{1t} and y_{3t} are necessarily also cointegrated. Yet, a cointegration rank test may not reject rank zero for $(y_{1t}, y_{3t})'$. Again, that result could be due to a lack of power. It must be kept in mind that not rejecting a null hypothesis does not establish the validity of the null hypothesis, but only that there is not enough sample information to be sure beyond a reasonable doubt that the null is false. If looking at the sample information from some other angle makes the data speak more clearly, then there is nothing wrong with relying on that information. Hence, in that example working with a cointegrating rank of $r = 2$ for the three-dimensional subsystem $(y_{1t}, y_{2t}, y_{3t})'$ and with $r = 3$ for the full four-dimensional system would be a sensible choice.

3.4 Diagnostic Tests

Most of the diagnostic tools discussed in the previous chapter for VAR models estimated in levels are also applicable to cointegrated VAR models and VECMs. We already mentioned tests for residual autocorrelation. Although Portmanteau tests are not suitable for models with unknown cointegrating rank because their distribution depends on r , they can be used to test for autocorrelation in the innovations of a given VECM that is assumed to be valid under the null hypothesis. In that case, the cointegrating rank is assumed to be correctly specified, and hence the Portmanteau test has an approximate $\chi^2(K^2h - K^2(p - 1) - Kr)$ distribution. As mentioned earlier, no adjustments are necessary for the Breusch-Godfrey LM test for error autocorrelation in a VECM or in a VAR model with integrated variables (see Brüggemann, Lütkepohl, and Saikkonen 2006).

The tests for nonnormality mentioned in the previous chapter can also be applied to the residuals of a VECM. The asymptotic distributions of the test statistics are not affected by $I(1)$ variables in the model. This result follows from the superconsistency of the estimator for the cointegration matrix and the properties of the empirical moment matrices of the integrated model variables (see Kilian and Demiroglu 2000).

It is also straightforward to extend tests for structural change such as the Chow test to the case of cointegrated processes. Suppose that a change in the parameters of the VECM (3.2.6) is suspected after period $T_1 < T$. In the context of VECMs one may then be interested in testing

$$\mathbb{H}_0 : \beta_{(1)} = \beta_{(2)}, \alpha_{(1)} = \alpha_{(2)}, \boldsymbol{\Gamma}_{(1)} = \boldsymbol{\Gamma}_{(2)}, \quad (3.4.1)$$

where the subscripts (1) and (2) refer to the first and second subperiods, respectively, against the alternative that at least one of the equalities is violated. The relevant Wald and LR tests have asymptotic χ^2 distributions. The degrees of freedom have to account for the fact that a nonsingular asymptotic distribution for the estimator of β is only obtained upon suitable normalization. Thus, $\beta_{(1)} = \beta_{(2)}$ implies only $r(K - r)$ restrictions. The corresponding LR statistic for testing the null hypothesis (3.4.1) hence has a limiting χ^2 distribution with $r(K - r) + rK + (p - 1)K^2$ degrees of freedom. Tests for constancy of only a subset of the parameters can be constructed analogously (see Hansen 2003). Of course, these tests can be extended to models with deterministic terms.

3.5 The Benefits of the VECM Representation

Ultimately, there are two reasons for users of structural VAR models to be interested in the VECM representation. One reason is the efficiency gains in estimating the reduced-form VAR model when the VECM is correctly specified. Once the VECM has been estimated, however, it is often convenient to represent the estimates as a VAR model in levels, as shown in Section 3.1. This representation facilitates the construction of forecasts and impulse responses, in particular. Expressing forecasts and impulse responses instead in terms of the VECM parameters does not provide any new insights. Similarly, imposing known unit roots and imposing known cointegration restrictions on the VAR model may improve the power of statistical tests such as Granger causality tests (see Lütkepohl and Reimers 1992b).

The other reason for considering VECMs is that they facilitate the imposition of restrictions on the long-run effects of structural shocks in the VAR model, which extends the range of identifying assumptions used for structural impulse response analysis. This point is discussed in Chapter 10.

3.6 Practical Issues

In practice, it is rarely clear when to use the VECM framework as opposed to a VAR in levels or in differences. Obviously, if we were sure of the existence of a unit root, we ought to impose it in estimation. If we are not sure about the presence of a unit root, as is typically the case in practice, the situation changes. On the one hand, incorrectly imposing a unit root results in overdifferencing of the data, rendering the VAR estimator inconsistent under standard assumptions. Failing to impose a unit root when the unit root is correct, on the other hand, preserves consistency. It only causes a reduction in the precision of the LS estimator and worsens its small-sample bias. Thus, the consequences of correctly imposing a unit root and of incorrectly imposing a unit root are asymmetric. This conclusion also extends to the question of whether to impose the cointegrating rank or the cointegrating vector in estimation. One practical strategy used in the literature is to rely on VAR models in levels that tend to be

robust to alternative specifications of the cointegrating rank and vectors. The potential cost of this approach, of course, is that we may not exploit all the economic structure in the DGP. In addition, the use of a level representation prevents us from using certain types of identification schemes for structural shocks that have been popular in the literature (see Chapter 10). An alternative strategy used in the literature is to specify a VECM if that model can be economically motivated and if the data do not object to this specification. This approach requires verifying that unit root and cointegration tests do not contradict the properties of the VECM. It also requires establishing that other diagnostic tests do not raise concerns about the model specification and showing that key features of the VECM are robust to relaxing the VECM specification. This strategy is not without risks, however, as discussed in the next section.

3.6.1 Limitations of Tests for Unit Roots and Cointegration

It may seem that the question of whether the model should be specified as a VAR model in levels, as a VAR model in differences, or as a VECM could be resolved by implementing a battery of pretests for unit roots and cointegration. This is not the case. There are two distinct problems. First, these tests cannot be used to confirm the features specified under the null hypothesis. For example, many empirical studies rely on evidence that unit root tests fail to reject the null hypothesis of a unit root as justification for imposing the unit root in estimation. However, a non-rejection of the null hypothesis only means that there is insufficient evidence to rule out unit roots beyond a reasonable doubt. This outcome may arise because the unit root is true or because the test lacks power against the alternative. All we can say is that the data are consistent with a unit root, just as they are consistent with the absence of a unit root. Much the same type of problem afflicts commonly used tests of the $I(0)$ null hypothesis. Similar problems of interpretation also arise in tests for cointegration.

Second, not only do these pretests suffer from low power, but Elliott and Stock (1994) and Cavanagh, Elliott, and Stock (1995), among others, demonstrate that the use of unit root pretests is invalid in environments when the dominant root is local-to-unity. Second-stage inference based on these pretests for unit roots exhibits substantial size distortions in empirically plausible situations. The same concern applies to pretests for the cointegration rank (see Elliott 1998). Although the local-to-unity model merely represents a thought experiment that may or may not approximate the DGP well, these studies show that inference drawn after doing pretests for unit roots and cointegration can be very misleading even for DGPs close to a VECM.

3.6.2 Alternative Approaches

There are two main alternatives in practice. One approach is to make explicit that the author wishes to impose some unit roots or some cointegration

relationship in estimation, while recognizing that such an assumption could potentially invalidate the empirical results. One may also examine the sensitivity of the results to alternative modeling assumptions. Of course, there is no guarantee that the results will be quantitatively or qualitatively consistent across specifications. The other approach is to specify the VAR model in levels. The use of the levels specification not only avoids the unit root issue. It also avoids the controversial issue of which cointegration restrictions to impose in estimation. It is useful to keep in mind, however, that the levels VAR model as well is only an approximation to the DGP.

Specifying the VAR model in levels is not without drawbacks either. First, this specification cannot be used for imposing restrictions on the long-run behavior of the data, as discussed in Chapter 10. Second, there is the question of how to conduct inference in that case. One option is to rely on the alternative asymptotic theory for near-unit root processes. Existing results within this theoretical framework, however, are limited to impulse responses, and the implementation of these methods of inference can be computationally challenging in some cases. Another option is to rely on the Bayesian methods of estimation and inference discussed in Chapter 5. The latter approach is not without its own challenges, one of which is the specification of the prior. The key difference is that standard Bayesian distribution theory is invariant to the knife-edge case of an exact unit root, which simplifies the analysis, whereas classical distribution theory is discontinuous at the unit circle. Precisely because the data are fairly uninformative about the presence of unit roots and cointegration, however, Bayesian estimates are sensitive to prior information about the long-run behavior of the data, leaving the unit root question unresolved. The third option is to exploit the fact that for higher-order autoregressive models in some cases standard methods of inference remain asymptotically valid even in the possible presence of unit roots and cointegration, as discussed in Section 3.2.3 (see Sims, Stock, and Watson 1990). Dealing with the singularities in the asymptotic covariance of many estimators may require the use of lag-augmented VAR models, however (see, e.g., Dolado and Lütkepohl 1996).

Even when the estimator of the VAR model in levels remains asymptotically valid in the presence of possible unit roots and cointegration, its small-sample properties may be unsatisfactory. The main concern in practice is the higher small-sample bias in the levels specification compared with models that impose unit roots. How severe this small-sample bias is depends on the model and on the sample size. This bias problem is greatly exacerbated by the inclusion of a deterministic time trend. A Monte Carlo study in Inoue and Kilian (2002a) demonstrates that even in univariate AR(2) models, the first-order approximation for the slope parameter in the levels model with deterministic time trend remains poor until the sample size exceeds $T = 500$. In contrast, without the deterministic time trend, $T = 300$ is perfectly adequate.

In practice, few researchers in empirical macroeconomics include a deterministic time trend in the VAR model. One reason is that many macroeconomic time series are highly persistent, and a deterministic time trend in conjunction with a (near) unit root in the autoregressive lag polynomial implies a (near) quadratic trend in y_t . This specification seems implausible when the trend in the data appears linear, as is the case for standard macroeconomic aggregates such as U.S. industrial production or real GDP. Moreover, the inclusion of a time trend in such models invites overfitting in small samples. This realization prompted Sims (1987) to argue for the exclusion of the deterministic trend term. This argument is uncontroversial from a Bayesian point of view. From a frequentist perspective, it requires taking a stand on the trend model in favor of a (near) stochastic trend. This is effectively what many non-Bayesian users of VAR models have done when adopting the same levels specification as Sims, although this point is rarely made explicit. While this additional restriction is immaterial in modeling nontrending series such as interest rates, for example, it is not innocuous when dealing with trending data such as real GDP or industrial production. The exclusion of the deterministic time trend effectively means that nonstochastic trending behavior in the data can only arise from a drift term in the VAR model. In this sense, the unrestricted VAR model in levels nests $I(1)$ processes and local-to-unity processes. Even if we are willing to grant the existence of unit roots, however, this approach is still useful in that it frees the user from the need to commit to a specific cointegration structure in estimating the VAR model.

Whether one prefers the VECM approach or the VAR in levels in the end comes down to how certain one is of the implied restrictions. Although the alternative of working with the VAR in levels is not without its own drawbacks, tentative simulation evidence in Gospodinov, Herrera, and Pesavento (2013) suggests that typically estimates based on the VAR model in levels are more accurate than estimates from models selected based on pretests. The relevance of such simulation results for applied work, of course, depends on how closely the assumed underlying DGP matches the features of the actual DGP.

4 Structural VAR Tools

This chapter provides an overview of the uses one can put a structural VAR to. We review the construction of structural impulse responses, forecast error variance decompositions, historical decompositions, forecast scenarios, and counterfactual outcomes including policy counterfactuals.

Consider the structural VAR(p) model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t,$$

where the $K \times 1$ vector y_t is presumed to be zero mean for expository purposes. The dimension of B_i , $i = 0, \dots, p$, is $K \times K$. The $K \times 1$ vector w_t is assumed to be white noise. The model is structural in that the elements of w_t are mutually uncorrelated and have clear interpretations in terms of an underlying economic model. We further postulate that the K model variables are driven by K distinct shocks such that their variance-covariance matrix Σ_w is of full rank. Thereby we rule out that the structural model includes equations that are merely identities, as is common in traditional simultaneous equations models, rather than being subject to stochastic errors. We also rule out that the data are generated by economic models in which there are fewer than K structural shocks. This assumption excludes, for example, the standard real business cycle model, in which all macroeconomic aggregates are driven by a technology shock only such that the covariance structure of the data is singular.

This model can be expressed in reduced form as

$$y_t = \underbrace{B_0^{-1} B_1 y_{t-1}}_{A_1} + \cdots + \underbrace{B_0^{-1} B_p y_{t-p}}_{A_p} + \underbrace{B_0^{-1} w_t}_{u_t}.$$

We normalize the covariance matrix of the structural errors $\mathbb{E}(w_t w_t') \equiv \Sigma_w = I_K$ without loss of generality such that the reduced-form error covariance matrix is $\mathbb{E}(u_t u_t') \equiv \Sigma_u = B_0^{-1} B_0^{-1'}$. Going back and forth between the structural and the reduced-form representation requires knowledge of the matrix B_0 that governs the instantaneous relationships among the model variables or of its inverse, the structural impact multiplier matrix B_0^{-1} . Given that $u_t = B_0^{-1} w_t$,

this matrix allows us to express the typically mutually correlated reduced-form innovations (u_t) as weighted averages of the mutually uncorrelated structural innovations (w_t), with the elements of B_0^{-1} serving as the weights.

Much of the discussion in subsequent chapters of this book addresses the question of how to recover estimates of B_0^{-1} (and, hence, estimates of the structural model coefficients B_0, \dots, B_p) from estimates of the reduced-form VAR model, as discussed in Chapters 2, 3, and 5. In the current chapter we focus on the question of how to present the estimates of the structural model in ways that facilitate the interpretation of economic data. For expository purposes, we postulate for now that B_0^{-1} (and hence B_0) is known. A review of how to specify and estimate B_0 and B_0^{-1} in practice can be found in Chapters 8, 9, 10, 11, 13, 14, and 15.

4.1 Structural Impulse Responses

Given B_0 and u_t , we immediately obtain $w_t = B_0 u_t$. Having identified the structural shocks w_t , our interest usually is not in the shocks themselves, however, but in the responses of each element of $y_t = (y_{1t}, \dots, y_{Kt})'$ to a one-time impulse in $w_t = (w_{1t}, \dots, w_{Kt})'$,

$$\frac{\partial y_{t+i}}{\partial w'_t} = \Theta_i, \quad i = 0, 1, 2, \dots, H,$$

where Θ_i is a $K \times K$ matrix. The elements of this matrix for given i are denoted as

$$\theta_{jk,i} = \frac{\partial y_{j,t+i}}{\partial w_{kt}},$$

such that $\Theta_i = [\theta_{jk,i}]$. Usually the objective is to plot the responses of each variable to each structural shock over time. Since there are K variables and K structural shocks, there are K^2 impulse response functions, each of length $H + 1$, where H is the maximum propagation horizon of the shocks.

A useful starting point for determining the structural impulse responses ($\theta_{jk,i}$) are the responses of y_{t+i} to the reduced-form errors u_t . They can be obtained by considering the VAR(1) representation of the VAR(p) process,

$$Y_t = \mathbf{A} Y_{t-1} + U_t, \tag{4.1.1}$$

where

$$Y_t \equiv \begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}, \quad \mathbf{A} \equiv \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_K & 0 & & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{bmatrix}, \text{ and } U_t \equiv \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

(see Chapter 2, Section 2.2.1). By successive substitution for Y_{t-i} , equation (4.1.1) can be written as

$$Y_{t+i} = \mathbf{A}^{i+1} Y_{t-1} + \sum_{j=0}^i \mathbf{A}^j U_{t+i-j}.$$

Left-multiplying this equation by $J \equiv [I_K, 0_{K \times K(p-1)}]$ yields

$$\begin{aligned} y_{t+i} &= J \mathbf{A}^{i+1} Y_{t-1} + \sum_{j=0}^i J \mathbf{A}^j U_{t+i-j} \\ &= J \mathbf{A}^{i+1} Y_{t-1} + \sum_{j=0}^i J \mathbf{A}^j J' J U_{t+i-j} \\ &= J \mathbf{A}^{i+1} Y_{t-1} + \sum_{j=0}^i J \mathbf{A}^j J' u_{t+i-j}. \end{aligned}$$

Thus, the response of the variable $j = 1, \dots, K$ in the VAR(p) system to a unit shock u_{kt} , $k = 1, \dots, K$, i periods ago, is given by

$$\Phi_i = [\phi_{jk,i}] \equiv J \mathbf{A}^i J'.$$

The Φ_i are also sometimes referred to as responses to VAR forecast errors, as dynamic multipliers, or simply as reduced-form impulse responses.

If y_t is covariance stationary, y_t can be expressed as a weighted average of current and past shocks, with weights Φ_i that decline, the more distant a shock lies in the past. This multivariate MA representation is:

$$y_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i} = \sum_{i=0}^{\infty} \Phi_i B_0^{-1} B_0 u_{t-i} = \sum_{i=0}^{\infty} \Theta_i w_{t-i}, \quad (4.1.2)$$

where we made use of $w_{t-i} = B_0 u_{t-i}$ and defined $\Theta_i \equiv \Phi_i B_0^{-1}$. Under stationarity, it follows immediately that

$$\frac{\partial y_t}{\partial w'_{t-i}} = \frac{\partial y_{t+i}}{\partial w'_t} = \Theta_i.$$

The latter responses may be computed simply by post-multiplying Φ_i , $i = 0, \dots, H$, by B_0^{-1} :

$$\Theta_0 = \Phi_0 B_0^{-1} = I_K B_0^{-1} = B_0^{-1}$$

$$\Theta_1 = \Phi_1 B_0^{-1}$$

$$\Theta_2 = \Phi_2 B_0^{-1}$$

⋮

etc. If the VAR in question is not stable, the same approach to computing Φ_i and Θ_i will work, but the impulse responses may not approach zero for $i \rightarrow \infty$ and will no longer represent the coefficients of the structural MA representation. The jk^{th} element of Θ_i , denoted by $\theta_{jk,i}$, represents the response of variable j to structural shock k at horizon $i = 0, 1, \dots, H$. There are K^2 impulse response functions

$$\theta_{jk,0}, \theta_{jk,1}, \dots, \theta_{jk,H}$$

that trace the responses over time of each variable j to each structural shock k .

It is also possible to construct linear combinations of structural impulse responses. For example, a VAR model may generate responses $\theta_{lk,i}$ for the nominal interest rate and $\theta_{mk,i}$ for the inflation rate, in which case the implied response of the real interest rate may be computed as $\theta_{lk,i} - \theta_{mk,i}$, $i = 0, \dots, H$. Likewise, for example, we may infer from the response of the inflation rate, say Δp_t , the implied response of the log price level,

$$p_{t+h} = p_{t-1} + \Delta p_t + \Delta p_{t+1} + \dots + \Delta p_{t+h},$$

by cumulating the responses of the inflation rate. In other words, the price level response to a shock in period t at horizon h is obtained as the sum

$$\sum_{i=0}^h \theta_{mk,i}$$

(see Lütkepohl 2005, section 2.3.2). Finally, it is also possible to compute non-linear transformations of impulse response functions such as the half-life of responses, defined as the time that elapses until the response has fallen to half of the impact response in absolute value (see Kilian and Zha 2002).

A direct implication of the linearity of the VAR model is that responses to negative shocks are the mirror image of responses to positive shocks. Another implication of the linearity of the VAR model is that the magnitude of the structural shock does not matter in constructing impulse response functions, because rescaling the shock merely rescales the entire impulse response function.¹

It is customary to choose B_0^{-1} such that the structural shocks represent one standard deviation of the time series of structural shocks, because such a shock is viewed as a shock of typical magnitude, but nothing hinges on this convention. It is also important to keep in mind that structural shocks in general are unit free and cannot be expressed in terms of the units of measurement applied to the model variables. Only in special cases will structural shocks be associated with a particular model variable.

¹ If one does not like these implications, one has to turn to nonlinear VAR models instead (see Chapter 18).

To estimate the impulse responses in practice, all we need to do is replace the unknown parameters in the reduced-form VAR(p) model by consistent estimates. Given estimates of the VAR parameters, \hat{A}_j , $j = 1, \dots, p$, and $\hat{\Sigma}_u$ and the implied estimate of B_0^{-1} , we can construct $\hat{\Phi}_i$ (and thus $\hat{\Theta}_i$) recursively for $i = 0, \dots, H$. For example, consider a VAR model proposed by Kilian and Park (2009) to shed light on the relationship between the global market for crude oil and the U.S. stock market. The model consists of the growth rate in global crude oil production, $\Delta prod_t$, a measure of the global business cycle in industrial commodity markets, rea_t , the real price of crude oil, $rpoil_t$, and U.S. real dividend growth, Δrd_t . Let $y_t = (\Delta prod_t, rea_t, rpoil_t, \Delta rd_t)'$. The objective is to decompose the reduced-form innovations into structural shocks representing oil supply shocks ($w_{1t}^{\text{oil supply}}$), shocks to the aggregate demand for all industrial commodities including crude oil ($w_{2t}^{\text{aggregate demand}}$), demand shocks that are specific to the oil market ($w_{3t}^{\text{oil-specific demand}}$), and a residual shock that captures all other determinants of U.S. real dividends (w_{4t}^{other}). By imposing a particular recursive ordering on B_0^{-1} such that the elements above the diagonal are zero, the remaining elements of B_0^{-1} can be recovered uniquely from Σ_u , as discussed in more detail in Chapter 8. Given an estimate of the A_j and hence Φ_i , $i = 0, \dots, H$, knowledge of B_0^{-1} allows us to estimate the corresponding structural impulse response matrices Θ_i .

Now suppose that we are interested in quantifying the effects of oil demand and oil supply shocks on the level of U.S. real dividends (measured in percent deviations from the baseline). In other words, we are interested in the cumulative effects of oil supply and oil demand shocks on the fourth variable in the VAR model. These effects may be computed by cumulating the estimates of $\theta_{4,1,i}$, $\theta_{4,2,i}$, and $\theta_{4,3,i}$ for $i = 0, \dots, H$. Figure 4.1 illustrates that a positive shock to global aggregate demand for all industrial commodities in the current month tend to raise U.S. real dividends by about 1% one year later, consistent with the view that a global demand boom has positive effects on the U.S. economy. In contrast, positive oil-specific demand shocks and negative oil supply shocks tend to lower real dividends by about 0.5% and 1%, respectively, consistent with the view that supply disruptions and other adverse events in the global oil market are detrimental to the U.S. economy. Figure 4.1 not only confirms prevailing views about the sign of these responses but allows us to quantify the effect of each shock.

4.2 Forecast Error Variance Decompositions

A second practically important question that a structural VAR model can answer is how much of the forecast error variance or prediction mean squared error (MSPE) of y_{t+h} at horizon $h = 0, 1, \dots, H$ is accounted for by each structural shock w_{kt} , $k = 1, \dots, K$. In a stationary model, the limit of the

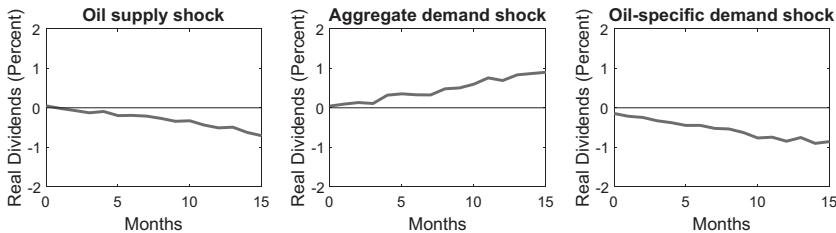


Figure 4.1. Point estimates of responses of U.S. real dividends to selected structural shocks

Source: Kilian and Park (2009).

forecast error variance decomposition, as $h \rightarrow \infty$, is the variance decomposition of y_t because the forecast error covariance matrix or MSPE converges to the unconditional covariance matrix of y_t (see Chapter 2). Thus, for stationary systems one may construct an MSPE decomposition for horizon infinity. In integrated systems the MSPE diverges when the forecast horizon goes to infinity, but the forecast error variance decomposition remains valid up to a finite maximum horizon of H . All we need to compute this decomposition is the Θ_i matrices, which we already computed for the structural impulse response analysis.

Recall from Chapter 2, Section 2.4.1, that for a VAR process the h -step ahead forecast error is

$$y_{t+h} - y_{t+h|t} = \sum_{i=0}^{h-1} \Phi_i u_{t+h-i} = \sum_{i=0}^{h-1} \Theta_i w_{t+h-i},$$

where $u_t = B_0^{-1} w_t$ allows us to replace $\Phi_i u_{t+h-i}$ by $\Theta_i w_{t+h-i}$. Hence, the MSPE at horizon h is

$$\begin{aligned} \text{MSPE}(h) &\equiv \mathbb{E}[(y_{t+h} - y_{t+h|t})(y_{t+h} - y_{t+h|t})'] = \sum_{i=0}^{h-1} \Phi_i \Sigma_u \Phi_i' \\ &= \sum_{i=0}^{h-1} \Theta_i \underbrace{\Sigma_w}_{I_K} \Theta_i' \\ &= \sum_{i=0}^{h-1} \Theta_i \Theta_i'. \end{aligned}$$

Let $\theta_{kj,h}$ be the kj^{th} element of Θ_h . Then the contribution of shock j to the MSPE of y_{kt} , $k = 1, \dots, K$, at horizon h is

$$\text{MSPE}_j^k(h) = \theta_{k,j,0}^2 + \dots + \theta_{k,j,h-1}^2$$

Table 4.1. *Forecast Error Variance Decomposition for U.S. Real Dividend Growth*

Horizon	Percent of h -Step Ahead Forecast Error Variance Explained by:			
	Oil supply shock	Aggregate demand shock	Oil-specific demand shock	Residual shock
1	0.2	0.2	1.7	98.0
2	0.6	0.4	2.1	97.0
3	0.8	0.5	2.1	96.6
12	2.8	6.8	4.5	85.8
∞	6.6	8.4	7.9	77.1

Source: Kilian and Park (2009).

and the total MSPE of y_{kt} , $k = 1, \dots, K$, at horizon h is:

$$\text{MSPE}^k(h) = \sum_{j=1}^K \text{MSPE}_j^k(h) = \sum_{j=1}^K (\theta_{kj,0}^2 + \dots + \theta_{kj,h-1}^2).$$

Dividing

$$\text{MSPE}^k(h) = \sum_{j=1}^K \frac{\text{MSPE}_j^k(h)}{\text{MSPE}^k(h)}$$

by $\text{MSPE}^k(h)$ yields the following decomposition for given h and k :

$$1 = \frac{\text{MSPE}_1^k(h)}{\text{MSPE}^k(h)} + \frac{\text{MSPE}_2^k(h)}{\text{MSPE}^k(h)} + \dots + \frac{\text{MSPE}_K^k(h)}{\text{MSPE}^k(h)},$$

where each ratio gives the fraction of the contribution of the j^{th} shock to the $\text{MSPE}(h)$ of variable k for $j = 1, \dots, K$. In other words, $\text{MSPE}_j^k(h)/\text{MSPE}^k(h)$ is the fraction of the contribution of shock j to the forecast error variance of variable k . By multiplying the fractions by 100 we obtain percentages.

The use of forecast error variance decompositions again is best illustrated by example. Returning to the four-variable VAR model considered by Kilian and Park (2009), an obvious question raised by their analysis is to what extent variability in U.S. real dividend growth can be explained by oil demand and oil supply shocks. This question may be answered based on a forecast error variance decomposition for U.S. real dividend growth. Kilian and Park focus on horizons of 1, 2, 3, and 12 months. Since real dividend growth (Δrd_t) is $I(0)$, a forecast horizon of $h = \infty$ is also considered. The results are presented in Table 4.1.

Ignoring rounding error, the entries in each row of the table sum to 100% by construction. The entries for horizon ∞ represent the variance decomposition

of U.S. real dividend growth. In practice, we can approximate ∞ by a very large number. This number is determined by showing that further increments to the horizon do not change the results up to the desired degree of accuracy.

In studying forecast error variance decompositions, one often is interested in the patterns across horizons. In this example we learn that the oil supply shock and the two oil demand shocks combined account for only 2% of the MSPE of U.S. real dividend growth at the one-month horizon, but that their explanatory power increases to 23% in the long run. This is evidence that the relationship between the global oil market and the U.S. stock market is weak at best. One may also be interested in the relative contribution of different shocks at a given horizon. For example, whereas at the one-month horizon oil specific demand shocks are much more important than oil supply shocks or aggregate demand shocks in explaining the forecast error variance of real dividend growth, each oil demand and oil supply shock accounts for about the same share of the unconditional variance.

4.3 Historical Decompositions

Structural forecast error variance decompositions and structural impulse response functions describe the average movements in the data. They represent unconditional expectations. Sometimes we are interested instead in quantifying how much a given structural shock explains of the historically observed fluctuations in the VAR variables. In other words, we would like to know the cumulative effect of a given structural shock on each variable at every given point in time. For example, we may not be interested in the average contribution of monetary policy shocks to the variability of real GDP growth over the last decades, but in the question of whether monetary policy shocks caused the 1982 recession.

Such historical decompositions may be computed from covariance stationary VAR models as follows. Suppose that we have data from 1 to t . Then, for any t ,

$$y_t = \sum_{s=0}^{t-1} \Theta_s w_{t-s} + \sum_{s=t}^{\infty} \Theta_s w_{t-s}.$$

In other words, the value of y_t depends on shocks w_1, \dots, w_t that can be estimated and shocks that predate the start of the sample at $t = 1$ and hence cannot be estimated. Given the fact that the MA coefficients die out, as we move further into the past, the second term (corresponding to the presample period) will have a steadily diminishing effect on y_t as t increases. Even after dropping the second term, it will be the case that

$$y_t \approx \sum_{s=0}^{t-1} \Theta_s w_{t-s}, \tag{4.3.1}$$

except for a number of periods early in the sample. We denote this approximation by

$$\hat{y}_t = \sum_{s=0}^{t-1} \Theta_s w_{t-s}. \quad (4.3.2)$$

We start by plotting \hat{y}_t and the (suitably demeaned and detrended) actual data y_t in the same plot. We discard the initial observations (also known as transients), for which the two series do not effectively coincide. How many periods it takes for the approximation to work well depends on the dominant root of the VAR process (defined as the root closest to the unit circle). The closer the dominant root is to unity, the more persistent is the effect of the shocks on y_t and the longer the period of transition. For the remaining sample period, we can decompose the sum in (4.3.2) to isolate the cumulative contribution of each shock to each element of \hat{y}_t , as discussed next.

Historical Decompositions as Time Series Plots. In practice, the construction of historical decompositions involves three simple steps. Step 1 is to compute the structural MA coefficient matrices $\Theta_0, \dots, \Theta_{T-1}$. Step 2 is to compute the structural shocks $w_t = B_0 u_t$, $t = 1, \dots, T$. Step 3 is to match up each structural shock, say shock j , with the appropriate impulse response weight, as required by the structural moving average representation, to form $T \times 1$ vectors of fitted values for variable k , denoted $\hat{y}_k^{(j)}$ for $j = 1, \dots, K$.

For example, let $K = 5$ and suppose that we are interested in the cumulative effect of each of the five structural shocks on the fourth variable of the VAR system. In that case, we compute the following weighted sums for $t = 1, \dots, T$:

$$\begin{aligned}\hat{y}_{4t}^{(1)} &= \sum_{i=0}^{t-1} \theta_{41,i} w_{1,t-i}, \\ \hat{y}_{4t}^{(2)} &= \sum_{i=0}^{t-1} \theta_{42,i} w_{2,t-i}, \\ \hat{y}_{4t}^{(3)} &= \sum_{i=0}^{t-1} \theta_{43,i} w_{3,t-i}, \\ \hat{y}_{4t}^{(4)} &= \sum_{i=0}^{t-1} \theta_{44,i} w_{4,t-i}, \\ \hat{y}_{4t}^{(5)} &= \sum_{i=0}^{t-1} \theta_{45,i} w_{5,t-i},\end{aligned}$$

where $\theta_{jk,i}$ denotes the response of variable j to shock k at horizon i and $w_{k,t}$ is the k^{th} structural shock at time t . Each vector $\widehat{y}_4^{(j)} = (\widehat{y}_{41}^{(j)}, \dots, \widehat{y}_{4T}^{(j)})'$ shows the cumulative contribution of shock j on the fourth variable in the VAR model over time. By construction, the value for \widehat{y}_{4t} is obtained as the sum

$$\widehat{y}_{4t} = \sum_{j=1}^K \widehat{y}_{4t}^{(j)}.$$

In practice, historical decompositions are computed by replacing the unknown quantities $\theta_{jk,i}$ and w_t by the usual estimates.

It is important to remember that historical decompositions involve an approximation error. This approximation error arises because we truncate the moving average representation. For example, y_{k1} depends on the structural shocks at date 1 as well as the infinite history of structural shocks. With much of the history of shocks unobserved, the approximation is bound to be poor initially. As we recursively update \widehat{y}_{kt} , however, more and more of the recent structural shocks that receive high impulse response weights are captured, and the weights of earlier unobserved shocks decline. Thus, \widehat{y}_{kt} approaches y_{kt} . How fast this convergence takes place depends on the persistence of y_{kt} .

The best way to determine the point T^* beyond which the historical decomposition is reasonably accurate is to plot both \widehat{y}_{kt} and y_{kt} against time. If we plot the historical decomposition along with actual data, we must first remove all deterministic components in the actual data, because the data generated from the structural MA representation (4.1.2) are zero mean by construction. All data predating the point of convergence must be discarded. Only $\widehat{y}_{kt}^{(j)}$, $j = 1, \dots, K$, for $t = T^*, T^* + 1, \dots, T$ can be used for the analysis of the historical decomposition. When the sample is short and the data are persistent, this fact may preclude the use of this tool.

The way to interpret historical decompositions is that each $\widehat{y}_{kt}^{(j)}$, $j = 1, \dots, K$, shows the cumulative effect of shock j on y_{kt} up to this point in time. This cumulative effect may be positive or negative relative to the benchmark of zero, which represents the mean value of the data. Each time series $\widehat{y}_{kT^*}^{(j)}, \dots, \widehat{y}_{kT}^{(j)}$, $j = 1, \dots, K$, shows how our approximate variable \widehat{y}_{kt} would have evolved if all other structural shocks had been turned off. Plotting each time series $\widehat{y}_{kT^*}^{(j)}, \dots, \widehat{y}_{kT}^{(j)}$, $j = 1, \dots, K$, against the actual data helps assess which structural shock(s) alone or in combination account for the fluctuations in $\widehat{y}_{kT^*}, \dots, \widehat{y}_{kT}$ during specific historical episodes of interest. By construction, the sum

$$\sum_{j=1}^K \widehat{y}_{kt}^{(j)}$$

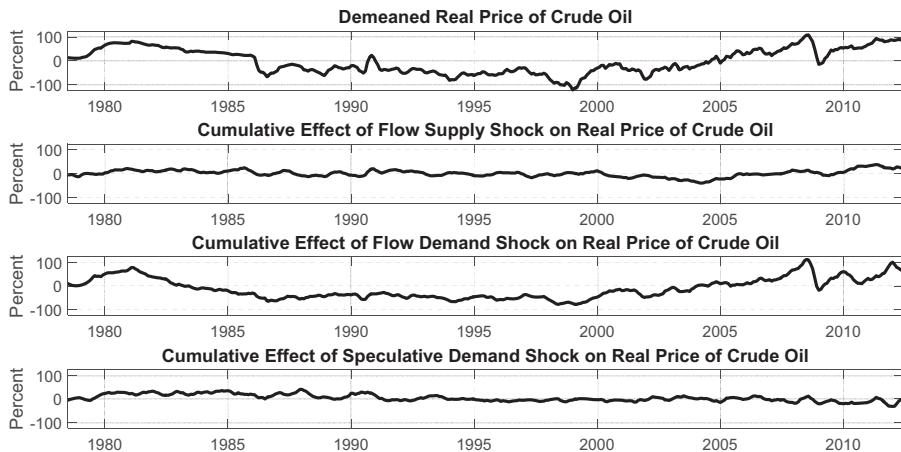


Figure 4.2. Historical decomposition of the real price of crude oil in percent deviations from the mean.

Source: Kilian and Lee (2014).

must equal the demeaned actual data except for model estimation error and approximation error arising from the truncation of the infinite sum. Because the structural MA representation is based on a parametric estimate of the mean of the data, defined as the k^{th} element of $(I_K - \hat{A}_1 - \dots - \hat{A}_p)^{-1} \hat{v}$, and that estimate in finite samples may differ from the sample average of the data, defined as the k^{th} element of \bar{y} , \hat{y}_{kt} may in practice differ slightly from the demeaned y_{kt} by a constant, even after disregarding the transients. When plotting both \hat{y}_{kt} and the demeaned y_{kt} against time, this discrepancy may be removed by demeaning \hat{y}_{kt} .

Figure 4.2 illustrates the use of historical decompositions in understanding the evolution of the real price of oil from the late 1970s to early 2012. The example is based on a global oil market model studied in Kilian and Lee (2014) which is discussed in more detail in Chapter 13. This structural model attributes variation in the real price of oil to shocks to the flow supply of oil, shocks to the flow demand for oil, a speculative oil demand shock, and a residual shock designed to capture various idiosyncratic shocks.

Suppose that we are interested in understanding what happened in the global oil market during a particular historical episode. For example, we may be interested in explaining the surge in the real price of oil between 2003 and mid-2008, keeping in mind that different historical episodes may be explained by different combinations of structural shocks. Figure 4.2 focuses on the role of the structural shocks that have an explicit economic interpretation under the maintained assumption that the real price of oil is an $I(0)$ time series during the estimation period. It shows that much of this surge (as well as the

collapse of the real price of oil in late 2008 and its recovery since then) must be attributed to the effects of flow demand shocks. Neither flow supply shocks nor speculative demand shocks are able to explain the surge in the real price of oil during this period. This result could not have been inferred from the structural impulse responses that trace out the average effect of a hypothetical one-time structural shock, or from forecast error variance decompositions that measure the extent to which a structural shock explains the variability of a variable on average.

The importance of conducting historical decompositions is not always appreciated in empirical macroeconomics. An occasional mistake in the business cycle literature is to view evidence of large and persistent impulse responses of real output to a structural shock as evidence of this shock's ability to explain the business cycle. This conclusion is unwarranted because impulse responses trace out the response to a one-time positive shock only, whereas business cycle variation in real output is driven by a sequence of shocks of different magnitude and signs. It is not uncommon for the effects of a positive shock in one period to be eclipsed by negative shocks in subsequent periods. Only the historical decomposition allows us to assess the cumulative effect of these shocks on the business cycle and the relative importance of different shocks in explaining particular recessions or expansions.

Extensions to Integrated Time Series. Unlike structural impulse responses, historical decompositions are designed for stationary VAR variables. They cannot be applied to integrated or cointegrated variables in levels without modifications because they rely on the existence of a stationary MA representation of the DGP. A case in point is a VAR model containing real GDP. Suppose we think of this time series as containing a unit root. One way of proceeding would be to express this time series in growth rates and to apply the historical decomposition to the growth rate. The disadvantage of that approach is that it involves a loss of information. We cannot learn anything about the level of real GDP predicted by the model because that information is lost when converting the data to growth rates. We can, however, quantify the ability of a given shock to explain the cumulative change in real GDP since a given point in time.

If the possibly integrated real GDP variable is included in levels in the VAR model, as is common in applied work, it is not possible to construct a historical decomposition for real GDP at all. Sometimes VAR users report the results of an exercise in which, starting at some date T^* , all subsequent realizations of shock j are replaced by zero. The difference between the observed values of real GDP after T^* and those obtained under this conditional counterfactual are then taken as the cumulative effect of this structural shock since T^* . While this approach has merit, if we are interested in the incremental role played by shock j during a particular episode in the data, it differs from a historical

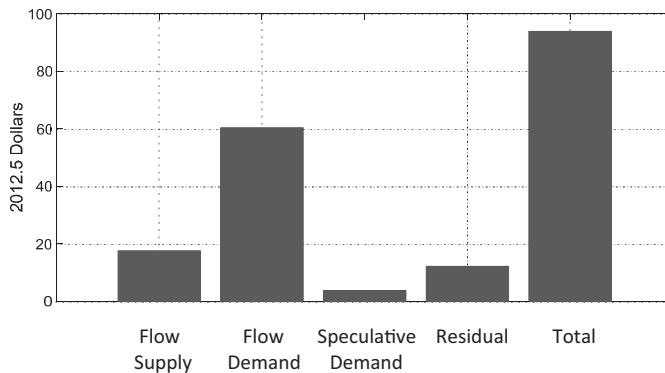


Figure 4.3. Contribution of each structural shock to the cumulative change in the real price of oil from January 2003 to June 2008.

Source: Adapted from Kilian and Lee (2014).

decomposition in that it ignores the cumulative effects of shock j prior to T^* on the subsequent real GDP realizations.

Historical Decompositions as Bar Charts. Kilian and Lee (2014) proposed two alternative simpler formats for presenting the information conveyed by historical decompositions. The first format allows one to express the cumulative change in a stationary variable explained by each structural shock in the form of a bar chart. To continue with the earlier example, let $\hat{y}_{4T_1}^{(j)}$ denote the cumulative contribution of shock j to variable y_{4t} at date T_1 and $\hat{y}_{4T_2}^{(j)}$ the corresponding result for date $T_2 > T_1$. Then the cumulative change in y_{4t} between dates T_1 and T_2 contributed by shock j is approximated by:

$$\hat{y}_{4T_2}^{(j)} - \hat{y}_{4T_1}^{(j)},$$

with the total change in y_{4t} measured as:

$$\sum_{j=1}^5 (\hat{y}_{4T_2}^{(j)} - \hat{y}_{4T_1}^{(j)}) = \hat{y}_{4T_2} - \hat{y}_{4T_1} \approx y_{4T_2} - y_{4T_1}.$$

This allows us to provide a quick summary of the evidence for any subperiod of interest. Kilian and Lee (2014) use this tool to summarize the determinants of the surge in the real price of oil between January 2003 and June 2008 on the one hand and since the peak of the real price of oil on the other. Figure 4.3 shows an example of such a bar chart. The bar on the right shows that between 2003 and mid-2008 the price of oil in total increased by 95 dollars in real terms. The four bars on the left indicate how much of this 95-dollar increase must be attributed to each of the four structural shocks. They show that 18 dollars of this increase are explained by the flow supply shock, 61 dollars by the flow

demand shock, 4 dollars by the speculative demand shock, and 12 dollars by the residual shock, providing evidence against the hypothesis that speculative demand was a quantitatively important determinant of the surge.

Historical Decompositions as Counterfactuals. An alternative way of assessing the cumulative contribution of each shock to y_{kt} , also proposed in Kilian and Lee (2014), is to construct the counterfactual

$$y_{kt} - \hat{y}_{kt}^{(j)},$$

where y_{kt} denotes the k^{th} actual time series variable (including any deterministic means or trends) and $\hat{y}_{kt}^{(j)}$ is the cumulative contribution of shock j to the evolution of variable k up to date t , as defined earlier. This representation avoids the impossible task of having to attribute the deterministic component of y_{kt} to individual shocks j . The counterfactual series indicates how the variable of interest would have evolved had one been able to replace all realizations of shock j by zero while preserving the remaining structural shocks in the model. If the counterfactual exceeds y_{kt} , this means that structural shock j lowered y_{kt} . A counterfactual below the observed y_{kt} means that the shock in question raised y_{kt} . The vertical distance between the actual y_{kt} and the counterfactual tells us how much shock j affected y_{kt} at this point in time.

It is useful to contrast this approach with the earlier approach of constructing cumulative increases in $\hat{y}_{kt}^{(j)}$ and y_{kt} . That approach focused on changes in the variables over time explained by a given structural shock rather than the component of y_{kt} at a given point in time driven by that structural shock. To move from a plot of the counterfactual to the cumulative increase measure, one would have to compare the difference between the counterfactual and y_{kt} on the first and on the last date of the counterfactual and construct the rate of change over time in this difference. Thus, these representations are mutually consistent but focus on different aspects of the same data.

Figure 4.4 provides an example of some historical counterfactuals based on structural VAR models. It characterizes the evolution of the real price of oil from January 2003 to June 2008 based on the global oil market model of Kilian and Lee (2014), again maintaining the assumption that the real price of oil is $I(0)$. The counterfactuals show how the real price of oil would have evolved, had one been able to replace all realizations of a given structural shock by zero while preserving the remaining structural shocks in the model. The first panel shows, for example, that in the absence of flow supply shocks, the level of the real price of oil would have been higher between 2003 and mid-2005. In other words, positive flow supply shocks helped keep the real price of oil down. For the remainder of the sample, flow supply shocks played only a minor role. The second panel shows that, in the absence of flow demand shocks, the real price of oil would have been somewhat higher in 2003, but much lower starting

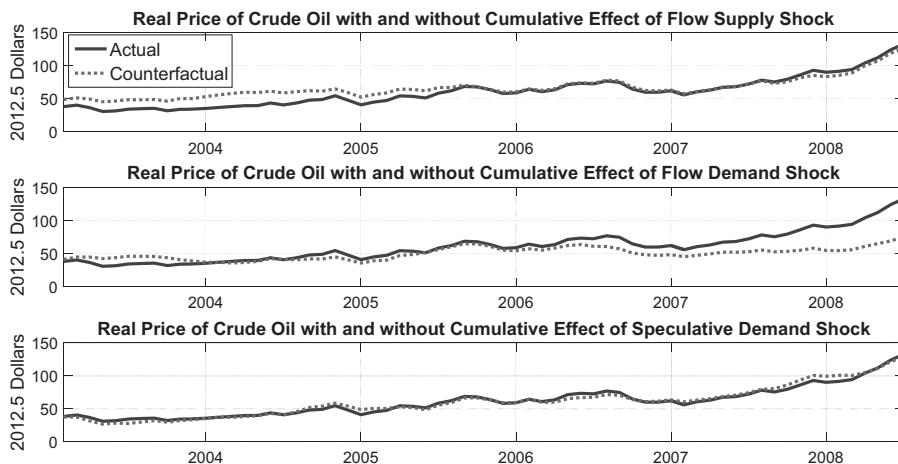


Figure 4.4. Historical counterfactuals for the real price of crude oil from January 2003 to June 2008

Source: Kilian and Lee (2014).

in 2006. Finally, the third panel demonstrates that speculative demand shocks made little difference overall. In fact, they slightly lowered the real price of oil in late 2007 and early 2008, according to this model.

4.4 Forecast Scenarios

The objective of forecast scenarios is to assess the sensitivity of reduced-form VAR forecasts to hypothetical future events. Constructing such forecast scenarios requires a structural VAR model, the reduced-form representation of which generates accurate out-of-sample forecasts. It is important to keep in mind that the objective of constructing forecast scenarios is not to improve the accuracy of the baseline reduced-form VAR forecast. Indeed, that forecast by construction already provides the best possible out-of-sample prediction from a given forecasting model.

Unlike a conventional forecast, forecast scenarios are not designed to characterize the most likely outcomes given past observations, but to characterize the risks associated with unlikely possible outcomes in the future in the form of a hypothetical “what-if” question. For example, we may ask how the occurrence of a certain future event, no matter how unlikely, would affect the future values of the variable that is being forecast. This is a question often raised by policymakers.

The simplest form of a forecast scenario is a conditional point forecast. More generally, the sensitivity of a forecast to alternative assumptions

about future structural shocks or observables may be captured by probability-weighted conditional forecast densities and summarized by formal risk measures (see Baumeister and Kilian 2014). Risk analysis along these lines allows users of VAR forecasts to explore by how much and at what forecast horizon downside and upside risks change as a function of the probability weights attached to different scenarios. In conjunction with historical decompositions, this approach helps satisfy the needs of policymakers not only for accurate out-of-sample forecasts but for a coherent economic interpretation of both historical data and forecasts.

There are two closely related but distinct approaches to the construction of conditional forecasts from structural VAR models. One conditions on sequences of future structural shocks. This type of conditional forecast was developed by Baumeister and Kilian (2014). The other approach relies on forecasts conditional on the future path of observables. The latter approach was proposed by Waggoner and Zha (1999).

4.4.1 Conditional Forecasts Expressed in Terms of Sequences of Structural Shocks

Following Baumeister and Kilian (2014), forecast scenarios are constructed from the structural moving average representation of the VAR model. We first iterate that representation forward to obtain:

$$y_{t+h} = \sum_{i=0}^{\infty} \Theta_i w_{t+h-i} = \sum_{i=0}^{h-1} \Theta_i w_{t+h-i} + \sum_{i=h}^{\infty} \Theta_i w_{t+h-i},$$

where y_{t+h} denotes the dependent variable h periods in the future, which can be written as the sum of two terms. The second term captures the cumulative effects on y_{t+h} of all structural shocks that already occurred between $-\infty$ and t . This term is predetermined at t . The first term captures the cumulative effect on y_{t+h} of all structural shocks that have yet to occur between $t+1$ and $t+h$. When computing forecasts, the conventional approach is to set this first term to zero in the forecast equation because each structural shock in the model is unconditionally zero in expectation. It follows that the second term represents the unconditional model-based expectation (or forecast of y_{t+h} at date t),

$$y_{t+h|t} = \sum_{i=h}^{\infty} \Theta_i w_{t+h-i},$$

which is obtained by setting the future shocks, w_{t+1}, \dots, w_{t+h} , to zero. This forecast is identical by construction to the date t forecast that would be obtained from iterating forward the autoregressive reduced-form representation of the VAR model.

Feeding in a sequence of nonzero future structural shocks and conditioning on them in computing the forecast, in contrast, provides a conditional point

forecast. Such forecast scenarios may be based on purely hypothetical shock sequences, as long as the shock sequence considered is within the range of historical experience. This caveat is reminiscent of the discussion of modest policy interventions in econometric models in Leeper and Zha (2003). Shocks that are unusually large by historical standards not only may induce changes in behavior of the type emphasized by Lucas (1976), but there also would be reason to doubt the adequacy of standard linear model approximations when considering shock sequences involving extremely large shocks.

Forecast scenarios may also be based on sequences of structural shocks that occurred during selected episodes in the past. A historical decomposition can help identify such shock sequences. For example, Baumeister and Kilian (2014) identify the sequence of structural shocks in the oil market associated with events such as the Asian crisis of 1997 or the collapse of oil prices in 2008, and explore the consequence of a recurrence of these events for forecasts of the real price of oil. When postulating shock sequences directly based on historical precedent, the Lucas (1976) critique does not apply by construction, given the maintained assumption of a well-specified structural VAR model.

A final caveat is that forecast scenarios by construction involve sequences of unanticipated shocks. This does not mean that one cannot construct forecast scenarios involving anticipation, but that doing so requires a structural VAR model that allows for anticipation. An example is provided in Baumeister and Kilian (2014). For further discussion, see Chapter 17.

Conditional Point Forecasts. It is convenient to normalize all conditional forecasts relative to the baseline forecast from the structural model obtained by setting all future structural shocks to zero, which eliminates the dependence of the forecast scenario on t . The plot of this normalized conditional forecast represents the upward or downward adjustments of the baseline forecast that would be required if a given hypothetical scenario were to occur. More formally, for a vector of future structural shocks $\{w_{t+1}^{\text{scenario}}, w_{t+2}^{\text{scenario}}, \dots, w_{t+h}^{\text{scenario}}\}$ with nonzero mean, the revision required in the baseline forecast of y_{t+h} , $h = 1, 2, \dots$, in expectation would be:

$$\begin{aligned} & \mathbb{E} \left(\sum_{i=0}^{h-1} \Theta_i w_{t+h-i} + \sum_{i=h}^{\infty} \Theta_i w_{t+h-i} \middle| \{w_{t+h-i} = w_{t+h-i}^{\text{scenario}}\}_{i=0}^{h-1}, \Omega_t \right) \\ & - \mathbb{E} \left(\sum_{i=0}^{h-1} \Theta_i w_{t+h-i} + \sum_{i=h}^{\infty} \Theta_i w_{t+h-i} \middle| \{w_{t+h-i} = w_{t+h-i}^{\text{baseline}}\}_{i=0}^{h-1}, \Omega_t \right) \\ & = \sum_{i=0}^{h-1} \Theta_i w_{t+h-i}^{\text{scenario}}, \end{aligned}$$

where $\Omega_t = \{w_{t+h-i}, i = h, h+1, \dots\}$ denotes the information set of the forecaster at date t and $\sum_{i=0}^{h-1} \Theta_i w_{t+h-1}^{\text{baseline}} = 0$ because $w_{t+h-i}^{\text{baseline}} = 0$ for $i = 0, \dots, h-1$. This approach is similar to the construction of impulse response functions. The key difference is that an impulse involves a one-time structural shock $w_{t+1}^{\text{scenario}} \neq 0$ followed by $w_{t+i}^{\text{scenario}} = 0$ for $i = 2, 3, \dots, h$, whereas forecast scenarios tend to involve sequences of nonzero structural shocks extending over several periods. Once the forecast scenarios have been expressed in percent deviations from the baseline forecast, the scenario forecasts are constructed by scaling the baseline reduced-form VAR forecast. This simply involves adding the deviation from the baseline forecast to the baseline forecast such that the scenario forecast reduces to

$$y_{t+h|t} + \sum_{i=0}^{h-1} \Theta_i w_{t+h-i}^{\text{scenario}}.$$

In practice, forecast scenarios are constructed by replacing Θ_i by a consistent estimate. An important difference between conditional and unconditional forecasting methods is that, as with impulse response analysis, in generating conditional forecasts, the objective is to identify to the best of our ability the true dynamic effects of a sequence of structural shocks, which necessitates the use of fully revised data. Accordingly, the estimates of Θ_i should be based on fully revised data, even if the unconditional forecasts are based on real-time data. Moreover, given the linearity of the VAR model, forecast scenarios (like impulse response functions) are time invariant. In other words, a given hypothetical sequence of shocks will cause the same revisions of the baseline forecast at each point in time. This also means that deviations from the baseline practice do not necessarily have to be recomputed every month, except to the extent that a longer sample offers efficiency gains in estimating the structural model.

An example of the use of forecast scenarios is shown in Figure 4.5. Based on a model of the global oil market similar to that used in Kilian and Lee (2014), Baumeister and Kilian (2014) investigate a wide range of real-time forecast scenarios for the real price of crude oil, including a return of Iraqi oil production to full capacity, a supply disruption in Libya, a strong recovery of the global economy, a financial meltdown similar to the collapse of Lehman Brothers, and two contagion scenarios in which expectations of rising oil prices are triggered by political events in the Middle East. Some of these scenarios are based on historical precedent, while others are purely hypothetical. All scenarios involve sequences of structural shocks within the range of historical experience. Figure 4.5 shows how the forecast of the real price of oil would deviate from the baseline real-time VAR forecast as of December 2010 if one were willing to condition on each one of these events occurring in isolation. Such evidence allows policymakers to gauge the potential effects of unlikely but high-impact events on the real price of crude oil.

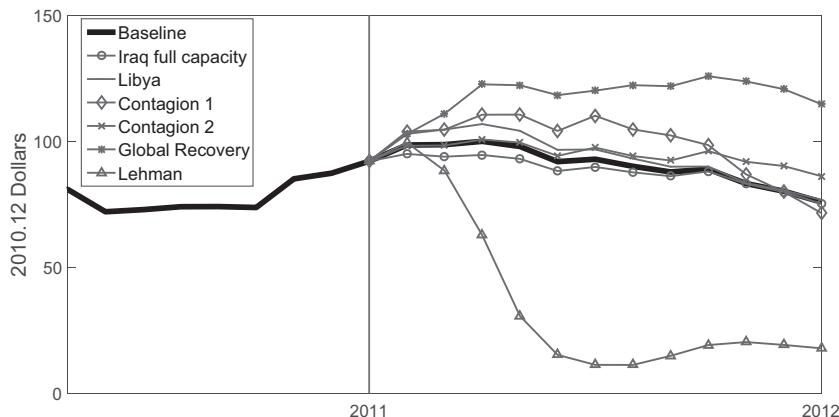


Figure 4.5. Selected real-time forecast scenarios for the real price of crude oil as of December 2010.

Source: Adapted from Baumeister and Kilian (2014).

Conditional Forecast Densities. So far our analysis has focused on conditional point forecasts and has abstracted from the uncertainty surrounding these forecasts. A more formal evaluation of the risks inherent in forecasts requires the construction of conditional density forecasts. We postulate that the scenario errors, $w_{t+1}^{\text{scenario}}, w_{t+2}^{\text{scenario}}, \dots, w_{t+h}^{\text{scenario}}$, have the same distribution as the baseline errors, $w_{t+1}^{\text{baseline}}, w_{t+2}^{\text{baseline}}, \dots, w_{t+h}^{\text{baseline}}$, during the forecast period except that the mean of this distribution has changed to the alternative shock sequence associated with the scenario. This allows one to construct the conditional forecast density for a given forecast scenario by shifting the unconditional forecast density and centering it on the conditional point forecast implied by the forecast scenario.

The unconditional forecast density may be obtained by bootstrap methods. For example, if the VAR errors are iid, we can generate sequences of random shocks, $w_{t+1}, w_{t+2}, \dots, w_{t+h}$, by drawing with replacement from the set of model residuals, allowing us to simulate the future path of the data conditional on the information set at date t and the estimated VAR model parameters. Having simulated a large number of realizations of the sequence $y_{k,t+1}, y_{k,t+2}, \dots, y_{k,t+h}$, an estimate of the forecast density for $y_{k,t+h}$ at a given forecast horizon h may be constructed using a Gaussian kernel density smoother. An example is the baseline 1-year ahead forecast density shown in Figure 4.6. For a discussion of more sophisticated bootstrap methods for predictive inference that allow for estimation uncertainty in the VAR model parameters the reader is referred to Chapter 12.

The conditional forecast density is all we need when considering one forecast scenario at a time. Alternatively, several forecast scenarios may be combined into one conditional point or density forecast if one is willing to

attach probability weights to each individual scenario. These probabilities reflect how much weight the end user wishes to attach to each one of a set of hypothetical events. They allow one to explore the consequences of conditioning on combinations of events occurring in expectation with given probability weights. One example would be an analysis of two scenarios in conjunction. In this case, it can be useful to consider the implications of the respective probabilities of only one of these two scenarios occurring, or of both occurring in conjunction, instead of postulating that only one of the three possible scenarios occurs. Alternatively, one may wish to explore how the increasing probability of an extreme event affects the forecast compared with the baseline forecast.

The construction of probability-weighted scenarios is a natural generalization of the idea of conditioning on one scenario at a time, which amounts to assigning probability one to this scenario and probability zero to all other scenarios. Conditioning on hypothetical events, with the mean of the distribution expressed as a probability-weighted average, serves to explore the sensitivity of the forecast to these events. A useful tool in constructing the probability weights is a Venn diagram that illustrates the relationship between alternative scenarios. The ultimate objective of constructing probability-weighted conditional density forecasts is to summarize the results in the form of formal risk measures. This allows the policymaker to explore how changes in the probability weights attached to different scenarios affect the upside and downside risks of the forecast, providing insights into the determinants of the baseline forecast.

Figure 4.6 illustrates how alternative scenarios may affect the predictive density generated from the baseline VAR forecasting model for the real price of oil. All results take the information set available to forecasters in December 2010 as given. The baseline density implies substantial uncertainty about the real price of oil one year into the future. Figure 4.6 also examines how the density changes under three alternative scenarios that differ primarily in how optimistic the user is about the future state of the global economy. The more optimistic the user is about the future state of the global economy, the higher the probability weight on the global recovery scenario compared with the other scenarios discussed earlier. More optimism translates to higher demand for crude oil in expectation and hence shifts the oil price density to the right without affecting its shape. Increased pessimism shifts the density to the left. Figure 4.6 shows that the baseline predictive density almost coincides with the density under the moderately pessimistic scenario, but expectations of a global recovery may shift the predictive density substantially to the right.

The extent to which predictive densities evolve under alternative scenarios can be difficult to convey to policymakers. It is therefore useful to summarize these results based on formal measures of oil price risks, building on the work

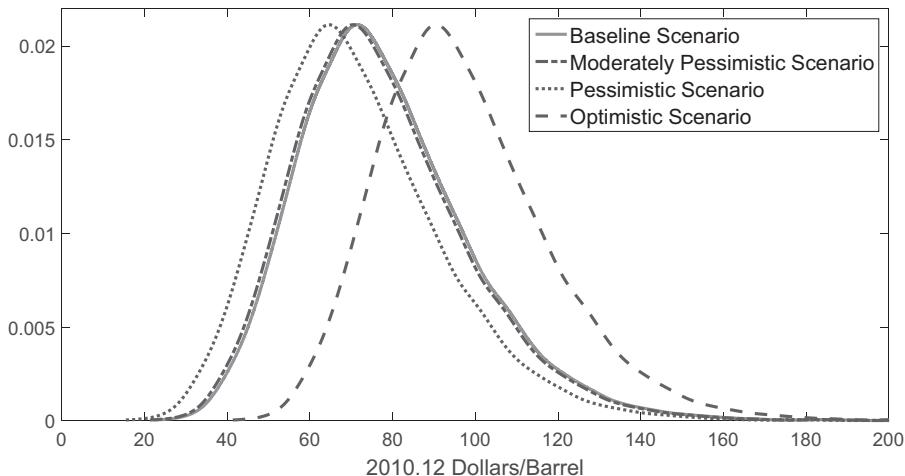


Figure 4.6. Real-time probability weighted 1-year ahead density forecasts as of December 2010 under different scenarios about the future state of the global economy.
Source: Adapted from Baumeister and Kilian (2014).

of Kilian and Manganelli (2007). Let R_{t+h} denote the real price of oil h months from now measured in December 2010 dollars and F the predictive distribution of R_{t+h} . For expository purposes, consider the event of R_{t+h} exceeding an upper threshold of \$100. Then, under conventional assumptions about the degree of risk aversion, the upside risk in the oil price forecast, defined as

$$\int_{100}^{\infty} (R_{t+h} - 100) dF,$$

can be expressed as

$$\mathbb{E}(R_{t+h} - 100 | R_{t+h} > 100) \times \mathbb{P}(R_{t+h} > 100),$$

where the first term of the expression is referred to as the tail-conditional expectation and the second term as the tail probability. The first column of Table 4.2 reports the tail probability of the real price of oil exceeding \$100; the second column shows the tail conditional expectation (or expected excess) which tells us by how much one would expect to exceed \$100 conditional on exceeding that threshold, and the last column reports the weighted expected excess defined as the product of the first two terms. Similar risk measures could also be computed for the risk of the real price of oil falling below \$80, for example, depending on the event the user wishes to guard against. It may also be of interest to study the balance of the upside and downside risks.

Table 4.2. *Real-Time Risk Measures as of December 2010*

Scenario	h	$\mathbb{P}(R_{t+h} > 100)$	$\mathbb{E}(R_{t+h} - 100 R_{t+h} > 100)$	$\mathbb{E}(R_{t+h} - 100 R_{t+h} > 100) \times \mathbb{P}(R_{t+h} > 100)$
Baseline	3	0.50	11.03	5.56
	6	0.36	15.25	5.42
	12	0.14	15.40	2.22
Moderately Pessimistic	3	0.56	11.51	6.42
	6	0.31	14.81	4.60
	12	0.14	15.29	2.08
Pessimistic	3	0.44	10.49	4.57
	6	0.19	13.69	2.59
	12	0.10	14.76	1.41
Optimistic	3	0.86	16.78	14.48
	6	0.63	18.41	11.60
	12	0.40	17.82	7.10

Source: Adapted from Baumeister and Kilian (2014). The more optimistic we are about the future state of the global economy, the larger the upside risk for the real price of oil.

Table 4.2 shows that under the baseline scenario the probability of the real price of oil exceeding \$100 is 50% at the 3-month horizon, but declines to 36% at the 6-month horizon and to 14% at the 12-month horizon. The expected excess is about \$11, but rises to \$15 at longer horizons. Overall, the upside risk as measured by the weighted expected excess declines with the horizon. Changes to the baseline scenario are reflected in somewhat different risk estimates. Conditioning on the optimistic scenario, for example, increases the probability of the real price of oil surpassing \$100 to 86% at the 3-month horizon, coupled with an expected excess of \$17.

4.4.2 Conditional Forecasts Expressed in Terms of Sequences of Observables

In closely related work, Waggoner and Zha (1999) proposed the construction of forecasts conditional on the path of the observables. This approach makes sense when modeling the path of a policy instrument such as the short-term interest rate under the premise that the central bank can control the policy instrument by executing a sequence of interest rate shocks. In that case, forecast scenarios are based on the sequence of monetary policy shocks implied by the hypothesized path of the interest rate. Waggoner and Zha (1999) show how to construct conditional point forecasts and conditional forecast densities in the context of a monetary policy VAR model using a Bayesian approach to estimation and inference (see Chapter 5).

The key difference between Waggoner and Zha's forecasts conditional on the interest rate and Baumeister and Kilian's forecast scenarios for the real price of oil is that in oil markets none of the many structural shocks affecting the variable of interest can be controlled by the policymaker. Moreover, there would be no point in constructing scenarios in terms of a prespecified path of the real price of oil, given that there is no unique mapping from the real price of oil to the underlying structural shocks. The nature of the structural shocks determining the path of the future real price of oil, however, is what determines the implied future path of the other model variables. Thus, without specifying the precise sequence of structural shocks, the scenario analysis would be ill-defined.

4.5 Simulating Counterfactual Outcomes

The construction of historical decompositions and of conditional forecasts is closely related to the construction of counterfactuals, as discussed in Kilian (2017). A counterfactual refers to a simulation of the path of the VAR model variables under a different sequence of structural shocks than observed in the actual data. The use of counterfactuals is best illustrated by an empirical example.

There has been a remarkable surge in the production of unconventional crude oil in the United States since November 2008. This surge has been tied to the rapid expansion of the production of shale oil (see Kilian 2016). The U.S. shale oil boom has important ramifications for the global market for crude oil. It represents a shift in the global supply curve along the global demand curve, so, all else equal, one would expect the shale oil boom to have lowered the global price of crude oil, as measured by the Brent price of crude oil. What is not obvious is how much higher the Brent price would have been in the absence of the shale oil boom. Answering this question requires the construction of a counterfactual based on a structural dynamic model of the global oil market.

In short, the question of interest is how different the global price of crude oil would have been if all producers other than the United States had maintained their observed oil production levels but the U.S. shale oil boom had never happened. Kilian (2017) provides an answer to this question based on a thought experiment in which increased U.S. shale oil production is entirely due to exogenous oil supply shocks rather than being caused by oil demand shocks. This assumption is consistent with the view that in the absence of technological innovation, the shale oil boom would not have taken place, but it makes no allowance for any additional effects working through oil price expectations.

The analysis involves three steps. First, we construct the counterfactual level of global oil production. Second, we specify a structural VAR model of the global oil market and estimate this model on the full sample. Given the change in the quantity of oil produced in equilibrium under the counterfactual, it is

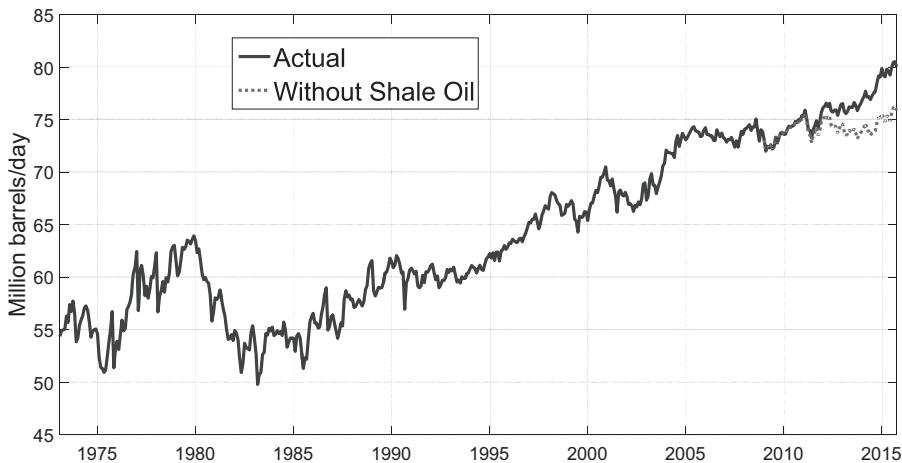


Figure 4.7. Actual level of global oil production and counterfactual level in the absence of the U.S. shale oil boom.

Source: Kilian (2017).

straightforward to infer from this model how much the supply curve must shift each month, conditional on past data, in order to produce the counterfactual equilibrium quantity of global oil production. Third, we replace the original sequence of flow supply shocks by the resulting counterfactual sequence of flow supply shocks into the structural VAR model and simulate the evolution of the global price of crude oil under this new sequence, while leaving unchanged all other structural shocks in the model.

The construction of the counterfactual level of global oil production in step 1 relies on direct measures of the volume of shale oil production available from the U.S. Energy Information Administration. Figure 4.7 shows the actual evolution of global oil production as well as the counterfactual path obtained by subtracting the barrels of shale oil produced since November 2008 from the actual data.

In step 2, Kilian (2017) employs an updated version of the global oil market model of Kilian and Lee (2014) that already served as an empirical example in Sections 4.3 and 4.4. Let $y_t = (y_{1t}, y_{2t}, y_{3t}, y_{4t})'$ denote the vector of variables in this model, where y_{1t} stands for the growth in world oil production and y_{3t} for the log of the real price of crude oil, in particular. Consider the structural VAR model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_{24} y_{t-24} + w_t,$$

where the 4×1 vector y_t is assumed to be stationary and the deterministic regressors have been suppressed for expository purposes. The dimension of

B_i , $i = 0, \dots, 24$, is 4×4 . The 4×1 vector $w_t = (w_{1t}, w_{2t}, w_{3t}, w_{4t})'$, where w_{1t} denotes a shock to the global flow supply of crude oil, is assumed to be zero mean white noise with a diagonal 4×4 covariance matrix Σ_w that is of full rank and that, without loss of generality, can be normalized to equal the identity matrix. This structural model can be expressed in reduced form as

$$y_t = A_1 y_{t-1} + \dots + A_{24} y_{t-24} + u_t,$$

where $A_i = B_0^{-1} B_i$, $i = 1, \dots, 24$, and $u_t = B_0^{-1} w_t$ with variance-covariance matrix $\Sigma_u = B_0^{-1} (B_0^{-1})'$. Let the 4×4 matrix $\partial y_{t+i}/\partial w_t' = \Theta_i$ denote the responses of the model variables to each of the structural shocks at horizon $i = 0, 1, \dots, H$. The matrix $\Theta_i = [\theta_{jk,i}]$ consists of the elements $\theta_{jk,i} = \partial y_{j,t+i}/\partial w_{kt}$ that denote the response of variable j to structural shock k at horizon i . These responses may be computed as $\Theta_i = J \mathbf{A}^i J' B_0^{-1}$, where $J = [I_4, 0_{4 \times 4(24-1)}]$ and \mathbf{A} denotes the matrix of slope parameters obtained by expressing the VAR(24) model in its VAR(1) companion format.

In constructing the counterfactual in question, we make use of the fact that, after removing the deterministic terms,

$$y_t \approx \sum_{i=0}^{t-1} \Theta_i w_{t-i},$$

as already noted in the section on historical decompositions. As a result, the fitted value for the log real price of oil can be written as

$$\hat{y}_{3t} = \sum_{j=1}^4 \hat{y}_{3t}^{(j)},$$

where $\hat{y}_{3t}^{(j)}$ denotes the cumulative contribution since the beginning of the sample of structural shock j to the third model variable at time t , defined as

$$\begin{aligned}\hat{y}_{3t}^{(1)} &\equiv \sum_{i=0}^{t-1} \theta_{31,i} w_{1,t-i}, \\ \hat{y}_{3t}^{(2)} &\equiv \sum_{i=0}^{t-1} \theta_{32,i} w_{2,t-i}, \\ \hat{y}_{3t}^{(3)} &\equiv \sum_{i=0}^{t-1} \theta_{33,i} w_{3,t-i}, \\ \hat{y}_{3t}^{(4)} &\equiv \sum_{i=0}^{t-1} \theta_{34,i} w_{4,t-i}.\end{aligned}$$

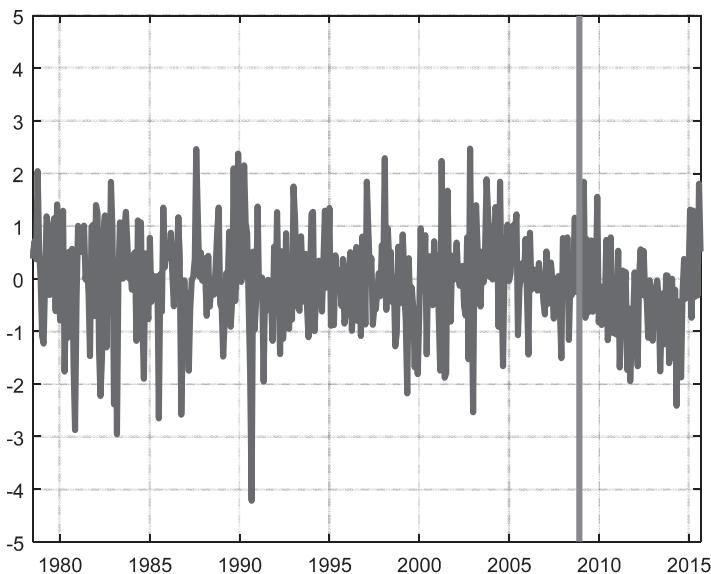


Figure 4.8. Counterfactual sequence of flow supply shocks in the absence of the U.S. shale oil boom.

Notes: The vertical line marks the beginning of the shale oil boom in November 2008.

Source: Kilian (2017).

In practice, $\theta_{jk,i}$ and w_t are replaced by consistent estimates. An analogous decomposition also exists for the fitted value \hat{y}_{1t} of the growth rate of global oil production. These decompositions play an important role in constructing the counterfactual.

Having recovered estimates of B_0^{-1} , $\theta_{jk,i}$ and $w_t = B_0 u_t$, $t = 1, \dots, T$, from the full-sample estimates of the Kilian and Lee (2014) model, as discussed in Section 4.3, we retain the realizations of the flow supply shock, w_{1t} , up to November 2008, but replace the estimates of w_{1t} for the remainder of the sample by counterfactual values chosen to ensure that the path of global oil production corresponds to the counterfactual path shown in Figure 4.7. Given the demeaned growth rate of world oil production implied by the counterfactual, these values may be easily computed by an iterative procedure. Having computed what the expected value of y_{1t} , given $\hat{y}_{1t}^{(j)}$, $j = 2, 3, 4$, would have been next month if this month's supply shock had been zero, the difference between that model prediction and the target level of oil production growth, scaled by the impact response of world oil production to a flow supply shock, will be the magnitude of the flow supply shock, w_{1t} , required to reach the target. Figure 4.8 shows the sequence of flow supply shocks that would have to be imposed in the oil market VAR model to make global oil production follow

the counterfactual path of world oil production in Figure 4.7. By construction, this shock sequence is identical to the observed structural residuals, \hat{w}_{1t} , until November 2008.

An important concern is whether the flow supply shocks required to implement the prespecified path of global oil production after November 2008 are too large by historical standards or too predictable to maintain the assumption of a time-invariant structural VAR model. Figure 4.8 illustrates that the counterfactual shock sequence involves no flow supply shocks that are unusually large by historical standards, addressing the first concern. Nor is there reason to believe that economic agents would have been able to predict these counterfactual flow supply shocks. This concern would be warranted if there were an unusually large number of consecutive flow supply shocks of the same sign (referred to as a “run”) relative to past data. The longest run of flow supply shocks observed under the counterfactual after December 2008 is fourteen months. Although the longest run in the flow supply shock sequence prior to December 2008 lasted only nine months, a run of fourteen months is not entirely unprecedented. Runs of similar length can be found in other structural shock sequences in the same model, suggesting that this counterfactual is not unreasonable.

To answer the question of how much the global price of oil would have changed in the absence of the shale oil boom, in step 3 we recompute $\hat{y}_{3t}^{(1)}$ under the counterfactual sequence of flow supply shocks while retaining the other three structural shock sequences, as originally estimated, in computing $\hat{y}_{3t}^{(2)}$, $\hat{y}_{3t}^{(3)}$, and $\hat{y}_{3t}^{(4)}$. Comparing the implied sequence of \hat{y}_{3t} to the demanded value of y_{3t} provides a measure of the cumulative impact of the shale oil boom on the log real price of oil.

In order to map the results for the log real price of oil (defined as the real U.S. refiners’ acquisition cost for oil imports in the Kilian and Lee model) to the nominal Brent price, we proceed as follows. Given the implied sequence of \hat{y}_{3t} , the counterfactual Brent price is constructed from the real price of Brent crude oil, scaled up by the percent deviation between the counterfactual real price of oil and the actual real price of oil in the VAR model. The counterfactual real Brent price is then converted back to dollars using the U.S. consumer price index. Figure 4.9 shows how much higher the nominal Brent price of crude oil would have been in the absence of the shale oil boom.

The gap between the actual and the counterfactual path of the Brent price measures the cumulative effect of the shale oil boom on the Brent price. This effect became important starting in 2011 and reached nearly 10 dollars per barrel of crude oil in early 2014, before declining toward the end of the sample.

This example illustrates how counterfactuals may be used to answer policy-relevant questions based on structural VAR models. The same approach could be used, in principle, to simulate the performance of the economy under a

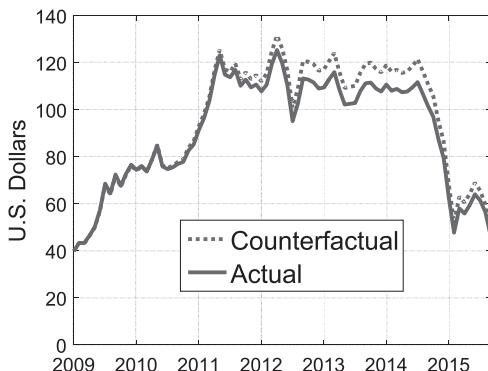


Figure 4.9. Evolution of the nominal Brent price of crude oil with and without shale oil.

Source: Kilian (2017).

counterfactual path of the interest rate for a given policy reaction function.² The construction of policy counterfactuals becomes more complicated when the counterfactual involves changing the policy rule rather than merely the path of the interest rate conditional on the estimated policy reaction function. This situation is considered in the next subsection.

4.6 Policy Counterfactuals

An important question in structural VAR analysis dating back to Bernanke, Gertler, and Watson (1997) and Sims and Zha (2006a), among others, has been how to construct policy counterfactuals. For example, Bernanke et al. were interested in assessing the effects of the systematic response of monetary policy to exogenous oil price shocks. For this purpose they augmented the monetary policy reaction function in an otherwise standard monetary policy VAR model to allow the policymaker to respond directly to oil price shocks. Their premise was that this systematic policy response was responsible in part for the subsequent recessions. The policy counterfactual Bernanke et al. proposed involved the central bank holding the interest rate constant following the oil price innovation. The question of interest was how recessionary the consequences of oil price shocks would have been, had the central bank followed a constant interest rate rule.

A different policy counterfactual for the same class of models was proposed by Kilian and Lewis (2011). They observed that – if we are interested in the

² Note that one could also construct a forecast conditioning on the path of the interest rate in much the same way, as discussed in Section 4.4.2, except that in the latter case future values of the remaining structural shocks would be set to zero in expectation rather than being held fixed at their actual realizations.

question of how much of a difference the direct response of the central bank to oil price shocks makes – the relevant benchmark should be a conventional model of monetary policy in which the central bank reacts to fluctuations in other macroeconomic variables such as inflation and real output with only the direct response to oil price shocks shut down.

In what follows, we contrast the construction of these two counterfactuals, referred to as the BGW and KL counterfactual for convenience. For illustrative purposes, we consider a VAR model that includes only real non-oil commodity prices, the real price of oil, U.S. real output, U.S. inflation, and the federal funds rate. Under the KL counterfactual of only shutting down the direct response to the real price of oil, the objective is to construct a sequence of hypothetical shocks to the federal funds rate that offsets the contemporaneous and lagged effects of including the real price of oil in the policy reaction function. Such a shock sequence may be recovered from the estimated VAR, allowing us to compare the actual evolution of the VAR variables during a given episode with that under the counterfactual policy shock sequence.

In contrast, for the original BGW counterfactual, which Bernanke et al. refer to as the Sims-Zha counterfactual based on a proposal dating back to 1996 and published in Sims and Zha (2006a), one simply constructs a hypothetical path of the policy shock that offsets all endogenous dynamics in the federal funds rate such that the federal funds rate remains unchanged over time. For a more detailed description of the construction of these counterfactuals, see Kilian and Lewis (2011). For an alternative description of the BGW counterfactual, see Hamilton and Herrera (2004).

Figure 4.10 provides an illustration of the two counterfactuals based on the VAR model of monetary policy estimated in Kilian and Lewis (2011) on monthly data for 1967m5–1987m7. It compares the evolution of real output (measured by the Chicago Fed National Activity Index or CFNAI), of consumer price inflation, and of the federal funds rate in response to a one-time unexpected increase in the real price of oil under each of the two counterfactuals. It also shows the response of these observables under the policy regime maintained in the unrestricted VAR model. Figure 4.10 shows that shutting down the direct response to oil price shocks has virtually no effect on inflation and little effect on real output. The Federal Reserve still would have raised interest rates by roughly the same number of basis points in response to an exogenous positive oil price shock, but the bulk of that response would have occurred three months later. Holding the interest rate constant, in contrast, as proposed by Bernanke et al., again would have had virtually no effect on inflation and real output would have been only slightly higher than observed.

An obvious concern in constructing policy counterfactuals from structural VAR models is that such counterfactuals may run afoul of the Lucas critique (see Lucas 1976). The concern is that agents may change their behavior in

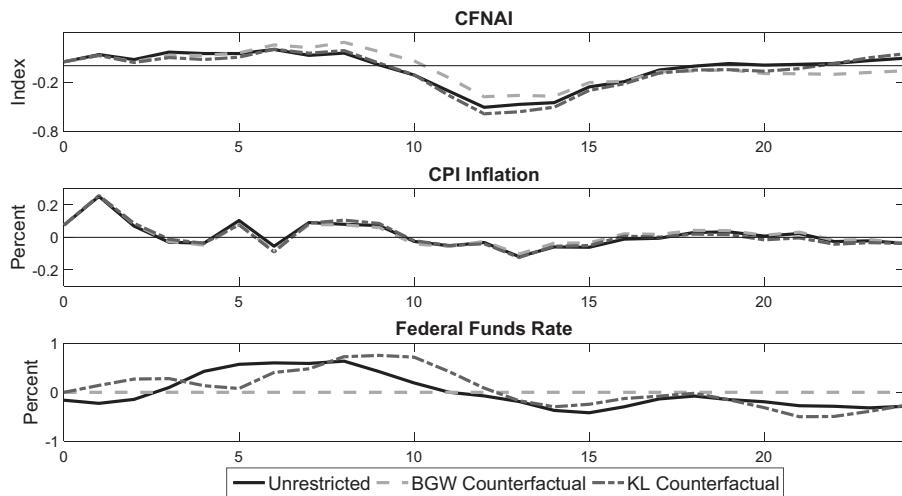


Figure 4.10. Counterfactual paths of key observables under alternative policy counterfactuals.

Source: Kilian and Lewis (2011).

anticipation of shifts in policy, causing the structure of the VAR model to change over time, thus invalidating the analysis. A reasonable presumption is that a policy shock sequence outside of the range of historical experience is likely to be recognized by agents and, if implemented, would cause changes in the structure of the underlying model. Leeper and Zha (2003) make the case that modest policy interventions, in contrast, do not significantly shift agents' beliefs about the policy regime and hence are not subject to the Lucas critique.

Hamilton and Herrera (2004) discuss diagnostics for evaluating whether the shock sequences required to implement counterfactual paths for policy instruments in VAR models are subject to the Lucas critique. They focus on both the magnitude of the policy surprises required to implement the counterfactual and the extent to which agents would have been able to anticipate these surprises. Their conclusion is that the constant interest rate counterfactual of Bernanke et al. involves large policy shocks outside of the range of historical experience and requires agents to be fooled by repeated policy surprises in the same direction for an extended period. Applying these criteria to the counterfactual proposed by Kilian and Lewis (2011) and illustrated in Figure 4.10 does not reveal the same problem. This point is illustrated in Figure 4.11 which plots the sequences of policy shocks required to implement this counterfactual, given the monetary policy regime in the unrestricted structural VAR model. Not only are the hypothesized interest rate shocks well within the range of historical

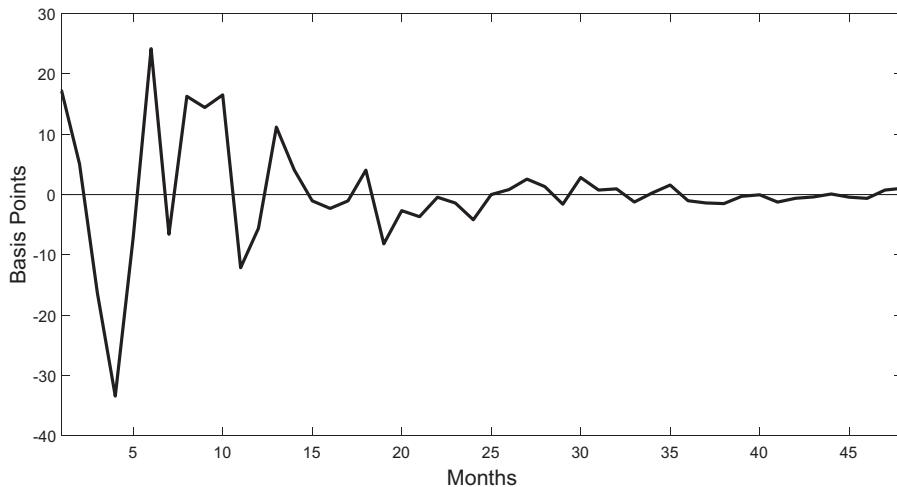


Figure 4.11. Sequence of policy shocks required to implement the KL policy counterfactual.

Source: Adapted from Kilian and Lewis (2011).

policy shocks but there is no evidence of serial correlation in the policy shock sequence that would allow agents to predict and anticipate the counterfactual path of monetary policy.

These examples illustrate that there are obvious limits to the types of policy counterfactuals that may be entertained based on a structural VAR model. Nevertheless, when used with care, policy counterfactuals may provide valuable insights.

5 Bayesian VAR Analysis

Given the fact that VAR models are estimated on comparatively short samples and hence tend to be imprecisely estimated, many users of VAR models are sympathetic to the idea of imposing additional structure in estimation. Shrinking parameter estimates toward specific values, in particular, may help reduce the variance of unrestricted LS estimators. The Bayesian approach provides a formal framework for incorporating such extraneous information in estimation and inference. It also facilitates the inclusion of extraneous economic information about the VAR model parameters that would be difficult to incorporate in frequentist analysis.

The ideas behind the Bayesian approach differ fundamentally from the frequentist approach. Broadly speaking, frequentists condition on the existence of a parameter vector, say θ , that governs the DGP from which the observed sample of data is thought to have been obtained. The objective is to infer the value of θ from this sample. Whereas the data are stochastic, the parameter vector θ is nonstochastic. Probability statements refer to the properties of the estimator of θ in repeated sampling.

In contrast, in Bayesian analysis, the parameter vector θ is treated as stochastic, whereas the data are considered nonstochastic. Bayesians are concerned with modeling the researcher's beliefs about θ . These beliefs are expressed in the form of a probability distribution. The Bayesian probability concept is a subjective probability concept. It does not require a repeated sampling experiment. Moreover, the nature of the DGP is immaterial for inference on the parameter of interest because inference conditions on the observed data.

The researcher's subjective beliefs about the value of the parameter vector θ before inspecting the data are summarized in the form of a prior probability distribution (or prior for short) for θ . This prior information is formally combined with the information contained in the data, as captured by the likelihood function of the model, to form the posterior probability distribution (or simply posterior) for θ . This posterior conveys everything the researcher knows about the model parameters after having looked at the data.

Frequentists are aware that they do not know the DGP for a given set of variables, but they evaluate the properties of their methods under the assumption that there exists a true parameter vector θ that can be objectively defined. They postulate a DGP and then conduct their analysis as if this model structure including any distributional assumptions were correct. In contrast, Bayesians do not need to take a stand on the DGP. However, their formal framework for deriving and evaluating the posterior requires users to articulate their prior beliefs in the form of a prior probability distribution. It also involves assuming a specific distribution (or family of distributions) for the data when evaluating the likelihood.

The choice between frequentist and Bayesian methods depends on the preferences of the user. In practice, convenience may also play an important role. In some situations frequentist methods are easier to deal with, and in other cases Bayesian methods are more convenient. It has to be kept in mind, however, that these approaches not only may produce numerically different answers, but that their interpretation is fundamentally different, even when the estimates coincide numerically. Since Bayesian methods are frequently used in VAR analysis, it is essential to have at least a basic understanding of this approach.

Although Bayesian methods often require extensive computations, they have become quite popular for VAR analysis because the cost of computing has decreased dramatically over the past decades. Moreover, new methods and algorithms have broadened the applicability of Bayesian methods. In Section 5.1, we briefly review some basics of the Bayesian methodology and terminology and then discuss Bayesian methods commonly used in VAR analysis. Section 5.2 reviews some commonly used prior specifications for the reduced-form VAR parameters. There are many good introductory treatments of Bayesian methodology in general and for macroeconomic analysis in particular. Canova (2007) and Koop and Korobilis (2009) fall in the latter category. Recent surveys include, for example, Ciccarelli and Rebucci (2003), Del Negro and Schorfheide (2011), and Karlsson (2013). The present chapter follows in part Lütkepohl (2005, section 5.4), Geweke and Whiteman (2006), and Canova (2007).

5.1 Basic Terms and Notation

5.1.1 Prior, Likelihood, Posterior

The Bayesian approach treats the data, $\mathbf{y} = (y'_1, \dots, y'_T)'$, as given and the parameter of interest, θ , as unknown. Inference about θ is conditional on the data. Prior information on θ is assumed to be available in the form of a density. Suppose that the prior information is summarized in the prior probability density function (pdf) $g(\theta)$ and that the sample pdf conditional on a particular

value of the parameter θ is $f(\mathbf{y}|\theta)$. The latter function is algebraically identical to the likelihood function $l(\theta|\mathbf{y})$. The two types of information are combined by applying Bayes' theorem, which states that the joint density is

$$f(\theta, \mathbf{y}) = g(\theta|\mathbf{y})f(\mathbf{y}),$$

and, hence,

$$g(\theta|\mathbf{y}) = \frac{f(\mathbf{y}|\theta)g(\theta)}{f(\mathbf{y})}.$$

Here $f(\mathbf{y})$ denotes the unconditional sample density which is just a normalizing constant for a given sample \mathbf{y} and does not depend on θ . Scaling the posterior by $f(\mathbf{y})$ ensures that the posterior density integrates to 1. In other words, the joint sample and prior information can be summarized by a function that is proportional to the likelihood function times the prior density $g(\theta)$,

$$g(\theta|\mathbf{y}) \propto f(\mathbf{y}|\theta)g(\theta) = l(\theta|\mathbf{y})g(\theta), \quad (5.1.1)$$

where \propto indicates that the right-hand side is proportional to the left-hand side. The conditional density $g(\theta|\mathbf{y})$ is the posterior pdf. It contains all the information that we have on the parameter vector θ after having updated our prior views by looking at the data. Thus, the posterior pdf is the basis for estimation and inference. In the next subsection we discuss Bayesian inference within this general framework in more detail.

5.1.2 Bayesian Estimation and Inference

Point Estimators. Bayesian inference on the parameter vector θ is based on the posterior distribution. Often moments of the posterior distribution are of interest. For example, one may be interested in $\mathbb{E}(\theta|\mathbf{y})$. The posterior mean is often used as a point estimator for θ if the posterior distribution is Gaussian or at least symmetric. More generally one is often interested in expected values of functions of θ . If we are interested in constructing a point estimate of some function $h(\theta)$ that may be vector-valued or a scalar, we minimize the expected loss of $h(\theta)$ based on some loss function. Denoting this loss function by $\mathcal{L}(h^\dagger, h(\theta))$, the point estimate, say \tilde{h}^\dagger , is chosen such that

$$\tilde{h}^\dagger = \operatorname{argmin}_{h^\dagger} \int \mathcal{L}(h^\dagger, h(\theta))g(\theta|\mathbf{y})d\theta, \quad (5.1.2)$$

where h^\dagger denotes an element of the range of $h(\theta)$. For example, if the loss function is quadratic and the first two moments of the posterior distribution exist, the point estimate corresponds to the posterior mean. Under alternative loss functions, the median or the mode of the posterior distribution may be

used as point estimates. These estimates coincide if the posterior distribution is Gaussian.

Credible Sets. The Bayesian concept corresponding to a confidence set in classical inference is a credible set. For $0 < \gamma < 1$, a $(1 - \gamma)100\%$ credible set for θ with respect to the posterior $g(\theta|y)$ is a set Ω such that

$$\mathbb{P}(\theta \in \Omega|y) = \int_{\Omega} g(\theta|y)d\theta = 1 - \gamma. \quad (5.1.3)$$

This set can be estimated by sampling from the posterior distribution. In general it will not be unique, however. To make the credible set unique, one may choose Ω such that $g(\theta|y) \geq c_{\gamma}$ for all $\theta \in \Omega$, where c_{γ} is the largest number such that $\mathbb{P}(\theta \in \Omega|y) = 1 - \gamma$. In other words, one may choose the highest posterior density (HPD) set.

Depending on the context, this set may differ numerically from a frequentist confidence interval or region. Given that the posterior density is partly determined by the prior, the posterior for a specific parameter may, for example, be bimodal, in which case an HPD region may consist of two disjoint sets. Even if the HPD set and the frequentist confidence set coincide numerically, their interpretation is quite different. A $(1 - \gamma)100\%$ confidence interval estimator asymptotically contains the true parameter value in $(1 - \gamma)100\%$ of the cases in repeated sampling. This means that the true parameter is either in a specific interval estimate or it is not in this interval. Frequentists cannot say whether the true parameter is inside or outside a specific confidence interval estimate, but only that an interval constructed by their method will include the true value in repeated sampling with probability $1 - \gamma$. In contrast, the Bayesian credible set specifies a region where $(1 - \gamma)100\%$ of the probability mass of the posterior distribution is concentrated. This allows them to make probability statements about the parameter of interest, θ , given the data and the stochastic model of the data. There is no true value of θ for a Bayesian, nor is it considered important what the true value of θ is. The construction of credible sets easily generalizes to smooth functions $h(\theta)$, the posterior of which may be evaluated by simulation.

Testing Statistical Hypotheses and Model Comparison. Hypothesis testing is another important part of frequentist statistics. In frequentist hypothesis testing, the model specified under the null hypothesis is maintained unless there is overwhelming evidence against it. Bayesians are typically not interested in hypothesis testing, but in quantifying the empirical support for alternative models. Typically, Bayesians choose between models based on their posterior odds ratios. Suppose that we want to compare two models M_1 and M_2 , for

example. Then the posterior odds ratio is

$$\frac{g(M_1|\mathbf{y})}{g(M_2|\mathbf{y})} = \frac{g(M_1)}{g(M_2)} \times \frac{f(\mathbf{y}|M_1)}{f(\mathbf{y}|M_2)}, \quad (5.1.4)$$

where $g(M_1)/g(M_2)$ reflects the prior odds and $f(\mathbf{y}|M_1)/f(\mathbf{y}|M_2)$ is the Bayes factor. The Bayes factor is the ratio of the marginal likelihoods or marginal data densities,

$$f(\mathbf{y}|M_j) = \int f(\mathbf{y}|M_j, \boldsymbol{\theta})g(\boldsymbol{\theta})d\boldsymbol{\theta}, \quad j = 1, 2,$$

that are obtained by integrating the parameters out of the likelihood function. If the posterior odds ratio is greater than 1, this is evidence in favor of model M_1 , whereas a posterior odds ratio smaller than 1 favors M_2 . Obviously, if the prior for all models is the same, the ratio of the marginal likelihoods, known as the Bayes factor, determines which model is preferred. More generally, if there are m alternative models that are regarded as a priori equally likely, then a Bayesian solution would be to select the model with the highest marginal likelihood. This procedure is also applicable if nested models are considered, as would be the case, for example, in choosing between VAR models of different lag orders.¹

Note that Bayesian model comparison and frequentist hypothesis testing have quite different interpretations. Frequentist hypothesis testing treats the parameter vector as fixed. A rejection of the null hypothesis in favor of the alternative hypothesis occurs when the sample gives rise to a value of the test statistic that exceeds the critical value at the chosen significance level. The choice of the significance level reflects the user's definition of reasonable doubt. If there is no evidence against the null hypothesis beyond a reasonable doubt, the test is not informative. Thus, the choice of the null hypothesis matters. The null hypothesis is always protected from rejection rather than both hypotheses being treated symmetrically.

Although one could make the case for choosing model M_1 over M_2 if the posterior odds ratio is much larger than 1, implementing this rule in practice would require choosing a threshold value beyond which a Bayesian would choose M_1 for further analysis and discard M_2 . Choosing such a threshold would be analogous to choosing a significance level in frequentist hypothesis testing because eliminating one model from further consideration corresponds to rejecting the null hypothesis in frequentist testing. Thus, Bayesian model comparison is more akin to model selection procedures in frequentist statistics (such as the information criteria discussed in Chapter 2) than

¹ Even some subsets of the parameter space having measure zero is not a problem for Bayesian analysis, as long as the prior probability of the parameters is nonzero, given that models with zero prior probability are of no interest from a Bayesian perspective.

to classical hypothesis testing in that it treats all models under consideration symmetrically.

More generally, many Bayesians view all attempts at choosing between alternative models as misguided. A common alternative is Bayesian model averaging. Consider m candidate models M_1, \dots, M_m . If interest centers on $\mathbb{E}(h(\theta))$, for example, one averages over the candidate models such that

$$\mathbb{E}(h(\theta)) = \sum_{i=1}^m \mathbb{E}(h(\theta)|\mathbf{y}, M_i)g(M_i|\mathbf{y}). \quad (5.1.5)$$

Alternatively, a Bayesian may select the median model. Another choice would be to select the a posteriori most likely model, i.e., the model with the largest $g(M_i|\mathbf{y})$, $i = 1, \dots, m$. Each of these choices reflects a different loss function.

5.1.3 Simulating the Posterior Distribution

Bayesian inference is based on the posterior distribution. That distribution may not be available in closed form, but often can be simulated using numerical methods. In this subsection some numerical algorithms are discussed that are useful for this purpose.

Typically, the objective is to generate random draws $\theta^{(1)}, \dots, \theta^{(n)}$ from the posterior distribution of the parameters or some function of these parameters such as the impulse responses associated with a VAR model. Generating draws from potentially complicated and high-dimensional distributions is an important part of Bayesian analysis. Suppose that we are interested in the expectation of $h(\theta)$ which may be a vector or a scalar function. Laws of large numbers suggest that we can obtain a good approximation to this quantity by drawing at random $\theta^{(1)}, \dots, \theta^{(n)}$ from the posterior distribution and noting that

$$\frac{1}{n} \sum_{i=1}^n h(\theta^{(i)}) \xrightarrow{a.s.} \mathbb{E}(h(\theta)|\mathbf{y}) = \int h(\theta)g(\theta|\mathbf{y})d\theta. \quad (5.1.6)$$

The approximation precision increases with the number n of random draws.

Simulating the posterior distribution is difficult when this distribution is nonstandard or unknown. Next, we discuss a range of tools designed for such situations. We only outline the general ideas to enable the reader to determine which algorithm is suitable in a specific situation, without providing the full details. Some of these methods are used in later chapters in the context of specific applications. Further details can be found, for example, in Canova (2007, section 9.5) or Geweke and Whiteman (2006, section 3).

Direct Sampling. If the joint posterior distribution of all parameters is from a known distribution family, random samples can be generated easily using the random number generators provided in standard software packages. A case in

point is Gaussian posterior distributions. Unfortunately, in practice, often the posterior distribution is not from a known family, in which case more sophisticated sampling techniques are required.

Acceptance Sampling. Sometimes the posterior density is not available in closed form, but what is known is a function that is proportional to the posterior density. A function that is proportional to a given density function, but does not integrate to one, is called a kernel of this density function. For example, the product of the prior and the likelihood function, $g(\boldsymbol{\theta})l(\boldsymbol{\theta}|\mathbf{y})$, is a kernel of the posterior density. It is not a density because the area under the function usually does not integrate to one. Denote the kernel of the posterior density by $g^*(\boldsymbol{\theta}|\mathbf{y})$.

If the underlying distribution is difficult to sample from, we may choose a density $g^{AS}(\boldsymbol{\theta})$ that is easy to sample from and that satisfies $g^{AS}(\boldsymbol{\theta}) > 0$ whenever $g^*(\boldsymbol{\theta}|\mathbf{y}) > 0$. The function $g^{AS}(\boldsymbol{\theta})$ is called the source density. To ensure strictly positive values of the source density whenever $g^*(\boldsymbol{\theta}|\mathbf{y}) > 0$, one may, for example, choose a normal density for $g^{AS}(\boldsymbol{\theta})$ because this density is strictly positive on the whole real line. The maximum (or supremum) of the ratio of the posterior kernel and the source density is defined as

$$\varrho = \sup_{\boldsymbol{\theta} \in \{\boldsymbol{\theta} | g^*(\boldsymbol{\theta}|\mathbf{y}) > 0\}} \frac{g^*(\boldsymbol{\theta}|\mathbf{y})}{g^{AS}(\boldsymbol{\theta})}.$$

Then the i^{th} sample value, $\boldsymbol{\theta}^{(i)}$, from the posterior can be obtained by proceeding as follows:

- Step 1.** Draw a random number u from $\mathcal{U}(0, 1)$, the uniform distribution on the interval $(0, 1)$, and draw $\boldsymbol{\theta}^+$ from the distribution corresponding to $g^{AS}(\boldsymbol{\theta})$.
- Step 2.** Retain $\boldsymbol{\theta}^{(i)} = \boldsymbol{\theta}^+$ if $u < g^*(\boldsymbol{\theta}^+|\mathbf{y})/(\varrho \cdot g^{AS}(\boldsymbol{\theta}^+))$. Otherwise return to Step 1.

It can be shown that this algorithm provides a random sample from the posterior distribution. Clearly, if $g^{AS}(\boldsymbol{\theta}|\mathbf{y}) = g^*(\boldsymbol{\theta}|\mathbf{y})$, then $\varrho = 1$ and

$$\frac{g^*(\boldsymbol{\theta}^+|\mathbf{y})}{\varrho \cdot g^{AS}(\boldsymbol{\theta}^+)} = 1.$$

Thus, we would retain all draws from the source distribution in Step 2. If, however, the source density is very different from the posterior density such that it assumes very small values when $g(\boldsymbol{\theta}|\mathbf{y})$ is large, then the ratio $g^*(\boldsymbol{\theta}^+|\mathbf{y})/(\varrho \cdot g^{AS}(\boldsymbol{\theta}^+))$ tends to be small, and only very few draws are accepted. Hence, a large number of draws from the source distribution may be necessary for generating a sufficiently large number of draws from the posterior. In other words,

although the algorithm is general, it can be computationally costly, and other algorithms may be preferable.

Importance Sampling. Importance sampling avoids the drawback of having to discard many posterior draws. Instead, we retain a suitably weighted average of all posterior draws. This proposal dates back at least to Hammersly and Handcomb (1964) and was first used in the econometrics literature by Kloek and van Dijk (1978). The idea is to generate a sample that facilitates the estimation of the posterior moments of functions of θ . Suppose again that the function of interest is $h(\theta)$ and let $g^S(\theta)$ be a proper source density, defined as a density that approximates $g^*(\theta|y)$ and has the same support. Furthermore, define a weighting function as the ratio of g^* and g^S ,

$$w(\theta) = \frac{g^*(\theta|y)}{g^S(\theta)}. \quad (5.1.7)$$

Then, for $\theta^{(1)}, \dots, \theta^{(n)}$ randomly drawn from the distribution corresponding to $g^S(\theta)$, under general conditions, we obtain

$$\frac{\sum_{i=1}^n h(\theta^{(i)})w(\theta^{(i)})}{\sum_{i=1}^n w(\theta^{(i)})} \xrightarrow{a.s.} \mathbb{E}[h(\theta|y)], \quad (5.1.8)$$

because

$$\int h(\theta)g^*(\theta|y)d\theta = \int h(\theta) \frac{g^*(\theta|y)}{\int g^*(\theta|y)d\theta} d\theta = \frac{\int \frac{h(\theta)g^*(\theta|y)}{g^S(\theta)} g^S(\theta)d\theta}{\int \frac{g^*(\theta|y)}{g^S(\theta)} g^S(\theta)d\theta}.$$

Thus, a properly weighted set of draws from the source distribution can be used to approximate the expectation based on the posterior distribution.

Clearly, the advantage of this algorithm over acceptance sampling is that only a sample of size n has to be drawn. Since the accuracy of the approximation of the expected value of interest improves with the sample size, it is an obvious advantage that all sample values can be used directly. It should be noted, however, that finding a good approximation $g^S(\theta)$ to the posterior density may not be easy. The weights may vary dramatically, undermining the convergence properties of the quantity in expression (5.1.8). Thus, importance sampling may not work well if the posterior distribution is complicated and difficult to sample from. More recent proposals for generating samples from the posterior therefore involve Markov Chain Monte Carlo methods.

Markov Chain Monte Carlo (MCMC) Methods. MCMC methods use a Markov chain algorithm to generate one long sample of the parameter vector, the distribution of which converges to the posterior distribution of the parameters if the chain runs long enough. The draws of the parameter vector are not

independent, however, but serially dependent. Laws of large numbers and central limit theorems for dependent samples can be invoked to justify the use of such samples for posterior inference. It should be understood, however, that the approximation precision for the posterior moment of interest tends to be lower when the posterior is computed from a dependent sample rather than a random sample. Thus, longer samples may be necessary for precise inference. In fact, for the construction of approximately independent samples from the joint posterior, one may want to work only with every m^{th} sampled vector, where m is a sufficiently large number. Moreover, a large number of initial sample values (also known as transients or the burn-in sample) are usually discarded to ensure a close approximation of the posterior. Diagnostic tools for assessing the convergence of the posterior to the target distribution are discussed in Chib (2001), among others. Despite their high computational demands, MCMC methods have become quite popular in recent years because they are often simpler and more computationally efficient than other sampling methods. There are several variations of this approach in the literature that differ in the way they choose the i^{th} draw of θ , denoted $\theta^{(i)}$, given $\theta^{(i-1)}$.

MCMC methods have been known in the literature for a long time. They have become increasingly popular in Bayesian analysis after the publication of an influential article by Gelfand and Smith (1990). More recent introductory expositions are Chib and Greenberg (1995) and Geweke (2005). In the following, the very general Metropolis-Hastings algorithm and the popular Gibbs sampler are presented. The Gibbs sampler can only be used if the posterior can be broken down in a suitable way. It is very convenient and very efficient if the conditions for its use are satisfied.

Metropolis-Hastings Algorithm. The Metropolis-Hastings algorithm is based on a conditional density, $\eta(\theta|\theta^{(i-1)})$. For a given $\theta^{(i-1)}$, a candidate θ^+ is drawn from the distribution corresponding to $\eta(\theta|\theta^{(i-1)})$, and $\theta^{(i)} = \theta^+$ is chosen with probability

$$\min \left\{ \frac{g(\theta^+|y)/\eta(\theta^+|\theta^{(i-1)})}{g(\theta^{(i-1)}|y)/\eta(\theta^{(i-1)}|\theta^+)}, 1 \right\}.$$

Otherwise, $\theta^{(i)} = \theta^{(i-1)}$. In other words, a draw is accepted with probability one if it increases the posterior. Otherwise, it is accepted with a probability less than 1, the precise value of which depends on how much lower the current posterior value is compared with the previous draw. It can be shown that this algorithm converges to the posterior under general conditions.

There are a number of practical questions related to the Metropolis-Hastings algorithm. Of prime importance is, of course, the choice of the conditional density $\eta(\theta|\theta^{(i-1)})$. There are different proposals in the literature. A review can be found in Chib and Greenberg (1995), for example. Another important issue

is the size of the burn-in sample. The question is how many initial draws should be discarded in the construction of the chain. Several criteria are discussed in Cowles and Carlin (1996), among others.

Clearly, one would not want to use this algorithm if computationally simpler methods exist. In practice, the Metropolis-Hastings sampler is only used for nonstandard problems when there is no alternative. A case in point is the estimation of VAR parameters that cannot be estimated using the Gibbs sampler of either Waggoner and Zha (2003) or Villani (2009). For example, Giannone, Lenza, and Primiceri (2015) use the Metropolis-Hastings algorithm because of the complicated form of the kernel of the posterior distribution of their hyperparameters. Likewise many time-varying coefficient VAR models require Metropolis-Hastings sampling.

Gibbs Sampler. If the posterior distribution of θ is difficult to sample from, the problem often may be simplified by partitioning the parameter vector such that the conditional posterior of one of the subvectors, given the remaining elements, has a known, conventional distribution from which samples can be drawn easily. Consider the simplest case where θ can be partitioned as $\theta = (\theta'_{(1)}, \theta'_{(2)})'$ such that $g(\theta_{(1)}|\theta_{(2)}, \mathbf{y})$ and $g(\theta_{(2)}|\theta_{(1)}, \mathbf{y})$ correspond to known distributions that can be simulated easily. This case arises in VAR models, for example, if we partition the parameters into the VAR coefficients and the residual covariance matrix. In this situation, for given $\theta_{(2)}^{(i-1)}$ we may choose $\theta_{(1)}^{(i)}$ by drawing from $g(\theta_{(1)}|\theta_{(2)}^{(i-1)}, \mathbf{y})$ and choose $\theta_{(2)}^{(i)}$ by drawing from $g(\theta_{(2)}|\theta_{(1)}^{(i)}, \mathbf{y})$, starting from some initial value $\theta_{(2)}^{(0)}$. This simple algorithm converges to the joint posterior under general conditions.

The Gibbs sampler can be generalized in several ways. First, a generalization to the case where θ needs to be partitioned into more than two subvectors to obtain standard conditional posterior distributions is straightforward. It should be noted, however, that the Gibbs sampler works very well only when the posterior distributions of the different subvectors are independent or at least not strongly correlated. This feature should be taken into account in grouping the parameters. Second, it may happen that some of the conditional posterior distributions are not from a standard family. As long as at least one conditional distribution is obtained that is easy to sample from, the Gibbs sampler is worth considering. The other conditional distributions may then be approximated by a Metropolis-Hastings step.

5.2 Priors for Reduced-Form VAR Parameters

In Bayesian analysis an important issue is the specification of the prior for the parameters of interest. Often a prior is specified that simplifies the posterior analysis. In particular, it is convenient to specify the prior such that the

posterior is from a known family of distributions. If, for a given likelihood, the prior is chosen such that the posterior has the same functional form as the prior, it is called a conjugate prior. If the conjugate prior is from the same distribution family as the likelihood, then it is called a natural conjugate prior. For example, it is shown in Section 5.2.2 that if the likelihood is Gaussian and the prior is also normal, then the posterior again has a normal distribution. The Litterman or Minnesota prior is discussed as a special case. The natural conjugate prior for all the parameters is known as the Gaussian-inverse Wishart prior. It is considered in Section 5.2.4, whereas the more computationally convenient independent Gaussian-inverse Wishart prior is considered in Section 5.2.5.

When using priors from a known family of distributions, it is still necessary to specify at least some of the parameters of the prior distribution. This task is often made easier by imposing additional structure on the prior, reducing the number of parameters to be chosen to a handful of so-called hyperparameters. We discuss this idea in Section 5.2.1. In Sections 5.2.2–5.2.5 we consider popular choices for the prior distribution of the reduced-form VAR parameters and discuss how these distributions may be parameterized with the help of a small set of hyperparameters.

5.2.1 General Procedures for Choosing the Parameters of Prior Densities

In practice, it is often prohibitively difficult to fully specify the prior distribution. It is therefore common to impose additional structure on a given family of prior distributions so as to reduce the number of parameters to be specified by the user to a small number of so-called hyperparameters. Let γ denote the vector of hyperparameters such that $g(\theta) = g_\gamma(\theta)$. Often γ is chosen such that the implied VAR model yields accurate out-of-sample forecasts (see Doan, Litterman, and Sims 1984; Litterman 1986). Alternatively, Bańbura, Giannone, and Reichlin (2010) suggest to choose these hyperparameters based on the in-sample fit of the model.

Yet another proposal for choosing the hyperparameters was made by Giannone, Lenza, and Primiceri (2015). If one views the prior as being conditioned on the hyperparameters γ , $g_\gamma(\theta) = g(\theta|\gamma)$, then the prior can be regarded as a hierarchical prior (see Koop 2003). Suppose that the prior density for γ is $g(\gamma)$. Then the posterior is

$$g^*(\gamma) \propto h(y|\gamma)g(\gamma),$$

where the sample density with respect to the hyperparameters is obtained as

$$h(y|\gamma) = \int f(y|\theta, \gamma)g(\theta|\gamma)d\theta.$$

This expression is also known as the marginal likelihood because the parameters of interest, θ , are integrated out. If an improper uniform prior, $g(\boldsymbol{\gamma}) = \text{constant}$, is specified, then the posterior of the hyperparameters is equal to the marginal likelihood, and it makes sense to choose the hyperparameters such that $h(\mathbf{y}|\boldsymbol{\gamma})$ is maximized. Of course, strictly speaking, an improper prior does not qualify as a prior density because for an unbounded parameter space a constant prior does not integrate to one.

Giannone, Lenza, and Primiceri (2015) stress two advantages of this approach. First, under certain conditions maximizing the marginal likelihood results in optimal out-of-sample forecasts (see also Geweke 2001; Geweke and Whiteman 2006). Second, they point out that their procedure also can be justified from a frequentist point of view.

5.2.2 Normal Prior for the VAR Parameters for Given Σ_u

An early approach to specifying the prior for the VAR slope parameters takes the innovation variance Σ_u as given. In practice, we may replace the unknown Σ_u by its LS or ML estimate (see, e.g., Litterman 1986). Understanding this approach is also useful for expository purposes.

Consider a normally distributed K -dimensional VAR(p) process y_t of the form

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where $u_t \sim \mathcal{N}(0, \Sigma_u)$ is a Gaussian white noise error term. For $t = 1, \dots, T$, the model can be written in matrix notation as

$$Y = AZ + U, \quad (5.2.1)$$

where $Y \equiv [y_1, \dots, y_T]$, $A \equiv [v, A_1, \dots, A_p]$ and $Z \equiv [Z_0, \dots, Z_{T-1}]$ with $Z_{t-1} \equiv (1, y'_{t-1}, \dots, y'_{t-p})'$. Vectorizing the matrix expression (5.2.1), one obtains

$$\mathbf{y} = (Z' \otimes I_K)\boldsymbol{\alpha} + \mathbf{u}, \quad (5.2.2)$$

where $\boldsymbol{\alpha} \equiv \text{vec}(A)$, $\mathbf{y} \equiv \text{vec}(Y)$, and $\mathbf{u} \equiv \text{vec}(U)$. Next we discuss two alternative ways for expressing a normal prior for $\boldsymbol{\alpha}$ under the assumption that the white noise covariance matrix Σ_u is known.

Prior Distribution. Suppose the prior distribution of $\boldsymbol{\alpha}$ is multivariate normal with known mean $\boldsymbol{\alpha}^*$ and covariance matrix $V_{\boldsymbol{\alpha}}$. We write

$$\boldsymbol{\alpha} \sim \mathcal{N}(\boldsymbol{\alpha}^*, V_{\boldsymbol{\alpha}})$$

and, hence, the prior density is

$$g(\boldsymbol{\alpha}) = \left(\frac{1}{2\pi} \right)^{K(Kp+1)/2} |V_{\boldsymbol{\alpha}}|^{-1/2} \exp \left[-\frac{1}{2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)' V_{\boldsymbol{\alpha}}^{-1} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*) \right]. \quad (5.2.3)$$

Combining this information with the sample information summarized in the Gaussian likelihood function,

$$\begin{aligned} l(\boldsymbol{\alpha}|\mathbf{y}) &= \left(\frac{1}{2\pi} \right)^{KT/2} |I_T \otimes \Sigma_u|^{-1/2} \\ &\times \exp \left[-\frac{1}{2} [\mathbf{y} - (Z' \otimes I_K) \boldsymbol{\alpha}]' (I_T \otimes \Sigma_u^{-1}) [\mathbf{y} - (Z' \otimes I_K) \boldsymbol{\alpha}] \right], \end{aligned}$$

yields the posterior density

$$\begin{aligned} g(\boldsymbol{\alpha}|\mathbf{y}) &\propto g(\boldsymbol{\alpha}) l(\boldsymbol{\alpha}|\mathbf{y}) \\ &\propto \exp \left\{ -\frac{1}{2} \left[[V_{\boldsymbol{\alpha}}^{-1/2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)]' [V_{\boldsymbol{\alpha}}^{-1/2} (\boldsymbol{\alpha} - \boldsymbol{\alpha}^*)] \right. \right. \\ &\quad + \left. \left. \{ (I_T \otimes \Sigma_u^{-1/2}) \mathbf{y} - (Z' \otimes \Sigma_u^{-1/2}) \boldsymbol{\alpha} \}' \right. \right. \\ &\quad \left. \left. \times \{ (I_T \otimes \Sigma_u^{-1/2}) \mathbf{y} - (Z' \otimes \Sigma_u^{-1/2}) \boldsymbol{\alpha} \} \right] \right\}. \quad (5.2.4) \end{aligned}$$

Defining

$$w = \begin{bmatrix} V_{\boldsymbol{\alpha}}^{-1/2} \boldsymbol{\alpha}^* \\ (I_T \otimes \Sigma_u^{-1/2}) \mathbf{y} \end{bmatrix} \text{ and } W = \begin{bmatrix} V_{\boldsymbol{\alpha}}^{-1/2} \\ Z' \otimes \Sigma_u^{-1/2} \end{bmatrix},$$

the exponent in (5.2.4) can be rewritten as

$$\begin{aligned} &- \frac{1}{2} (w - W \boldsymbol{\alpha})' (w - W \boldsymbol{\alpha}) \\ &= -\frac{1}{2} [(\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}})' W' W (\boldsymbol{\alpha} - \bar{\boldsymbol{\alpha}}) + (w - W \bar{\boldsymbol{\alpha}})' (w - W \bar{\boldsymbol{\alpha}})], \quad (5.2.5) \end{aligned}$$

where

$$\begin{aligned} \bar{\boldsymbol{\alpha}} &= (W' W)^{-1} W' w \\ &= [V_{\boldsymbol{\alpha}}^{-1} + (Z Z' \otimes \Sigma_u^{-1})]^{-1} [V_{\boldsymbol{\alpha}}^{-1} \boldsymbol{\alpha}^* + (Z \otimes \Sigma_u^{-1}) \mathbf{y}]. \quad (5.2.6) \end{aligned}$$

Note that if the precision matrix $V_{\boldsymbol{\alpha}}^{-1}$ were zero, the expression for $\bar{\boldsymbol{\alpha}}$ would simplify to that of the unrestricted LS estimator in equation (2.3.2) by

noting that

$$\begin{aligned}(ZZ' \otimes \Sigma_u^{-1})^{-1} (Z \otimes \Sigma_u^{-1}) \mathbf{y} &= ((ZZ')^{-1} Z \otimes I_K) \text{vec}(Y) \\ &= \text{vec}(YZ'(ZZ')^{-1}).\end{aligned}$$

The second term on the right-hand side of (5.2.5) does not contain α and therefore is a constant. Thus,

$$g(\alpha | \mathbf{y}) \propto \exp \left[-\frac{1}{2} (\alpha - \bar{\alpha})' \bar{\Sigma}_\alpha^{-1} (\alpha - \bar{\alpha}) \right],$$

with $\bar{\alpha}$ as given in (5.2.6) and

$$\bar{\Sigma}_\alpha = (W'W)^{-1} = [V_\alpha^{-1} + (ZZ' \otimes \Sigma_u^{-1})]^{-1}. \quad (5.2.7)$$

This posterior density is easily recognizable as the density of a multivariate normal distribution with mean $\bar{\alpha}$ and covariance matrix $\bar{\Sigma}_\alpha$. In other words, the posterior distribution of α is $\mathcal{N}(\bar{\alpha}, \bar{\Sigma}_\alpha)$. This distribution may be used for inference about α . Sampling from this posterior distribution is particularly easy because the distribution is of a known form that is easy to draw from.

An Alternative Representation of the Prior Distribution. The prior information on α can equivalently be written as

$$C\alpha = c + e \quad \text{with} \quad e \sim \mathcal{N}(0, I), \quad (5.2.8)$$

where C is a fixed matrix and c is a fixed vector. If C is a $K(Kp + 1) \times K(Kp + 1)$ nonsingular matrix, this expression implies

$$\alpha \sim \mathcal{N}(C^{-1}c, C^{-1}C^{-1}), \quad (5.2.9)$$

which is just the normal prior specified earlier with mean $C^{-1}c$ and covariance matrix $(C'C)^{-1}$. Using (5.2.6), the posterior mean can be expressed as

$$\bar{\alpha} = [C'C + (ZZ' \otimes \Sigma_u^{-1})]^{-1} [C'c + (Z \otimes \Sigma_u^{-1}) \mathbf{y}]. \quad (5.2.10)$$

Note that under this prior specification there is no need to invert the potentially large covariance matrix V_α . Another practical advantage of this form is that it does not require the inversion of C . In fact, C does not have to be a square matrix.

It is also worth mentioning that the estimator $\bar{\alpha}$ in (5.2.10) is precisely the GLS estimator obtained from a regression model

$$\begin{bmatrix} \mathbf{y} \\ c \end{bmatrix} = \begin{bmatrix} Z' \otimes I_K \\ C \end{bmatrix} \alpha + \varepsilon, \quad \varepsilon \sim \left(0, \begin{bmatrix} I_T \otimes \Sigma_u & 0 \\ 0 & I \end{bmatrix} \right), \quad (5.2.11)$$

as pointed out by Theil (1963). Thus, in this case, imposing the prior amounts to extending the sample. As in the frequentist setting, adding more observations tends to reduce the variance of the GLS estimator compared with the unrestricted LS estimator, and, hence, increases the estimation efficiency.

In order to make these concepts operational, the prior mean α^* and the prior covariance matrix V_α (or, equivalently, C and c) must be specified. If no specific prior knowledge about the parameters is available, then one may use a diffuse Gaussian prior by choosing the prior variances very large, so that V_α^{-1} (also known as the precision matrix) becomes very small. In that case the posterior mean is seen to be close to the LS estimator (see equation (5.2.6)).

The posterior mean can also be interpreted as a shrinkage estimator of α with the degree of shrinkage determined by V_α^{-1} . If $V_\alpha^{-1} = 0$ is used, the posterior mean reduces to the LS estimator. Of course, in that case Bayesian analysis loses its potential advantage of reducing the variance of the parameter estimates. In practice, there are a number of alternative ways of specifying nontrivial precision matrices. The next subsections discuss the most common choices.

Practical Considerations. An important concern in practice is that priors may be inadvertently informative about the parameters of interest. Hence, there has been interest in priors that are uninformative. Ideally, assigning equal prior probability to all possible parameter values would avoid such distortions because in that case the prior density is just a constant that cancels from the posterior density due to the requirement that the posterior density integrates to one. Unfortunately, there is no probability density that is constant over the entire Euclidean space. Any nonzero, positive constant would integrate to infinity over the whole space. Hence such a flat prior is not a proper prior. In other words, a truly uninformative prior for the VAR parameters does not exist.

An alternative is a Gaussian prior with V_α^{-1} approaching zero. Such a prior is also known as a diffuse prior for the VAR parameters. Note, however, that such priors, although they do not seem to restrict the slope parameters much, may still be unintentionally informative for nonlinear functions of the parameters such as impulse responses. It should always be kept in mind that a prior that seems uninformative in one dimension tends to be informative in other dimensions.

Given the difficulty of agreeing on subjective priors, it has become common practice to choose the priors for the VAR parameters by convention. For example, priors may be chosen such that the associated models have certain

desirable properties. Typically, in the VAR literature, priors have been chosen based on the forecast accuracy of the implied estimated models. One such prior that has been quite popular in applied work is discussed in the next subsection.

5.2.3 The Original Minnesota Prior

Litterman (1986) and Doan, Litterman, and Sims (1984) propose a specific Gaussian prior for the parameters of a VAR model that is often referred to as the Minnesota prior or the Litterman prior. The original proposal shrinks the VAR estimates toward a multivariate random walk model. This practice has been found useful in forecasting many persistent economic time series. The proposal is to specify the prior mean and covariance matrix as follows:

- In each equation, set the prior mean of the first lag of the dependent variable to one and set the prior mean of all other slope coefficients to zero. In other words, if the prior means were the true parameter values, each variable would follow a random walk.
- Set the prior variances of the intercept terms to infinity and the prior variance of the i^{th} element of A_l , denoted $a_{ij,l}$, to

$$v_{ij,l} = \begin{cases} (\lambda/l)^2 & \text{if } i = j, \\ (\lambda\theta\sigma_i/l\sigma_j)^2 & \text{if } i \neq j, \end{cases}$$

where λ is the prior standard deviation of $a_{ii,1}$, where $0 < \theta < 1$ controls the relative tightness of the prior variance in the other lags in a given equation compared to the own lags (with a smaller θ increasing the relative tightness of the other lags), and where σ_i^2 is the i^{th} diagonal element of Σ_u .

For example, in a bivariate VAR(2) model with all slope parameters evaluated at their prior mean, we would have

$$\begin{aligned} y_{1t} &= 0 + 1 \cdot y_{1,t-1} + 0 \cdot y_{2,t-1} + 0 \cdot y_{1,t-2} + 0 \cdot y_{2,t-2} + u_{1t}, \\ &\quad (\infty) \quad (\lambda) \quad (\lambda\theta\sigma_1/\sigma_2) \quad (\lambda/2) \quad (\lambda\theta\sigma_1/2\sigma_2) \\ y_{2t} &= 0 + 0 \cdot y_{1,t-1} + 1 \cdot y_{2,t-1} + 0 \cdot y_{1,t-2} + 0 \cdot y_{2,t-2} + u_{2t}. \\ &\quad (\infty) \quad (\lambda\theta\sigma_2/\sigma_1) \quad (\lambda) \quad (\lambda\theta\sigma_2/2\sigma_1) \quad (\lambda/2) \end{aligned}$$

Here the numbers in parentheses are the prior standard deviations. Each of the two equations specifies a random walk prior mean for the dependent variables. The nonzero prior standard deviations reflect the uncertainty regarding the validity of that model. The standard deviations decline with increasing lag length because more recent lags are assumed to be more likely to have nonzero values. The standard deviations for the intercept terms are set to infinity to capture our ignorance about the actual values of these parameters. Also, the

prior distribution imposes independence across the parameters. Therefore, V_α is diagonal. Its inverse is

$$V_\alpha^{-1} = \begin{bmatrix} 0 & & & & & & & \\ 0 & 0 & & & & & & \\ & \frac{1}{\lambda^2} & & & & & & \\ & & \frac{\sigma_1^2}{(\lambda\theta\sigma_2)^2} & & & & & \\ & & & \frac{\sigma_2^2}{(\lambda\theta\sigma_1)^2} & & & & \\ & & & & \frac{1}{\lambda^2} & & & \\ & 0 & & & & \frac{2^2}{\lambda^2} & & \\ & & & & & & \frac{2^2\sigma_1^2}{(\lambda\theta\sigma_2)^2} & \\ & & & & & & & \frac{2^2\sigma_2^2}{(\lambda\theta\sigma_1)^2} \\ & & & & & & & \\ & & & & & & & \frac{2^2}{\lambda^2} \end{bmatrix},$$

where 0 is also substituted for the inverse of the infinite prior standard deviation of the intercepts.

In terms of expression (5.2.9), the prior for the slope parameters may equivalently be specified by choosing

$$C = \begin{bmatrix} 0 & 0 & \frac{1}{\lambda} & & & & & 0 & & \\ 0 & 0 & & \frac{\sigma_1}{\lambda\theta\sigma_2} & & & & & & \\ 0 & 0 & & & \frac{\sigma_2}{\lambda\theta\sigma_1} & & & & & \\ 0 & 0 & & & & \frac{1}{\lambda} & & & & \\ 0 & 0 & & 0 & & & \frac{2}{\lambda} & & & \\ 0 & 0 & & & & & & \frac{2\sigma_1}{\lambda\theta\sigma_2} & & \\ 0 & 0 & & & & & & & \frac{2\sigma_2}{\lambda\theta\sigma_1} & \\ 0 & 0 & & & & & & & & \frac{2}{\lambda} \end{bmatrix} \text{ and}$$

$$c = \frac{1}{\lambda} \begin{bmatrix} 1 \\ 0 \\ 0 \\ 1 \\ 0 \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

where $c = \frac{1}{\lambda}(\text{vec}(I_2)', 0_{1 \times 4})'$ and C is an 8×10 matrix with two leading columns of zeros and the square roots of the reciprocals of the diagonal elements of V_α on the main diagonal of the remaining 8×8 block. Note that in this alternative representation no prior is specified for the intercept.

Practical Issues. The crucial advantage of the Minnesota prior is that it reduces the problem of specifying a high-dimensional prior distribution to one of selecting two parameters by imposing additional structure on the prior. In specifying the Minnesota prior, the user has to choose only the two hyperparameters λ and θ . The parameter λ controls the overall prior variance of all VAR coefficients, whereas θ controls the tightness of the variances of the coefficients of lagged variables other than the dependent variable in a given equation. Roughly speaking, θ specifies the fraction of the prior standard deviation λ attached to the coefficients of other lagged variables. A value of θ close to one implies that all coefficients of lag 1 have about the same prior variance except for a scaling factor intended to capture differences in the variability of each variable. For example, Litterman (1986) finds that $\theta = 0.3$ and $\lambda = 0.2$ works well when using a Bayesian VAR model for forecasting U.S. macroeconomic variables. For given θ , the shrinkage is determined by λ . Therefore λ is often referred to as the shrinkage parameter. A smaller λ implies a stronger shrinkage toward the prior mean.²

There are also a number of other practical problems that have to be addressed in working with the Minnesota prior. For example, the assumption of a known Σ_u is unrealistic in practice. In a strict Bayesian approach, a prior pdf has to be specified for the elements of Σ_u . This possibility is discussed in the next section. A simple alternative is to replace Σ_u by its LS estimator or its ML estimator,

$$\tilde{\Sigma}_u = Y(I_T - Z'(ZZ')^{-1}Z)Y'/T,$$

based on the full sample T .

Given that the computation of $\bar{\alpha}$ requires the inversion of the potentially rather large matrix $V_{\alpha}^{-1} + (ZZ' \otimes \Sigma_u^{-1})$, Bayesian estimation based on the Minnesota prior has often been performed for each of the K equations of the system separately. In that case,

$$\bar{a}_k = [V_k^{-1} + \sigma_k^{-2}ZZ']^{-1} (V_k^{-1}a_k^* + \sigma_k^{-2}Zy_{(k)})$$

is the estimator for the parameters a_k of the k^{th} equation, where $y_{(k)} \equiv (y_{k1}, \dots, y_{kT})'$. In other words, a'_k is the k^{th} row of $A = [v, A_1, \dots, A_p]$. Here V_k denotes the prior covariance matrix of a_k and a_k^* is its prior mean. The unknown σ_k^2 may be replaced by the k^{th} diagonal element of $\tilde{\Sigma}_u$.

Shrinking the parameters of models of macroeconomic variables toward a random walk is only plausible for economic time series with stochastic trends.

² Rather than selecting λ and θ based on rules of thumb, one could instead treat these hyperparameters as endogenous and set them to the values that maximize the marginal likelihood, as proposed by Giannone, Lenza, and Primiceri (2015).

When working with stationary variables, the VAR parameters may be shrunk toward zero instead, as proposed by Lütkepohl (1991, section 5.4) and implemented, for example, in Baumeister and Kilian (2012). In that case, mean-adjusting the data before fitting a VAR model may be useful to avoid having to specify a prior for the intercept term. Villani (2009) makes the case that a mean-adjusted model form may have advantages if prior information is to be imposed on the steady state of the variables.

Many other modifications of the Minnesota prior have been proposed, depending on the needs of the researcher (see, for example, Kadiyala and Karlsson 1997; Sims and Zha 1998; Waggoner and Zha 2003; Bańbura, Giannone, and Reichlin 2010; Karlsson 2013). The main advantage of the Minnesota prior is that it results in a simple analytically tractable normal posterior distribution, which explains why it has remained popular over the years, despite some disadvantages. There are alternatives, however. For example, Sims and Zha (1998), building on the representation (5.2.11), proposed imposing prior restrictions on the structural VAR parameters rather than the reduced-form VAR parameters (see Chapter 12).

Cointegration and Near Unit Roots. One potential disadvantage of the Litterman prior is that even if all variables have stochastic trends, it is not clear that shrinking toward a multivariate random walk as in the Litterman prior is optimal because there may be cointegration between the variables (see Chapter 3). This approach can be rationalized on the grounds that exact unit roots are events of probability zero in standard Bayesian analysis. Hence, there is no reason to pay special attention to cointegration relations from a Bayesian point of view. Nevertheless the importance of the concept of cointegration and of VECMs in frequentist analysis has prompted some Bayesians to develop alternative priors that explicitly refer to the parameters of the VECM form of the VAR.

Consider expressing the VAR model in the VEC representation introduced in Chapter 3,

$$\Delta y_t = \nu + \boldsymbol{\Pi} y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots + \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t,$$

where $\boldsymbol{\Pi} = -(I_K - A_1 - \cdots - A_p)$. A prior that shrinks $\boldsymbol{\Pi}$ to zero in the limit reduces the VECM to a VAR model in differences. Such a prior may also be suitable if there are near unit roots (see Chapter 3). A specific prior of this type, referred to as the sum-of-coefficients prior in Doan, Litterman, and Sims (1984), is implemented by augmenting the observations similar to expression (5.2.11). For the sum-of-coefficients prior we augment Y and Z in model (5.2.1) by

$$Y_* = \text{diag}(\mu_1, \dots, \mu_K)/\tau$$

and

$$Z_* = \left[\begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \otimes \text{diag}(\mu_1, \dots, \mu_K) / \tau \right]^{\top},$$

respectively, and consider the vectorized version of the model

$$[Y, Y_*] = A[Z, Z_*] + [U, \mathcal{E}]$$

instead of (5.2.11). If the shrinkage parameter τ is small, the prior shrinks the posterior mean of Π to zero. If τ is very large, the posterior mean of Π is close to the LS estimator. The μ_k are supposed to capture the potentially different levels of the y_{kt} . In practice, the sample mean is used as a proxy for μ_k , although that approach is not, strictly speaking, Bayesian (see Baíbura, Giannone, and Reichlin 2010). Of course, shrinkage may also be imposed on the Γ_i parameters. Such a prior can be imposed by adding further dummy observations, as described earlier, based on expression (5.2.11). For further motivation and discussion of the sum-of-coefficients prior, see Sims and Zha (1998).

Bayesian analysis of VECMs is discussed, for example, in Kleibergen and van Dijk (1994), Kleibergen and Paap (2002), Strachan (2003), and Strachan and Inder (2004). Surveys with many additional references are Koop, Strachan, van Dijk, and Villani (2005) and Karlsson (2013).

An Empirical Illustration. To illustrate the use of the Minnesota prior, we consider a model including quarterly U.S. GDP deflator inflation ($\Delta\pi_t$), the seasonally adjusted unemployment rate (ur_t), and the yield on the 3-month Treasury bills (r_t) for the period 1953q1 – 2006q3, as used by Koop and Korobilis (2009).³ The time series are plotted in Figure 5.1. All three series exhibit considerable persistence, so using the Minnesota prior with shrinkage to a random walk makes sense. To allow for the possibility that the time series have no unit roots, we alternatively consider shrinking all parameters to zero by means of a white noise prior mean.

We consider a VAR(4) model with intercept and impose a conventional Minnesota prior. Following the example of some earlier studies, the unknown error variances are replaced by estimates obtained from fitting univariate AR(4) models to the individual model variables. No estimates of the error covariances are required for the specification of the prior. In constructing the

³ The data are available on Gary Koop's webpage at http://personal.strath.ac.uk/gary.koop/bayes_matlab_code_by_koop_and_korobilis.html. This webpage in addition provides a set of Matlab code for BVAR analysis, modified versions of which have been used for this empirical illustration.

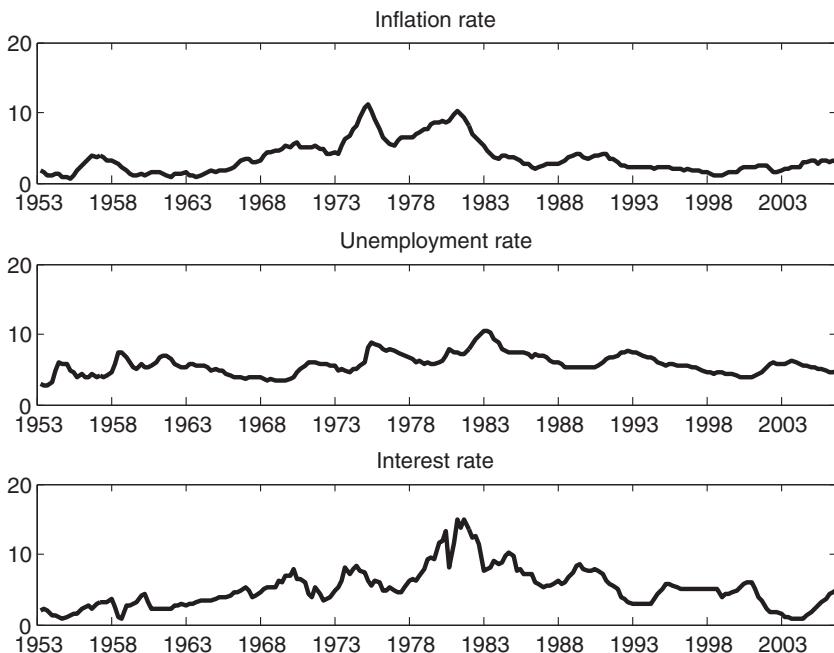


Figure 5.1. Quarterly U.S. inflation, unemployment rate, and interest rate for 1953q1–2006q3.

posterior, the error covariance matrix is treated as known and replaced by its LS estimate.

Figure 5.2 illustrates the impact of alternative specifications of the Minnesota prior on the posterior density of selected structural impulse responses (see Chapter 4). The structural responses are obtained by imposing a recursive structure on the impact multiplier matrix with the variables ordered as $y_t = (\Delta\pi_t, ur_t, r_t)'$. In particular, the interest rate is ordered last, so the shock to the interest rate equation may be interpreted as a monetary policy shock with no contemporaneous effect on inflation and unemployment (see Chapter 8). Figure 5.2 focuses on the responses of inflation to an unexpected increase in the interest rate (which represents a contractionary monetary policy shock). Following common practice, the figure plots the 10%, 50%, and 90% quantiles of the draws from the posterior distributions of the individual impulse response coefficients.

Figure 5.2 illustrates how the choice of the prior affects the structural impulse response estimates. The random walk prior mean is used for generating the panels on the left, and the white noise prior mean is used for the panels on the right. Obviously it makes a difference which prior mean specification is used, but the choice of the hyperparameters λ and θ , which control

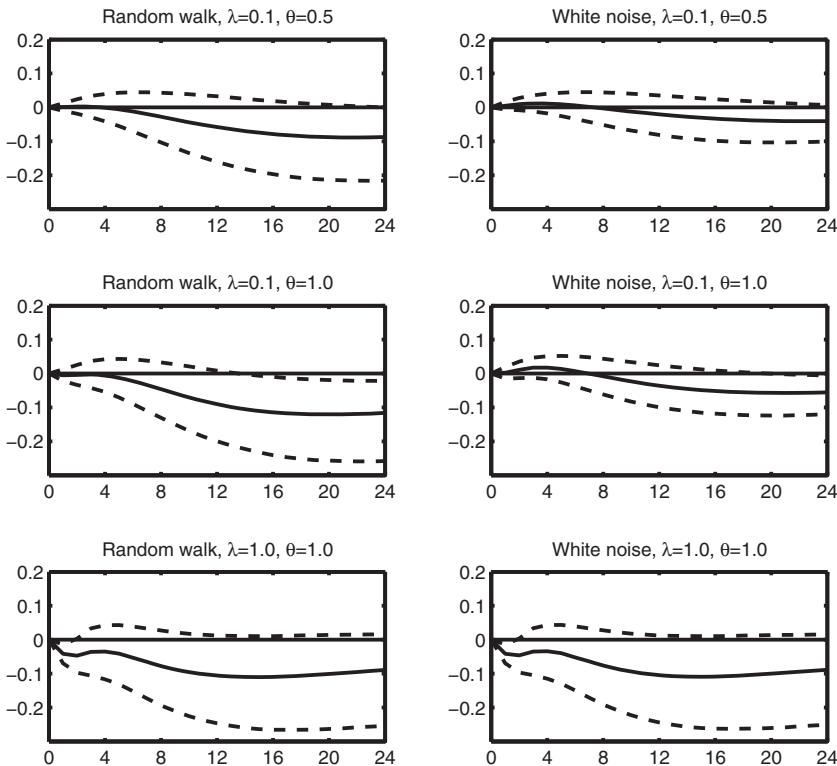


Figure 5.2. Simulated quantiles of inflation responses to monetary policy shocks for different Minnesota priors (pointwise median and 10% and 90% quantiles of the posterior distribution computed from 10,000 draws). The prior standard deviation for the constant terms is set to 1000 throughout; the standard deviations of the innovations are replaced by estimates obtained from fitting univariate AR(4) models to each model variable.

the prior variance, is also important. A tight prior variance about the random walk prior mean results in smaller bands around the pointwise medians than a prior with larger λ and θ parameters. Shrinkage to a white noise process makes an even larger difference, unless the prior variance is large ($\lambda = 1.0, \theta = 1.0$). In the first two panels on the right, inflation increases in response to a contractionary monetary policy shock. This phenomenon is usually referred to as the price puzzle and has been observed in many structural VAR studies. It is often attributed to omitted variables. This example illustrates that this puzzle can also be an artifact of the choice of the prior. In this example, shrinking the slopes to a white noise mean implies a very different posterior than shrinking to the random walk mean. More generally, one could allow for different

prior means in each equation of the VAR model, depending on the order of integration of the individual variables.

5.2.4 The Natural Conjugate Gaussian-Inverse Wishart Prior

So far we have replaced the unknown Σ_u by an estimate. This approach can be improved upon by specifying a prior not only for the slope parameters but also for Σ_u . The prior distribution of the innovation covariance matrix must satisfy the constraint that Σ_u is positive definite. This section discusses such a prior and derives the corresponding posterior.

Let $x_i \sim \mathcal{N}(0, \Sigma_x)$, $i = 1, \dots, n$, be K -dimensional independent, identically distributed normal random vectors. Then the distribution of $\sum_{i=1}^n x_i x_i'$ is called a (K -dimensional) Wishart distribution with parameters Σ_x and n . We write

$$\sum_{i=1}^n x_i x_i' \sim \mathcal{W}_K(\Sigma_x, n). \quad (5.2.12)$$

For univariate standard normal random variables x_i , $\sum_{i=1}^n x_i^2$ has a $\chi^2(n)$ distribution, which illustrates that the Wishart distribution can be viewed as a multivariate generalization of a χ^2 distribution with n degrees of freedom. If $\Omega \sim \mathcal{W}_K(\Sigma, n)$, then the distribution of Ω^{-1} depends on Σ^{-1} and n only. The latter distribution is an inverted Wishart or inverse Wishart distribution with parameters Σ^{-1} and n , and is abbreviated as

$$\Omega^{-1} \sim \mathcal{IW}_K(\Sigma^{-1}, n).$$

Suppose that for the VAR(p) model (5.2.1) with Gaussian innovations, $u_t \sim \mathcal{N}(0, \Sigma_u)$, we specify the priors

$$\alpha | \Sigma_u \sim \mathcal{N}(\alpha^*, V_\alpha = V \otimes \Sigma_u) \quad (5.2.13)$$

and

$$\Sigma_u \sim \mathcal{IW}_K(S_*, n), \quad (5.2.14)$$

such that $\Sigma_u^{-1} \sim \mathcal{W}_K(S_*^{-1}, n)$.⁴ Expressing the prior covariance matrix of α as a Kronecker product $V \otimes \Sigma_u$ simplifies the posterior distribution, and we obtain a Gaussian-inverse Wishart distribution

$$\alpha | \Sigma_u, \mathbf{y} \sim \mathcal{N}(\bar{\alpha}, \bar{\Sigma}_\alpha), \quad \Sigma_u | \mathbf{y} \sim \mathcal{IW}_K(S, \tau), \quad (5.2.15)$$

where (5.2.7) implies that

$$\bar{\Sigma}_\alpha = [(V^{-1} \otimes \Sigma_u^{-1}) + (ZZ' \otimes \Sigma_u^{-1})]^{-1} = (V^{-1} + ZZ')^{-1} \otimes \Sigma_u.$$

⁴ Sometimes in the literature the inverse Wishart distribution with parameters S_*^{-1} and n is denoted as $\Sigma_u \sim \mathcal{IW}_K(nS_*, n)$ such that $\Sigma_u^{-1} \sim \mathcal{W}_K(S_*^{-1}/n, n)$ (see, e.g., Uhlig 2005). This difference in notation leaves the definition of the inverse Wishart distribution unaffected.

Substituting $V \otimes \Sigma_u$ for V_α into expression (5.2.6) yields

$$\begin{aligned}\bar{\alpha} &= [(V^{-1} \otimes \Sigma_u^{-1}) + (ZZ' \otimes \Sigma_u^{-1})]^{-1} \\ &\quad \times [(V^{-1} \otimes \Sigma_u^{-1}) \alpha^* + (Z \otimes \Sigma_u^{-1}) y] \\ &= [(V^{-1} + ZZ')^{-1} \otimes \Sigma_u] [V^{-1} \otimes \Sigma_u^{-1}, Z \otimes \Sigma_u^{-1}] \begin{bmatrix} \alpha^* \\ y \end{bmatrix} \\ &= ((V^{-1} + ZZ')^{-1} [V^{-1}, Z] \otimes I_K) \text{vec}[A^*, Y].\end{aligned}$$

Hence, the posterior mean can be written in matrix notation as

$$\bar{A} = (A^* V^{-1} + Y Z')(V^{-1} + ZZ')^{-1}. \quad (5.2.16)$$

Equation (5.2.16) illustrates why Bayesian estimation methods may be used even when the number of regressors exceeds the sample size. In this case, ZZ' is not invertible and, hence, LS estimation is infeasible. In contrast, Bayesian estimation remains feasible. Adding the invertible precision matrix V^{-1} to ZZ' allows us to invert the sum $V^{-1} + ZZ'$, as required for the construction of the posterior mean \bar{A} . Of course, the solution \bar{A} in this case heavily depends on the choice of V^{-1} .

The parameters of the inverse Wishart distribution in (5.2.15) are

$$S = T \tilde{\Sigma}_u + S_* + \hat{A} Z Z' \hat{A}' + A^* V^{-1} A^{*'} - \bar{A} (V^{-1} + ZZ') \bar{A}' \quad (5.2.17)$$

and $\tau = T + n$ (see Koop and Korobilis 2009 or Uhlig 1994, 2005). Here A^* and \bar{A} are $K \times (Kp + 1)$ matrices such that $\alpha^* = \text{vec}(A^*)$ and $\bar{\alpha} = \text{vec}(\bar{A})$, $\hat{A} = Y Z'(ZZ')^{-1}$, and $\tilde{\Sigma}_u = (Y - \hat{A} Z)(Y - \hat{A} Z)' / T$. Since the posterior is from the same distributional family as the likelihood function, the prior (5.2.13)–(5.2.14) is a conjugate prior. Given that the prior is also from the same distributional family as the likelihood, it is more specifically a natural conjugate prior.

The advantage of using a natural conjugate prior is that a known posterior distribution is obtained that can be used for inference on α without additional simulations. In fact, the marginal posterior distribution of α is a multivariate t -distribution with $\tau = T + n$ degrees of freedom, mean $\bar{\alpha}$, and covariance matrix

$$\Sigma_{\alpha|y} = \frac{1}{\tau - K - 1} ((V^{-1} + ZZ')^{-1} \otimes S) \quad (5.2.18)$$

(e.g., Koop and Korobilis 2009).

The Gaussian-inverse Wishart posterior distribution can be used as a basis for inference on functions of α and Σ_u such as structural impulse responses. If the structural VAR model is just-identified, the structural parameters will be nonlinear functions of the reduced-form parameters considered thus far. In that case, the posterior distribution of the structural impulse responses may be simulated by drawing from the joint posterior distribution of the reduced-form parameters and substituting these draws into the formula of the structural

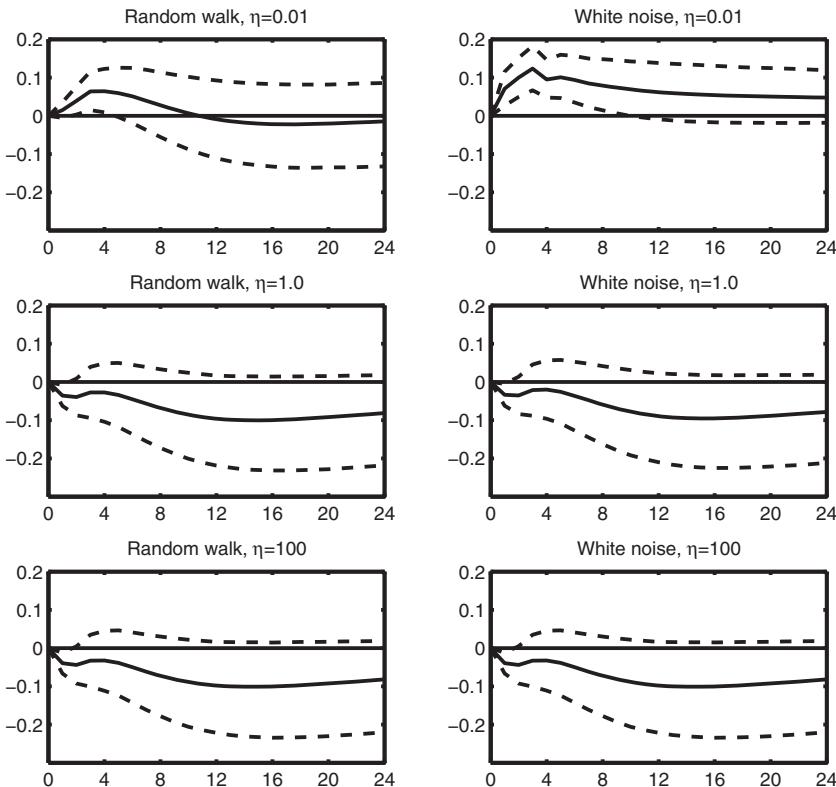


Figure 5.3. Simulated quantiles of inflation responses to monetary policy shocks for different Gaussian-inverse Wishart priors (pointwise median and 10% and 90% quantiles of the posterior distribution computed from 10,000 draws, given the prior parameters $V = \eta I$, $S_* = I_K$, $n = K + 1 = 4$).

impulse responses. In practice, such draws may be generated by first drawing τ independent vectors x_i , $i = 1, \dots, \tau$, from a K -dimensional normal distribution, $\mathcal{N}(0, S^{-1})$, and then conditioning on $\sum_{i=1}^{\tau} x_i x_i'$ for Σ_u^{-1} in simulating a draw from the posterior of α in (5.2.15). Note that S in (5.2.17) does not involve any unknown quantities, and hence can be computed when the prior is specified and the data are available. This means that when draws from the posterior are required, they can be obtained quickly and easily. Of course, the choice of the prior determines to some extent the posterior. If we choose $V^{-1} = 0$, for example, (5.2.16) implies that the posterior mean reduces to the LS estimator. More detailed discussion of this case can be found in Chapter 13.

An Empirical Illustration. We illustrate the use of the Gaussian-inverse Wishart prior based on the same empirical example already employed for the Minnesota prior. Figure 5.3 plots the inflation responses to a contractionary

monetary policy shock for different specifications of the prior. Our analysis is based on one of many possible configurations of this prior. The prior mean of the VAR parameters is either a random walk or white noise. The covariance matrix V is specified as ηI , where η is a prespecified constant. By varying η we can examine the effect of changing the prior variances. The hyperparameter η takes the place of λ in the Minnesota prior and determines the amount of shrinkage. A smaller value of η implies a smaller prior variance and, hence, more shrinkage whereas a larger η implies less shrinkage of the parameter estimates. The prior for Σ_u is chosen arbitrarily to be $S_* = I_K$ and $n = K + 1$.

Figure 5.3 illustrates that a small value of η (implying a tight prior variance) has a substantial effect on the estimated impulse responses. In fact, for a very small $\eta = 0.01$, inflation is estimated to respond positively to an interest rate shock. Thus, there is again a price puzzle. If η is increased, the posterior mean approaches the LS estimator as expected because all terms involving V^{-1} in the formulas for the posterior moments disappear if $\eta \rightarrow \infty$ and $V^{-1} \rightarrow 0$.

Extensions of a Gaussian-Inverse Wishart Prior. Giannone, Lenza, and Primiceri (2015) show that the Gaussian-inverse Wishart prior (5.2.13)–(5.2.14) implies a closed-form expression for the marginal likelihood that is easy to maximize with respect to the hyperparameters. They provide evidence that choosing the hyperparameters in this way results in models that tend to forecast more accurately and imply economically plausible impulse responses. This finding is based on quarterly VAR models. For monthly VAR models, there is evidence that Bayesian shrinkage estimation along these lines may actually worsen the accuracy of the forecasts compared with unrestricted LS estimation, as illustrated in Baumeister and Kilian (2012). Intuitively, this difference arises because forecasts from quarterly models tend to be smoother than forecasts from monthly models. Thus priors that smooth the dynamics of the VAR model are less likely to oversmooth in quarterly models.

Alternatively, one could also specify a proper prior for the hyperparameters and determine the posterior. Then a Gibbs sampler could be used for simulating draws from the joint posterior of α , Σ_u , and the hyperparameters. Because, for a given set of hyperparameters, the posterior of α and Σ_u is from a known distribution, it is easy to draw from the conditional posterior of the latter parameters (see Giannone, Lenza, and Primiceri 2015 for details).

Drawbacks of the Gaussian-Inverse Wishart Prior. A drawback of the natural conjugate prior is that it hinges on the regression matrix being $Z' \otimes I_K$. In other words, it requires that the regressors in each of the K equations of a K -dimensional VAR model be the same. In some applications, this condition is problematic because one may want to drop lags of some variable in one equation but not in others (see Chapter 2). Even though the Bayesian approach

can be viewed as an alternative to subset VAR models because it reduces the parameter variability by smoothing, there can be arguments for eliminating lags of one variable from some equation even in Bayesian analysis. This situation arises, for example, when one variable is specified to be Granger non-causal for some other variable. In such a case the posterior will no longer be Gaussian-inverse Wishart and we have to revert to simulation methods for generating draws from the posterior.

Another undesirable feature of the natural conjugate prior is the multiplicative covariance structure $V_\alpha = V \otimes \Sigma_u$. Notice that the Minnesota prior has a more general covariance that is not encompassed by this expression unless $\theta = 1$. Thus, the Kronecker product form in (5.2.13) is clearly restrictive. It implies that the prior covariance matrices for the lags of the k^{th} variable in different equations are proportional. More precisely, denoting the ij^{th} element of Σ_u by σ_{ij} , the lags of variable y_{kt} in equations i and j have prior covariances $\sigma_{ik}V$ and $\sigma_{jk}V$, respectively. Put differently, they differ by a multiplicative factor.

If these features are deemed too restrictive, one may, of course, specify a Gaussian-inverse Wishart prior of a more general type. This approach entails the loss of the known closed-form distribution of the posterior, however, and, hence, makes it necessary to use computationally more costly simulation techniques for inference.

In the next subsection a prior is discussed that is less restrictive than the natural conjugate prior and still makes it easy to sample from the posterior because it provides a natural basis for employing a Gibbs sampler.

5.2.5 The Independent Gaussian-Inverse Wishart Prior

In the natural conjugate Gaussian-inverse Wishart prior the distributions of the parameters α and Σ_u are not independent. The prior for α in (5.2.13) obviously depends on Σ_u . Alternatively, one may explicitly impose independence of the priors of α and Σ_u by specifying the joint prior pdf to be of the form

$$g(\alpha, \Sigma_u) = g_\alpha(\alpha)g_{\Sigma_u}(\Sigma_u). \quad (5.2.19)$$

This approach facilitates the use of a Gibbs sampler. The prior resulting from the marginal priors

$$\alpha \sim \mathcal{N}(\alpha^*, V_\alpha) \quad (5.2.20)$$

and

$$\Sigma_u \sim \mathcal{IW}_K(S_*, n), \quad (5.2.21)$$

is called independent Gaussian-inverse Wishart because of the independence assumption for the marginal prior distributions of α and Σ_u .

Assuming a Gaussian VAR process to start with, we know from Section 5.2.1 that the posterior of α given Σ_u is normal,

$$\bar{\alpha} | \Sigma_u, \mathbf{y} \sim \mathcal{N}(\bar{\alpha}, \bar{\Sigma}_\alpha), \quad (5.2.22)$$

where

$$\begin{aligned} \bar{\alpha} &= [V_\alpha^{-1} + (ZZ' \otimes \Sigma_u^{-1})]^{-1} [V_\alpha^{-1}\alpha^* + (Z \otimes \Sigma_u^{-1})\mathbf{y}] \\ &= \left[V_\alpha^{-1} + \sum_{t=1}^T \mathbf{Z}'_t \Sigma_u^{-1} \mathbf{Z}_t \right]^{-1} \left[V_\alpha^{-1}\alpha^* + \sum_{t=1}^T \mathbf{Z}'_t \Sigma_u^{-1} y_t \right] \end{aligned} \quad (5.2.23)$$

and

$$\bar{\Sigma}_\alpha = [V_\alpha^{-1} + (ZZ' \otimes \Sigma_u^{-1})]^{-1} = \left[V_\alpha^{-1} + \sum_{t=1}^T \mathbf{Z}'_t \Sigma_u^{-1} \mathbf{Z}_t \right]^{-1}. \quad (5.2.24)$$

Here \mathbf{Z}_t is $Z'_t \otimes I_K$ if the same lagged variables appear in all equations. If some lags are removed from some of the equations, these expressions may still be used after removing the corresponding element from α and redefining the rows of \mathbf{Z}_t accordingly. Thus, the expressions in terms of \mathbf{Z}_t are, in fact, more general than the expressions involving Z .

The conditional posterior of Σ_u , given α , is an inverse Wishart distribution,

$$\Sigma_u | \alpha, \mathbf{y} \sim \mathcal{IW}_K(S, \tau) \quad (5.2.25)$$

with

$$S = S_* + \sum_{t=1}^T (y_t - \mathbf{Z}_t \alpha)(y_t - \mathbf{Z}_t \alpha)'$$

and

$$\tau = T + n.$$

Both conditional posteriors are from known distribution families and therefore easy to sample from, facilitating the use of the Gibbs sampler for drawing samples from the joint posterior distribution.

Empirical Illustration. We now reexamine our empirical example using the independent Gaussian-inverse Wishart prior. The responses of inflation to a contractionary interest rate shock are depicted in Figure 5.4. They are computed by using a Gibbs sampler to draw from the posterior. The i^{th} iteration is based on the conditional distributions

$$\alpha | \Sigma_u^{(i-1)}, \mathbf{y} \sim \mathcal{N}(\bar{\alpha}^{(i-1)}, \bar{\Sigma}_\alpha^{(i-1)}) \quad \text{and} \quad \Sigma_u | \alpha^{(i)}, \mathbf{y} \sim \mathcal{IW}_K(S^{(i)}, \tau),$$

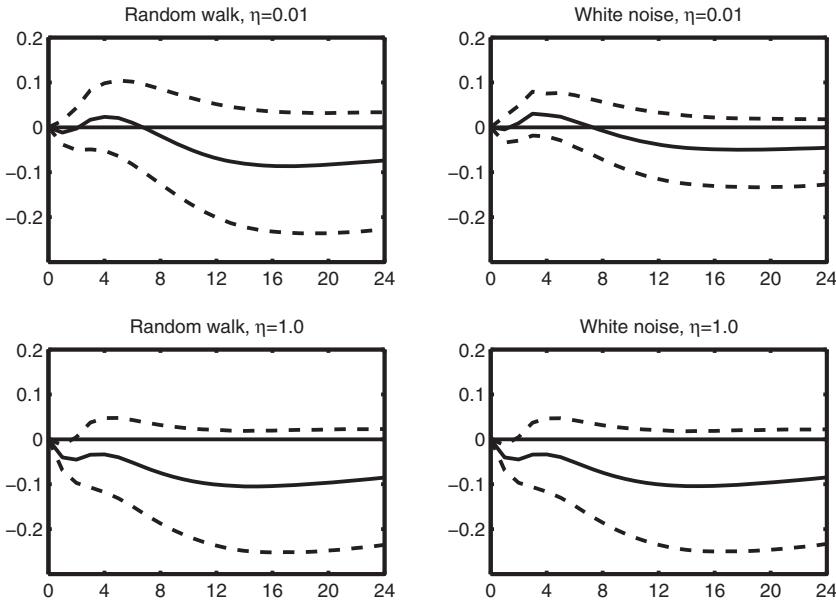


Figure 5.4. Simulated quantiles of inflation responses to monetary policy shocks for different independent Gaussian-inverse Wishart priors (pointwise median and 10% and 90% quantiles of the posterior distribution computed from 10,000 draws, given the prior parameters $V_\alpha = \eta I$, $S_* = I_K$, $n = K + 1 = 4$).

where

$$\bar{\alpha}^{(i-1)} = \left[V_\alpha^{-1} + ZZ' \otimes (\Sigma_u^{(i-1)})^{-1} \right]^{-1} \left[V_\alpha^{-1} \alpha^* + \left(Z \otimes (\Sigma_u^{(i-1)})^{-1} \right) y \right],$$

$$\bar{\Sigma}_\alpha^{(i-1)} = \left[V_\alpha^{-1} + ZZ' \otimes (\Sigma_u^{(i-1)})^{-1} \right]^{-1},$$

and

$$S^{(i)} = S_* + \sum_{t=1}^T (y_t - \mathbf{Z}_t \alpha^{(i)}) (y_t - \mathbf{Z}_t \alpha^{(i)})'.$$

A burn-in sample of 20,000 draws is discarded, and then 10,000 draws are computed to determine the quantiles of the pointwise distributions of the structural impulse responses. We use parameter settings for α^* , V_α , S_* , and n in the prior distributions similar to those for the natural conjugate Gaussian-inverse Wishart prior. Figure 5.4 illustrates once again that the posterior and, hence, the estimated impulse responses depend on the prior. The posterior quantiles look a little different from those for the other priors if the hyperparameter η is

small and, hence, the prior is tight, whereas they are similar to those obtained with other priors when η is larger ($\eta = 1.0$).

Although this example only serves as an illustration, it should alert the reader to the fact that the choice of priors in Bayesian estimation is not innocuous. It may substantially affect the estimates. Just how large this impact is may be difficult to determine in practice, especially when dealing with nonlinear functions of model parameters.

5.3 Extensions and Related Issues

So far we have primarily considered Gaussian likelihood functions in estimating the VAR model. Although this specification is commonly used in applied work, the assumption of unconditional normality is problematic in many macroeconomic applications (see, e.g., Kilian 1998b). For example, models with volatility clustering necessarily give rise to non-Gaussian unconditional distributions. Examples of such models are discussed in Chapters 14 and 18. While alternative distributions can be accommodated by the Bayesian framework, this generality usually increases the computational cost of Bayesian methods. Additional discussion of how to use Bayesian analysis in specific settings can, for example, be found in Chapters 12, 13, 16, and 18.

In this chapter, we have discussed priors for the parameters of the reduced-form VAR model. This approach continues to be a widely used approach in applied work. An alternative approach in structural VAR analysis is to impose priors on the parameters of the structural VAR representation. The latter approach is briefly discussed in Chapters 12 and 13 in the context of the question of how to conduct inference about structural impulse responses and related statistics.

Our analysis in this chapter has taken the lag order of the VAR model as given. Bayesians typically avoid the question of lag-order selection by choosing a conservative large order p , but they incorporate the prior belief that we are increasingly confident that the lagged coefficients are zero, the longer the lag length is. For example, the Minnesota prior postulates that the prior standard deviation shrinks by a factor of $1/l$ for $l = 1, 2, \dots, p$, as we have seen in this chapter. This device avoids having to truncate the lag structure at some order lower than p at the cost of imposing additional structure on the prior variances of all lagged coefficients. Although this approach is intuitively appealing, it is ad hoc. There is no guarantee that this prior will result in more accurate forecasts or impulse response estimates than estimating an unrestricted $\text{VAR}(p)$ model or for that matter estimating a $\text{VAR}(\hat{p})$ model obtained by conventional lag-order selection methods.

Although this approach is not common in applied work, it is also possible to consider restricted VAR models within the Bayesian framework. For example, Koop and Korobilis (2009) describe a so-called stochastic search variable

selection (SSVS) prior that may be useful in reducing the curse of dimensionality by eliminating some lags from some equations of a VAR model based on Bayesian procedures. Details on such priors can be found in the related Bayesian literature.

Finally, whereas in this chapter we have focused on priors motivated by the improved forecast accuracy of the estimated VAR model, there also have been efforts to construct priors for VAR model parameters that incorporate restrictions implied by dynamic macroeconomic models. For example, Ingram and Whiteman (1994), Del Negro and Schorfheide (2004, 2011), and Del Negro, Schorfheide, Smets, and Wouters (2007) discuss priors for VAR models derived from specific DSGE models.

6 The Relationship between VAR Models and Other Macroeconometric Models

This chapter puts the development of the VAR model in historical context and clarifies its relationship with other modeling frameworks used in empirical macroeconomics including dynamic simultaneous equations models (DSEMs) and dynamic stochastic general equilibrium (DSGE) models.

6.1 The Relationship between VAR Models and Traditional Dynamic Simultaneous Equations Models

Vector autoregressions as a tool for macroeconomic analysis were first introduced by Sims (1980a; 1980b). They quickly became the workhorse model in empirical macroeconomics as well as in empirical finance. Prior to 1980, the dominant model used in empirical macroeconomics was the DSEM. This class of models was developed by the Cowles Commission for Research in Economics (later renamed the Cowles Foundation), which concerned itself with linking economic theory to mathematics and statistics. Starting in the 1950s, macroeconomists at the Cowles Commission, including the subsequent Nobel Prize laureate Lawrence Klein, began constructing the first DSEMs of the U.S. macro economy (see Klein 1950; Klein and Goldberger 1955). Their work built on the pioneering work of Tinbergen (1939) and was inspired by the theoretical insights of Keynes' work. The objective of these researchers was to give quantitative content to Keynes' macroeconomic theory and to turn the DSEM into a tool for policymakers. Over time, DSEMs grew larger as economists strived for greater realism and sought to make the model more relevant for policy questions. These DSEMs often contained dozens or even hundreds of equations. They were driven by a large number of exogenous model variables, while including only a small number of endogenous model variables.

Exogeneity here is defined as the absence of contemporaneous or lagged feedback from the endogenous model variables to the exogenous model

variables. Suppose that a vector of observed variables y_t may be partitioned into two classes of variables, ‘exogenous’, x_t , and ‘endogenous’, z_t . Then, following Zellner and Palm (1974), a structural dynamic, multivariate simultaneous linear system may be written as

$$\begin{bmatrix} H_{11}(L) & H_{12}(L) \\ 0 & H_{22}(L) \end{bmatrix} \begin{pmatrix} z_t \\ x_t \end{pmatrix} = \begin{bmatrix} F_{11}(L) & 0 \\ 0 & F_{22}(L) \end{bmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix},$$

where $H_{ij}(L)$ and $F_{ij}(L)$, $i, j \in \{1, 2\}$, are matrix polynomials in the lag operator L and the elements of $w_t = (w'_{1t}, w'_{2t})'$ are mutually uncorrelated white noise. Such systems are known as structural vector ARMAX (or VARMAX) models, where the X refers to the set of strictly exogenous variables x_t . For further discussion of the concept of strict exogeneity, see Chapter 7. Conditions for the identification of the model parameters are discussed in Zellner and Palm (1974) and Hannan and Deistler (1988), for example. In practice, the structural equations are typically estimated by single-equation methods.

The implied reduced form of this system is obtained by expressing z_t as a function of lagged endogenous and current and lagged exogenous variables. Zellner and Palm (1974) show that this reduced-form representation takes the form of a VARMA (or, alternatively, VARMAX) model. The final form of this model is obtained by inverting the autoregressive lags for z_t and expressing z_t as a function of current and lagged values of the shocks and the exogenous variables only (see Zellner and Palm 1974; Wallis 1977; Lütkepohl 2005).

One advantage of large-scale DSEM^s is that they allow detailed analysis of alternative policies and scenarios involving changes in the path of the exogenous model variables. Another advantage is that their economic structure facilitates the communication of the results to policymakers. For that reason, descendants of these models survive to this day at some private forecasting firms. Their demise, at least in academic research, came when they failed to explain the poor macroeconomic performance of the 1970s. This failure seemed to support the Lucas critique of structural DSEM^s in the 1970s, which questioned the invariance of the estimated DSEM^s to policy interventions (see Lucas 1976). It was followed by the rise of rational expectations models in macroeconomics in the 1970s and the development of VAR models as well as DSGE models in the 1980s.

6.1.1 The VAR Representation of Traditional DSEM^s

There is a close connection between VAR models and the structural, reduced-form and final-form representations of DSEM^s. Zellner and Palm (1974) show that the reduced form of structural DSEM^s in general has a VARMA (or VARMAX) representation that can typically be approximated by a finite-order VAR (or VARX) model.

We can think of the structural representation of an econometric model as the data generating process underlying the observed economic data. For an econometric model to be structural it is necessary for the stochastic shocks (or errors) in each equation to be mutually uncorrelated. This feature allows us to consider thought experiments in which one structural shock moves while leaving all other shocks unchanged. The resulting responses of the observables then represent the causal effects of this structural shock. This interpretation would break down if the shock in question were correlated with other shocks.

To illustrate the relationship between structural DSEMs and VAR models, consider the very stylized example of a structural model involving only two equations: (1) a function for aggregate consumption (c_t), built on the assumption of adaptive expectations, and (2) an equation for national income (n_t):

$$\begin{aligned} c_t &= \eta_1 + \alpha n_t + \beta c_{t-1} + w_{1t}, \\ n_t &= \eta_2 + \gamma c_{t-1} + \delta n_{t-1} + w_{2t}, \end{aligned}$$

where w_{1t} and w_{2t} are mutually uncorrelated white noise structural shocks. Any such structural model has a reduced-form representation, which may be constructed by rewriting the system until only lagged dependent variables are left on the right-hand side (RHS). Here it suffices to substitute the complete RHS of the second equation for n_t in the first equation. We obtain

$$\begin{aligned} c_t &= (\eta_1 + \alpha\eta_2) + (\beta + \alpha\gamma)c_{t-1} + \alpha\delta n_{t-1} + (w_{1t} + \alpha w_{2t}), \\ n_t &= \eta_2 + \gamma c_{t-1} + \delta n_{t-1} + w_{2t}. \end{aligned}$$

This reduced form can be expressed as

$$\begin{aligned} c_t &= v_1 + a_{11,1}c_{t-1} + a_{12,1}n_{t-1} + u_{1t}, \\ n_t &= v_2 + a_{21,1}c_{t-1} + a_{22,1}n_{t-1} + u_{2t}, \end{aligned}$$

where, for example, in the first equation $v_1 = (\eta_1 + \alpha\eta_2)$, $a_{11,1} = (\beta + \alpha\gamma)$, $a_{12,1} = \alpha\delta$, $u_{1t} = (w_{1t} + \alpha w_{2t})$. Equivalently, this model may be written in matrix notation as a first-order VAR(1) model:

$$y_t = v + A_1 y_{t-1} + u_t,$$

where $y_t \equiv (c_t, n_t)'$, $v \equiv (v_1, v_2)'$, $u_t \equiv (u_{1t}, u_{2t})'$ is white noise with mean vector 0 and covariance matrix Σ_u , and

$$A_1 \equiv \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix}.$$

This example illustrates that the parameters of a reduced-form model in general are combinations of the parameters of the underlying structural model. Econometricians seek to learn about the parameters of the structural model from the parameters of the reduced-form model. If the structural parameters can be recovered uniquely from the reduced-form parameters

in population, we say that the structural parameters are exactly identified, point identified, or just identified. Determining the values of the structural parameters from the reduced form involves solving a system of equations. In the example above, this derivation is straightforward because the structure of the model is recursive: $\delta = a_{22,1}$, $\gamma = a_{21,1}$, and $\eta_2 = v_2$. Hence, $\alpha = a_{12,1}/a_{22,1}$, $\eta_1 = v_1 - v_2 a_{12,1}/a_{22,1}$, $\beta = a_{11,1} - a_{21,1} a_{12,1}/a_{22,1}$, $w_{2t} = u_{2t}$, and $w_{1t} = u_{1t} - u_{2t} a_{12,1}/a_{22,1}$.

In general, it is possible that some structural model parameters remain unidentified (which means that there are infinitely many possible solutions for these structural parameters even in a sample of infinite length) or over-identified (which means that there are more restrictions than needed to recover the structural parameters). Models in which not all structural parameters are exactly identified are called underidentified or partially identified. Finally, in the special case of models identified based on inequality restrictions, the structural parameters will be set-identified (which means that we can only bound the set of structural parameter values). We defer discussion of such models to Chapter 13.

With these definitions in mind, let us return to the previous example. Clearly, if we had written down a different structural model for c_t and n_t , we might have obtained a different reduced-form representation. Regardless of which structural model is the data generating process for c_t and n_t , however, if we had specified a reduced-form VAR(p) model for $y_t \equiv (c_t, n_t)'$ including a sufficient number of autoregressive lags such that

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

we would expect that reduced-form model to be logically consistent with a large set of different structural models. This intuition also applies to higher-dimensional models. In short, reduced-form VAR models encompass a range of structural DSEMs. This stylized example could be generalized to allow for the inclusion of exogenous variables. We do not consider this extension because exogenous variables are rare in the VAR literature for reasons discussed in the next subsection.

6.1.2 Incredible Restrictions in Traditional DSEMs

As we have seen, the vector autoregressive model is closely related to the reduced-form representation of traditional structural DSEMs. The key difference is that the interpretation of the vector autoregression as a structural DSEM requires additional cross-equation restrictions and possibly exclusion restrictions on the reduced-form VAR parameters. These restrictions may also affect the dynamic structure of the model. There are two types of restrictions on the

dynamic specification of the model. One involves treating some model variables as exogenous, consistent with the partial equilibrium nature of the IS-LM models of the 1950s and 1960s. This allows one to greatly reduce the number of parameters to be estimated.

Loosely speaking, a variable is exogenous if it is determined outside the system of equations under consideration. Such a variable is not subject to current or lagged feedback from other model variables, but only depends on its own lags (or on other exogenous variables). In practice, DSEMs treat a wide range of variables as exogenous with respect to the model variables of interest. Common examples include population growth, the money stock, and global commodity prices. More often than not, regarding a variable as exogenous is at best an approximation and at worst a heroic assumption. For example, the money stock is affected by the endogenous money creation of the banking system and hence not a plausibly exogenous model variable (see Chapter 7).

The other type of restriction involves constraining the dynamics of the endogenous variables and of the error term. Ideally, these dynamic restrictions should reflect economic theory, but traditional macroeconomic models such as the IS-LM model are static models. They offer no guidance for the specification of the dynamics of the econometric model. Such dynamics are important for two reasons. One reason is that adjustments to shocks are inherently sluggish, reflecting frictions in the economy. The other reason is that economic decisions depend on expectations which are driven at least in part by lagged model variables. In practice, therefore, econometricians in the 1950s and 1960s augmented the original static IS-LM model by adding dynamics, for example, in the form of partial adjustment models of investment, adaptive expectations models of consumption, and distributed-lag models for the regression errors or for exogenous model variables. The distributed-lag models often were restricted to enforce a smooth decay of the model coefficients at higher lags and to reduce the number of model parameters to be estimated freely. The purpose of these restrictions was (a) to enable researchers to fit large-dimensional models to the data and (b) to allow them to recover the structural parameters of the model. For further details see, e.g., Judge, Griffiths, Hill, Lütkepohl, and Lee (1985).

There are three problems with this traditional approach. First, commonly used dynamic specifications tend to violate the notion of rational expectations (which implies that economic agents do not make mistakes in expectation). This is a particular concern for partial adjustment and adaptive expectations specifications that prevent agents from adjusting fully on impact to policy changes. Second, the choice between different dynamic specifications often is arbitrary. Third, the building blocks of large-scale DSEMs are developed in isolation, with researchers focusing on one block of the model at a time (e.g., the consumption block, investment block, trade block, fiscal block, or

monetary block), taking as given all model variables not determined within their block. This partial-equilibrium approach ignores the feedback from one model block to another in general equilibrium. Dynamic general equilibrium macroeconomic models imply that every variable depends on every other variable in the economy, which contradicts the traditional notion that some variables may be treated as exogenous with respect to others.

6.1.3 Structural VAR Models as an Alternative to Traditional DSEMs

Since the restrictions on lagged model variables in the DSEM are rarely credible and since the precise form of the structural data generating process is not known in practice, Sims (1980a) made the case that we might as well replace the restricted reduced-form model by an unrestricted VAR model with a suitably chosen maximum lag order. This approach, of course, forces us to reduce the dimensionality of the model greatly because a VAR model involves many more parameters to be estimated. Instead of considering hundreds of equations, we will be able to include only a handful of equations in the model. Moreover, without an alternative set of identifying assumptions, VAR models merely provide reduced-form summaries of the data that are not particularly interesting from an economic point of view. Sims' key insight was that, starting with a general structural model of the form

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t,$$

with reduced-form representation

$$y_t = \underbrace{B_0^{-1} B_1 y_{t-1}}_{A_1} + \cdots + \underbrace{B_0^{-1} B_p y_{t-p}}_{A_p} + \underbrace{B_0^{-1} w_t}_{u_t},$$

after setting $\Sigma_w = I_K$ without loss of generality such that $B_0^{-1} B_0^{-1\prime} = \Sigma_u$, we can dispense with ad hoc restrictions on the slope parameters altogether, provided we are able to restrict enough elements of the structural impact multiplier matrix B_0^{-1} , so the remaining elements can be estimated from the data. Put differently, knowledge of B_0^{-1} suffices to trace out the dynamic response of the model variables to structural shocks even in the absence of restrictions on the slope parameters. Econometricians soon developed a range of methods of imposing such identifying restrictions on B_0^{-1} (or, equivalently, on B_0). We review this structural VAR literature in Chapters 8, 10, 13, 14, and 15. To the extent that these restrictions are more credible economically than the restrictions in traditional DSEMs, the structural VAR approach provides a superior alternative to traditional DSEMs.

It is important to reiterate that Sims' critique was not directed at the idea of DSEMs itself, but rather at how these models were specified in practice in the 1970s. In fact, structural VAR models of the type studied in this

book may be interpreted as small-scale DSEM_s in which restrictions have been imposed on the impact multiplier matrix and the lagged values are left unrestricted.

6.2 The Relationship between VAR Models and DSGE Models

VAR models are also closely related to DSGE models. DSGE models require the user to explicitly specify the microstructure of the economy. The premise is that this model structure is invariant to policy interventions. Agents maximize their objective functions in expectation subject to individual and aggregate resource constraints, while taking account of the behavior of other agents in a rational manner. They do so in a world subject to random variation in variables such as technology that are considered exogenous with respect to agents' decisions in that there is no contemporaneous or lagged feedback from the agents' actions to these variables. Changes in "technology" here refer broadly to changes in the efficiency with which firms combine the factors of production into final goods. After log-linearizing the DSGE model about the nonstochastic steady-state path of the model variables, agents' behavior and choices can be characterized in the form of log-linear decision rules that depend only on state variables known to the agents at each point in time. This log-linear structure allows us to simulate the dynamic behavior of the endogenous model variables subject to the arrival of random shocks.

The state-space representation of DSGE models, under suitable conditions, can be written as a reduced-form VARMA model for the endogenous model variables, which in turn, under suitable conditions, can be approximated by a finite-order reduced-form VAR model. The link between structural VAR models and DSGE models is less straightforward and depends on how the structural VAR shocks are identified. For a more detailed discussion of how to derive the approximate VAR representation of DSGE models the reader is referred to the survey by Giacomini (2013). For further discussion of the link between DSGE and VAR models, see Christiano, Eichenbaum, and Vigfusson (2006b, section 2.2), Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007), Ravenna (2007), Franchi and Vidotto (2013), and Guerron-Quintana, Inoue, and Kilian (2017). Next, we sketch the central issues in approximating DSGE models by finite-order VAR models and provide some intuition. Our discussion builds on Christiano (2002).

6.2.1 Basics

Let z_t be an $n \times 1$ vector of endogenous DSGE model variables (e.g., capital stock, output, consumption, investment). Let s_t be an $m \times 1$ vector of exogenous state variables (e.g., technology shocks, government spending shocks, money supply shocks, preference shocks). Without loss of generality, the law

of motion of s_t can be expressed as a VAR(1) process

$$s_t = Ps_{t-1} + \varepsilon_t,$$

where the $m \times 1$ vector ε_t is zero-mean white noise such that ε_t is uncorrelated with ε_{t-l} and s_{t-l} for $l > 0$.

This law of motion is quite general and accommodates an arbitrary ARMA(p, q) representation for the underlying shocks. Note that the common terminology of referring to s_t as “shocks” is misleading from an econometric point of view, because by construction the shock is ε_t rather than s_t . In reality, s_t refers to an exogenous state variable in the model. We nevertheless will follow the DSGE model literature in using this terminology in this section.

The leading example for an exogenous state variable is the technology shock process underlying the neoclassical growth model. We denote this variable by x_t . It is usually assumed that the process for technology (which determines aggregate productivity levels in the model economy) is highly persistent and follows an AR or ARMA process.

Example 1 Suppose a shock x_t follows an AR(1) process

$$x_t = \rho x_{t-1} + e_t,$$

where e_t is mean zero white noise. Then $s_t = Ps_{t-1} + \varepsilon_t$, with

$$s_t = x_t, \quad P = \rho, \quad \varepsilon_t = e_t.$$

Example 2 If x_t follows an ARMA(1,1) process

$$x_t = \rho x_{t-1} + e_t - \gamma e_{t-1},$$

instead, then $s_t = Ps_{t-1} + \varepsilon_t$, with

$$s_t = \begin{pmatrix} x_t \\ e_t \end{pmatrix}, \quad P = \begin{bmatrix} \rho & -\gamma \\ 0 & 0 \end{bmatrix}, \quad \varepsilon_t = \begin{pmatrix} e_t \\ e_t \end{pmatrix}.$$

It can be shown that a broad class of models, including real business cycle models with taxes and other distortions, limited participation models, models of labor hoarding, and models with sticky prices, has solutions for the endogenous model variables of the form:

$$z_t = Az_{t-1} + Ms_t,$$

where A is $n \times n$, M is $n \times m$, and $s_t = Ps_{t-1} + \varepsilon_t$. Hence, the reduced form representation of the vector process z_t will not in general have a finite-order VAR representation, unless $P = 0$ and s_t reduces to white noise. In fact, the process z_t need not even have a finite-order VARMA representation.

Economic theory provides no guidance for the specification of the time series process of the exogenous state variable s_t , but it is generally agreed

that s_t must be fairly persistent for models to generate persistence in z_t comparable to the actual data. Hence, s_t will not be white noise. Typical choices include AR(1), AR(2), or MA(2) models for technology, the growth in monetary aggregates and in government spending, and other variables treated as exogenous state variables.

The specification of the latent state variables, s_t , matters for the time series representation of the endogenous model variables, z_t . For example, choosing a process for the technology shock, x_t , has immediate implications for the time series representation of endogenous model variables such as the capital stock. In the neoclassical growth model, for example, the household's capital stock for next period, k'_t , is decided one period in advance and evolves according to the log-linear decision rule:

$$k'_t = \alpha_0 + \alpha_1 k'_{t-1} + \alpha_2 x_t, \quad (6.2.1)$$

where x_t denotes the technology shock. Thus, if x_t follows an MA(1) process, k'_t will follow an ARMA(1,1) process. If x_t follows an AR(1) process of the form $(1 - \rho L)x_t = e_t$, however, as is commonly assumed in the literature, k'_t will follow an AR(2) process. This result can be verified by premultiplying (6.2.1) by $(1 - \rho L)$.

The variable k'_t is, of course, only one of many endogenous variables in the DSGE model. In practice, we are interested in the joint n -dimensional vector time series process z_t , which raises additional complications. It may seem that z_t would follow a VARMA process if one of its elements follows an ARMA process, but this is not necessarily the case. A common problem is that the number of exogenous shocks in the DSGE model is smaller than the dimensionality of z_t . In this case, the time series process for z_t will be of reduced rank. For example, a textbook real business cycle (RBC) model has only one shock, so, when fitting a VAR(MA) model to output, investment, and consumption data from this RBC model, the error covariance matrix Σ_u would be of reduced rank. It follows immediately that z_t cannot be approximated by conventional VAR(MA) processes because the latter model requires the error covariance matrix Σ_u to be of full column rank. This condition means that there must be (at least) as many shocks as variables in the VAR(MA) model.

This rank-deficiency problem can be remedied in four ways: (a) by the use of reduced-rank estimation methods; (b) by adding noise shocks to the DSGE model (often referred to as measurement errors, although model approximation errors would be a more apt name) that are devoid of any economic interpretation, thereby undoing the benefits of relying on explicit microfoundations; (c) by reducing the number of observables in the VAR model; or (d) by augmenting the number of economically interpretable shocks in the model. Of course, this raises the question of why the latter shocks were not included in the model in the first place. One common response is to include preference shocks and efficiency shocks in the DSGE model that allow for departures from first-order

conditions that must hold in equilibrium. The difference between such shocks and noise shocks is largely a matter of degree. Neither type of shock is based on microfoundations. For now, let us suppose that the reduced-form representation of z_t is of full rank.

6.2.2 The Role of Data Transformations

In practice, it is almost never the case that we fit a parametric VAR model to the n -dimensional vector z_t . One reason is that there are often more variables in the DSGE model than could be reasonably included in a small-scale to medium-scale VAR(MA) model. Hence, we inevitably must integrate out one or more of the elements of z_t . Another reason is that not all time series in the model are actually observed by the econometrician. For example, there are no good data on the capital stock that could be included in a VAR model, which again requires the marginalization of z_t .

This marginalization affects the time series process for the remaining endogenous model variables, as discussed in Chapter 2. In particular, even if z_t followed a known parametric VAR(p) process, any K -dimensional subprocess or marginal process y_t of an n -dimensional process z_t , obtained by using transformation matrices of the form $F = [I_K, 0]$ with $K < n$, would not in general be a VAR process of finite order. This result follows from the following proposition in Lütkepohl (2005):

Proposition. Let z_t be an n -dimensional, stable, invertible VAR(p) process and let F be a $K \times n$ matrix of rank K . Then the process $y_t = Fz_t$ has a VARMA(p^* , q^*) representation with $p^* \leq np$ and $q^* \leq (n - 1)p$. \square

Hence, except in rare cases, DSGE models will not have a finite-order VAR representation. Moreover, even if the data from a particular DSGE model were to have a finite-order VAR representation, any aggregation of the observed data not accounted for in the DSGE model would similarly induce MA components of unknown form in the reduced-form representation of the data. One example is statistical agencies aggregating data across households or firms. Another example is the aggregation of data over time (say, from monthly to quarterly frequency). Thus, it would be surprising indeed if the quarterly VAR representation of y_t were truly of finite lag order (see also Section 2.2.3).

6.2.3 Why Not Use VARMA Models?

In light of these arguments against finite-order VAR models, it may seem that we simply should have estimated finite-order VARMA(p, q) models instead. There are two counterarguments, however. First, we have little confidence in the particular VARMA(p, q) specification implied by a given DSGE

model because that specification depends on inherently atheoretical assumptions about the dynamics of the exogenous state variables. Second, large-dimensional VARMA models are difficult to estimate reliably and hence are rarely used.

These considerations suggest the need for an alternative class of reduced-form models when approximating data generated by DSGE models. If the VARMA representation is invertible, z_t can be represented as a VAR process. More often than not, this VAR process is of infinite order. In this case, we can appeal to standard results in the literature about approximating $\text{VAR}(\infty)$ processes.

6.2.4 Autoregressive Sieve Approximations of $\text{VAR}(\infty)$ Processes

Assuming an exponential rate of decay of the coefficients of the autoregressive representation, the $\text{VAR}(\infty)$ process may be approximated by fitting a sequence of finite-order $\text{VAR}(p_T)$ models, as discussed in Chapter 2. The approximation error becomes arbitrarily small asymptotically, provided that p_T increases with the sample size T at a suitable rate. This semiparametric approach is also known as an autoregressive sieve approximation. The sieve idea refers to the fact that the choice of p_T determines how much information passes through the sieve. This sieve approach is not a panacea, however. First, it will only be suitable if the DSGE model is invertible. Second, a good linear approximation for given T may require a very large p_T .

One concern in the literature is that in some cases for small T no feasible choice of p_T may suffice for a good approximation of the structural impulse responses implied by a DSGE model. Given the difficulty of matching structural shocks in VAR models with structural shocks in DSGE models, the literature addressing this point has tended to focus on VAR models of the effects of technology shocks identified by long-run identifying restrictions (see Chapter 10). There are examples in which structural impulse response estimates, generated by fitting such VAR models to data generated from DSGE models, do not even come close to recovering the structural impulse responses implied by the underlying DGP. For example, Ravenna (2007) finds that even VAR models with 12 lags estimated on 50 years of quarterly data may provide a poor approximation to the structural responses of interest (see also Mertens 2012; Liu and Konstantinos 2012; Poskitt and Yao 2017).

Kascha and Mertens (2009) conclude that this problem is not so much related to the quality of the VAR approximation as to the use of long-run restrictions for identification (see Chapters 10 and 11). They show that VARMA models are no more accurate than VAR approximations as measured by the bias and MSE of the structural impulse responses. One possible explanation is that the VARMA model is close to being unidentified because the roots of the AR and MA polynomials are similar in magnitude, nearly causing these

roots to cancel (see Andrews and Cheng 2012). However, Kascha and Mertens find that equally inaccurate estimates are also obtained when estimating the reduced-form MA representation directly from the state-space equations of the DSGE model, which eliminates the problem of model misspecification and the problem of near root cancellations. How general this type of result is remains an open question. In related work, Pagan and Robinson (2016) provide evidence that the quality of low-order reduced-form VAR approximations to a DSGE model also depends on whether the stock variables in DSGE models (such as the capital stock), which are typically not observed in the real world, can be expressed as a function of a small number of lags of the observed variables.

A final concern is that none of these studies makes allowances for the small-sample bias in the VAR slope parameter estimator \hat{A} discussed in Chapter 2. This small-sample bias arises even in correctly specified models and may greatly undermine the accuracy of structural impulse response estimators, especially when working with model variables expressed in log-levels (see, e.g., Kilian 1998c).

6.2.5 Summary of Potential Problems in Approximating DSGE Models with VAR Models

To reiterate, there are several conditions required for approximating the reduced-form representation of DSGE models with VAR models. First, the number of shocks in the DSGE model must match the number of shocks in the reduced-form VAR model. We implicitly assumed that this condition is met, but this condition must be checked on a case-by-case basis. Second, the state-space representation of the DSGE model must be invertible for a VAR approximation to work. Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007) study conditions under which the state-space representation of DSGE models and VAR models match up. They show that invertibility cannot be taken for granted. This question is closely related to the observation going back to the early 1980s that models with forward-looking agents may not have a (backward-looking) VAR representation (see also Chapter 17). Despite these caveats, Fernández-Villaverde et al. conclude that many models of interest to macroeconomists are invertible. Moreover, Sims (2012) shows that, even when the conditions for the invertibility of the state-space representation fail, the degree of misspecification of the structural VAR responses may be small.

Even if these first two conditions hold, however, the question remains of how good of an approximation finite-dimensional VARs provide to the $\text{VAR}(\infty)$ data generating process. Simulation evidence shows that in some examples the $\text{VAR}(p_T)$ approximation to the $\text{VAR}(\infty)$ model appears adequate. In other examples, it may be poor for any feasible choice of p_T in realistic sample sizes, which may be an indication of the process being nearly noninvertible,

but also could be caused by other approximation errors. How important such counterexamples are depends on how relevant the latter DSGE models are for policy work. An obvious question for future research is how good the autoregressive approximation is for DSGE models commonly used by policymakers. Another important question is how to select the approximating lag order p_T in practice. The answer to this question also depends on which aspect of the fit of the DSGE model we are interested in. This is an open area of research. For example, Hall, Inoue, Nason, and Rossi (2012) propose a procedure for selecting p_T in conjunction with the horizon of the impulse responses to be matched.

All of the potential problems discussed so far relate to the question of whether reduced-form VAR models can approximate data generated by DSGE models. If we want to compare impulse responses from a structural VAR model to impulse responses in DSGE models, we must make sure in addition that the identifying assumptions in the VAR are fully consistent with the structure imposed in the DSGE model. This requirement is difficult to meet in practice. For example, the timing assumptions imposed in standard semistructural VAR models of monetary policy, as discussed in Chapter 8, are at odds with the structure of many DSGE models of monetary policy. As a result, caution must be exercised in comparing results from DSGE and structural VAR models. This caveat is likely to apply more to some methods of achieving identification in structural VAR models than to others. For example, structural models based on sign restrictions can be expected to be more robust than methods based on short-run exclusion restrictions. We will return to this discussion in Chapter 11.

6.3 DSGE Models as an Alternative to VAR Models?

The force of the Lucas critique of DSEMs led to considerable interest in rational expectations models in the 1970s and early 1980s. This movement was led by Thomas Sargent and Robert Lucas. Whereas there is little controversy among economists that agents are rational in the sense of not making systematic mistakes, that idea is devoid of empirical content if we do not know the objectives and constraints of agents. In practice, this fact gave rise to the Rational Expectations Hypothesis which states that agents' expectations must be consistent with the structure of the economic model, which also requires each agent to know and understand this structure. Thus, the question of testing rationality became intertwined with the question of testing the economist's ideas of what agents' preferences and constraints are. In other words, a rejection of this Rational Expectations Hypothesis may occur because agents are not rational or because the economist's model is wrong.

Early efforts to estimate small-scale rational expectations models were abandoned because the restrictions implied by rational expectations were

routinely rejected, and the resulting models performed worse than unrestricted models (see, e.g., Hansen and Sargent 1980; Wallis 1980). Rational expectations made a comeback in the mid-1980s and 1990s, however, when medium-scale DSGE models became the dominant form of macroeconomic models used by academics.

6.3.1 *Calibrated DSGE Models*

Initially, DSGE models were calibrated rather than estimated. Users simulated the dynamics of the model variables conditional on a set of so-called deep parameters, chosen in a process referred to as calibration. Deep parameters here refer to parameters governing technology and preferences that are presumed invariant to policy changes. The central idea of calibration is to pin down as many deep parameters as possible based on the long-run mean of the data (the first moment), before informally assessing the fit of the model by comparing the cross-autocorrelations (the second moment) in the model and in the data. In practice, this approach is insufficient, however. Some of the remaining model parameters therefore are chosen based on extraneous microeconomic estimates. This approach is controversial because such estimates need not be representative for the behavior of the economy at the aggregate level. Finally, the persistence of the exogenous technology shock is chosen to match the persistence of real GDP in the U.S. data, violating the premise of not relying on second moments in calibrating the parameters (see, e.g., Prescott 1986; Cooley and Prescott 1995). The latter procedure can be viewed as an informal application of the method of moments subject to the remaining parameter values having been restricted. Once the process of calibration is complete, however, all calibrated model parameters are treated as known when confronting the model with the observed data.

Proponents of the calibration methodology dismiss conventional measures of model fit and the idea of statistical testing. Their argument is that all models are inherently misspecified, freeing us from the need to formally test the model. After all, a statistical rejection of the model would only confirm what they had already conceded from the start. Moreover, proponents of calibration prefer not to articulate a formal loss function to be used in evaluating the fit of the DSGE model, allowing them to decide which dimension of model fit is most important on a case-by-case basis.

When calibrators confront their model with the data, they do so informally by inspecting tables of cross-autocorrelations in the data and comparing them with the cross-autocorrelations in the DSGE model. If the model appears at odds with the data in some dimension that is considered important, this fact is interpreted not as a rejection of the model, but as an indication that the model structure requires further refinements. The basic modeling paradigm

is taken as correct by assumption. The approach of trying to match unconditional second moments in the model and in the data may look at first sight like an informal application of the method of moments, but it is not because – rather than adjusting the values of the parameters in the DSGE model in response to indications of poor fit – calibrators tweak the DSGE model structure to achieve improved fit, conditional on the same set of calibrated parameters.

6.3.2 Estimated DSGE Models

The calibration approach to evaluating DSGE models, as envisioned originally, has not stood the test of time. Instead, users of DSGE models at central banks and most academic researchers nowadays rely on direct estimates of the state-space representation of DSGE models. The development of such estimation methods started about 10 years after the publication of Kydland and Prescott (1982). This does not mean that calibration is not a useful tool for macroeconomists. It may be reinterpreted as a method for evaluating the quantitative implications of economic theory, the focus being on understanding the dynamic implications of the DSGE model, not on fitting the actual data. If we are interested in the fit of DSGE models and in their ability to explain observed data, in contrast, there is no substitute for econometric methods. The first study to attempt to estimate the structural representation of a DSGE model directly by unrestricted ML estimation was Leeper and Sims (1994). This research program has been largely abandoned because of convergence problems arising from the large dimension of the parameter space.

An alternative estimation approach in the DSGE model literature is the method of moments. Given that the second moments of the data implied by the DSGE model can be expressed as a function of the deep model parameters, one may infer the values of these deep parameters by matching the moments in the DSGE model with the corresponding moments in the data. Based on a simulation study, Ruge-Murcia (2007) concludes that generalized method of moments (GMM) estimators and simulated method of moments estimators based on an auxiliary time series model compare favorably with the unrestricted ML estimator of the DSGE model. Another popular estimation method is impulse-response matching (see, e.g., Guerron-Quintana, Inoue, and Kilian 2017).

Although there continues to be interest in frequentist estimation methods for DSGE models, the dominant estimation approach since the 2000s has been Bayesian. For this purpose, the structural state-space representation of the DSGE model is typically written in the form of a Gaussian likelihood and combined with prior distributions for the structural DSGE model parameters. One reason for the appeal of the Bayesian framework is that Prescott's

calibration method can be viewed as Bayesian inference with a degenerate prior, conditional on the data, so full-fledged Bayesian methods of inference with proper priors on the structural parameters are a natural generalization of his ideas (see Del Negro and Schorfheide 2008). Another reason is that Bayesian priors can be used to restrict the search path of numerical optimization routines, facilitating the convergence of the estimates to economically plausible values. For a review of Bayesian estimation methods for DSGE models, see An and Schorfheide (2007).

6.3.3 Calibration versus Bayesian Estimation

There are some interesting parallels between calibrators and Bayesian econometricians. The Bayesian approach has a long tradition in time series econometrics. As discussed in Chapter 5, Bayesians think of the data as fixed and the DSGE model parameters as uncertain according to a prior distribution with suitable support. In contrast, calibrators may be viewed as members of a wayward Bayesian tribe whose members think of the data as fixed but of the DSGE model parameters as having a degenerate prior distribution obtained by calibration. In other words, they claim to be absolutely certain of their prior views about the model parameters before looking at the data.

This means that in principle they could simply work out the prior mean of the finite-sample distribution of the statistics of interest in the DSGE model (say, a coefficient of the cross-autocorrelation function of the model data) and compare it to the value of the same statistic in the actual data. Because in practice this finite-sample distribution is not known, calibrators tend to simulate the finite-sample distribution and to compute the average of the statistics of interest numerically. They then compare this prior mean to the corresponding statistic in the actual data.

Given that the prior distribution is degenerate, however, users of this methodology do not update their views about the model parameters based on the data, as one usually would in Bayesian analysis, by forming a posterior distribution. Indeed, calibrators would think of evidence of poor fit between the prior mean of the statistic of interest and the observed value of the same statistic in the data not as a reason to change the values of the model parameters, but at best as a reason for refining the DSGE model structure. This is a key difference between calibrators in the tradition of Kydland and Prescott and Bayesian econometricians in the tradition of Sims. Finally, the standard errors for the statistics of interest one sometimes sees reported in the literature on calibrated DSGE models do not relate to prior uncertainty about the parameters (because there is none), nor do they relate to the posterior uncertainty about the parameters (because no estimation is involved in their construction). Hence, they cannot be used to construct regions of highest posterior density or to construct confidence intervals in the frequentist sense.

6.3.4 Are Structural VAR Models Less Credible than DSGE Models?

A common misperception among macroeconomists in recent years has been that DSGE models are “structural”, whereas structural VAR models are not. The unspoken assertion is that estimates of structural VAR models are at least debatable, if not flawed, and inherently inferior to estimates of DSGE models. This confusion arises because the term “structure” referred to by these macroeconomists differs from the way this term is used by econometricians. In econometrics, the term “structural model” refers to the conditions required for identifying structural shocks, as formulated by the Cowles Commission in the early 1950s.¹ Among DSGE modelers it refers to explicit assumptions about the market microstructure in the economy, about agents’ constraints, and about the functional form of their objectives.

This DSGE microstructure is not required for the identification of structural shocks. Its potential benefit is that it implies additional cross-equation restrictions on the VAR model, which may enhance the efficiency of the model estimates compared with a VAR model whose lagged coefficients remain unrestricted. These same restrictions, however, may also render the VAR model estimates inconsistent if the microstructure in the DSGE model differs from that in the actual economy. This concern is heightened by the observation that the misspecification of any part of the DSGE model in a general equilibrium setting also invalidates the estimates of parameters in other parts of the DSGE model.

Microfoundations refer to explicit assumptions about economic structure at the level of household and firm decisions, but these assumptions often are merely artificial devices intended to induce more realistic macroeconomic behavior. Examples include lotteries that determine whether a worker is unemployed or whether a firm is allowed to change prices, assumptions that prices or wages remain fixed for a predetermined period of time, the assumption that money by itself generates utility, or that households are subject to cash-in-advance constraints. Often these assumptions are not even remotely plausible, but reflect the need for tractable models or the limits to our modeling skills. The reliance on the representative-agent framework is one example. Another example is the nature of the assumptions about the functional form of aggregate production functions and utility functions. A third example is the ad hoc price and wage setting behavior in standard New Keynesian DSGE models. A fourth example is the simplistic structure of the labor market in DSGE models. Even more troublesome is that not only technology, but also monetary aggregates and government spending are treated as exogenous processes in many DSGE models. Likewise, the dynamic specification of the processes for the

¹ For a review of the Cowles Commission’s approach to econometrics, see Christ (1994). Further discussion can be found in Cooley and Leroy (1985).

exogenous state variables is invariably ad hoc. In short, a model being rigorous in the sense of deriving results from explicit assumptions is not the same as that model being realistic. This does not mean that additional microstructure cannot be helpful in thinking about the economy, but that it is less than obvious that we would wish to impose this structure in estimating VAR models.

The same concern about model misspecification applies when estimating DSGE models directly as an alternative to the estimation of structural VAR models. This is not the only problem, however. There is also strong evidence that the structural parameters in the state-space representation of DSGE models are only weakly identified. Weak identification means that the likelihood is nearly flat across the parameter space, which implies that the posterior of the structural model parameters will be dominated by the prior. Put differently, whatever prior we impose in estimation is also for all practical purposes the posterior we obtain from the data. This observation tells us that we have learned nothing from the data. Evidence of weak identification casts doubt on standard frequentist and Bayesian approaches to estimating DSGE models. Econometricians are only beginning to develop methods of inference that are robust to weak identification problems in DSGE models (see, e.g., Guerron-Quintana, Inoue, and Kilian 2013; Dufour, Khalaf, and Kichian 2013; Qu 2014; Andrews and Mikusheva 2015). Many DSGE model estimates therefore have to be viewed with caution.

Proponents of DSGE models have at times made much of the fact that VAR models require auxiliary assumptions about the lag order and about the data transformations (see Cooley and Dwyer 1998). Because these auxiliary restrictions are atheoretical and effectively untestable, this argument goes, the resulting structural models are not as credible as DSGE models. It is easy to overlook, however, that much the same type of auxiliary assumptions are also required in specifying a DSGE model. For example, we may model the exogenous technology process in a DSGE model alternatively as a trend stationary process or as a process in differences, which is no different from deciding whether to express real GDP in a VAR model in growth rates or in deviations from a deterministic time trend. Moreover, this process may be specified as an AR(1) or an AR(2) process, for example, which is the same problem as choosing the lag order of the VAR model. None of these specification choices are grounded in economic theory.

Finally, an additional auxiliary assumption invoked by many users of DSGE models relates to how the simulated data should be filtered before the unconditional second moments can be computed (see also Chapter 19). DSGE models are designed explicitly as models of the business cycle and abstract from secular growth (see Kydland and Prescott 1982). Models of the business cycle should only be evaluated using data measured at business cycle frequencies,

but observed macroeconomic data fluctuate at higher and lower frequencies, creating a mismatch between the model data and the actual data. For this reason, users of DSGE models often insist on transforming both the actual and the simulated data to isolate the variation at business cycle frequencies. Depending on the specification of the technology process, the choice is usually between deterministic detrending and HP-filtering the seasonally adjusted data or between differencing and HP-filtering, but other forms of filtering the data could also be entertained. This choice is not only ad hoc, but even seemingly reasonable approaches to detrending the simulated model data and the actual data may bias statistical measures of model fit and mask the propagation mechanism at work in the model, as illustrated in Singleton (1988), Diebold and Kilian (2001), and Canova (2014). This means that the “stylized facts” users of DSGE models aim to emulate with calibrated DSGE models are as likely to be artifacts of the filtering method as features of the data. The same concern applies when estimating the parameters of DSGE models on filtered data.

6.3.5 Are DSGE Models More Accurate than VAR Models?

Much of the literature on comparing DSGE models and VAR models has taken the DSGE model as given, while asking whether the VAR model can recover the features of the data generated by the DSGE model. The specification of the VAR model in this type of exercise is intended to reflect the specification of the DSGE model that is presumed to be the DGP. The rationale for this approach is by no means self-evident. As observed by Canova and Ciccarelli (2013, p. 206),

tightly parameterized DSGE models are useful because they offer clear answers to policy questions and provide easy-to-understand welfare prescriptions. However, by construction, they impose a lot of restrictions that are not always consistent with the statistical properties of the data. Thus, the policy prescriptions are hardwired in the assumptions of the model, and must be considered more as a benchmark than a realistic assessment of the options and constraints faced by policymakers in real-world situations.

With equal justification one might ask whether the DSGE model is consistent with a VAR model chosen to fit the data. Reports in the literature about DSGE models fitting the data better than VAR or Bayesian VAR (BVAR) models hence have to be viewed with caution if the specification of the VAR model is artificially constrained to match the specification of the DSGE model.

There are three metrics for judging the relative accuracy of DSGE models and VAR models. One metric is in-sample fit measured, for example, by the marginal likelihood of these models (see Chapter 5). The marginal

likelihood, unlike measures such as R^2 , involves an implicit penalty for model complexity to avoid overparameterization (see Del Negro, Schorfheide, Smets, and Wouters 2007). Del Negro et al. report that the VAR specification subject to the cross-equation restrictions implied by the DSGE model fits the data better than the unrestricted VAR model. This conclusion, of course, applies to the VAR representation of this specific DSGE model only. No other VAR models are considered, so the result cannot be interpreted as this DSGE model being more accurate than the entire class of VAR models.

Another concern is that the priors used in Bayesian DSGE model estimation are heavily influenced by the data to be explained, violating the central premise of Bayesian analysis that the prior should not be influenced by the data (see Chapter 5). This problem arises because Bayesian DSGE models tend to rely on priors that have been found to work well in previous studies using nearly the same data set. Moreover, as the prior evolves more quickly than new data become available, there is a risk of converging to a posterior that fits the data well, perhaps even better than a VAR model, but is nevertheless spurious (see Kilian 2007).

One way of at least reducing this overfitting problem is to focus on the out-of-sample fit of the DSGE model. Until recently, the perception among proponents of DSGE models had been that DSGE models are the preferred structural models, but that reduced-form VAR models remain the benchmark when it comes to forecasting out of sample beyond the end of the estimation sample. Recent work by Del Negro and Schorfheide (2013) has challenged this view. Del Negro and Schorfheide presented evidence that Bayesian DSGE models are more accurate out-of-sample forecasting models than both unrestricted VAR models and BVAR models estimated in the Minnesota tradition. This point is important in that it suggests that the VAR model may no longer be the right benchmark for DSGE model evaluations.

Subsequent research has questioned this assessment. In particular, Gürkaynak, Kisacikoglu, and Rossi (2013) demonstrate that the relatively large-scale VAR and BVAR forecasting models considered by Del Negro and Schorfheide are not a credible forecasting benchmark because they are frequently outperformed by lower-dimensional autoregressive models. Gürkaynak et al. compare Bayesian DSGE models with a range of AR, VAR, and BVAR forecasting models for inflation, growth, and interest rates. They find that in real time no forecasting model is uniformly more accurate than the others. The relative performance of the models differs greatly over time, across forecast horizons, and across the variables to be forecast.

The third metric for comparing DSGE models and SVAR models is how close the implied structural impulse response functions are. There is in fact an entire literature on impulse response matching between DSGE models and VAR models based on minimum distance estimators (see, e.g., Rotemberg and Woodford 1997; Christiano, Eichenbaum, and Evans 2005; Dridi, Guay,

and Renault 2007; Hall, Inoue, Nason, and Rossi 2012; Guerron-Quintana, Inoue, and Kilian 2017). A comparison of the structural impulse responses makes sense, however, if and only if the identification of the structural shocks in the VAR model is valid and is consistent with the structure imposed in the DSGE model. In practice, it can be difficult to find identification schemes that are consistent with the DSGE model and simultaneously identify an entire vector of structural shocks in a high-dimensional VAR model.

Some researchers have addressed this problem by imposing structure on the impact multiplier matrix of the VAR that comes from the fitted DSGE model (see Del Negro and Schorfheide 2004; Del Negro, Schorfheide, Smets, and Wouters 2007). This approach bypasses the issue of the identification of the structural VAR shocks, but requires a high degree of confidence in the DSGE model estimate. Moreover, Sims (2007) observes that the results of Del Negro and Schorfheide's procedure can be sensitive to the ordering of the VAR variables. Other researchers have resorted to lower-dimensional models and/or models that are only partially identified. A popular choice are models that identify technology shocks based on long-run identifying restrictions. Such models, however, are not without their own limitations, as explained in more detail in Chapters 10 and 11.

6.3.6 Policy Analysis in DSGE Models and SVAR Models

Another common misperception among macroeconomists is that structural VAR models cannot be used to conduct meaningful policy experiments. The central question in this debate is whether estimates of structural VAR models remain stable following government policy interventions. The Lucas critique suggests that only deep parameters (such as parameters representing tastes, preferences, and perhaps technological constraints) that govern the behavior of economic agents remain unchanged in response to policy interventions. This result seems to imply that reduced-form VAR representations cannot be expected to remain stable in response to policy interventions. This line of reasoning is frequently invoked as an argument (a) against relying on SVAR models and (b) for relying on DSGE models. Neither argument is compelling in general.

The first point to keep in mind is that Lucas' argument may apply to changes in policy regimes within the VAR framework, but typically does not apply to policy surprises within a given policy regime. As long as the policy surprises are within the range of historical experience, agents will have formed expectations that allow for these policy surprises and there is no reason for the reduced-form model to change (see, e.g., Sims 1986; Leeper and Zha 2003). To the extent that policy counterfactuals and forecast scenarios based on structural VAR models can be mapped into sequences of structural shocks, as discussed in Chapter 4, this reasoning may be applied even to changes in policy regimes.

Some changes in the policy regime may be too dramatic to avoid the Lucas critique of reduced-form VAR models, but others will involve policy shocks that are well within the range of historical experience and are not predictable. The analysis of the latter policy regime shifts will not be in obvious conflict with the Lucas critique.

A second implication of the Lucas critique is that the analysis of monetary policy VAR models, in particular, ought to focus on samples during which the policy reaction function remained stable. Fitting monetary policy VAR models on data from the early 1970s until today, for example, cannot be expected to generate meaningful results. This problem may be addressed by splitting the sample. Limiting the sample to homogenous periods, of course, reduces the sample available for estimation and inference. Another possible response to this problem therefore is the use of time-varying coefficient VAR models or of Markov-switching VAR models that allow the policy regime to change over time, as discussed in Chapters 18 and 19 (see, e.g., Sims and Zha 2006b; Sims, Waggoner, and Zha 2008).

It is perhaps less widely appreciated that estimates of micro-founded DSGE models are not immune from the Lucas critique either. A case in point is the invalidity of estimates of time-invariant DSGE models in the presence of shifts in monetary policy regimes. For example, fitting a DSGE model with a standard Taylor policy rule on data ending in 2014 without accounting for the shift of monetary policy toward quantitative easing in 2008 would not produce meaningful estimates. Any misspecification of the monetary policy rule in the estimated DSGE models invalidates the estimates of all model parameters, just like in VAR models.

The sense in which a DSGE model may in principle be used to address the Lucas critique is that estimates of the deep parameters of a DSGE model subject to one policy regime (obtained from a sample devoid of any policy regime shifts) may be used to predict agents' behavior in a different policy regime. The prediction may be constructed by evaluating the DSGE model that incorporates the alternative policy regime conditional on the estimates of the deep parameters from the estimated model. This reasoning is based on the premise that deep parameters are invariant to changes in the policy regime and hence not subject to the Lucas critique. This premise is not uncontroversial (see Fernández-Villaverde and Rubio-Ramírez 2007). It also relies on the model structures in question being an accurate representation of reality before and after the policy regime shift. Finally, it hinges on the researcher's ability to estimate the deep parameters reliably. Evidence that deep parameters in commonly used DSGE models are only weakly identified casts doubt on the latter premise (see, e.g., Canova and Sala 2009).

We conclude that mechanical applications of either approach should be avoided. DSGE models and structural VAR models are complementary, with each approach having its own strengths and weaknesses. There is no basis for claims that one approach dominates the other.

6.4 An Overview of Alternative Structural Macroeconometric Models

The following table compares the three main approaches to empirical macroeconomics in the literature: the DSGE model, the DSEM, and the structural VAR (SVAR) model.

Features	DSGE	DSEM	SVAR
Exogeneity restrictions	Few	Many	None
Dynamic exclusion restrictions	Few	Many	Few
Number of variables	Large	Very large	Small
Number of shocks	Few	Many	Few
Trend treatment	Explicit	Implicit	Explicit
Microstructure required	Yes	No	No

The table highlights that each approach has its own strengths and weaknesses. For example, traditional DSEMs require many unrealistic exogeneity restrictions, whereas SVAR models require no exogeneity assumptions at all. DSGE models, on the other hand, require some unrealistic exogeneity restrictions. Likewise, DSEMs require many dynamic exclusion restrictions, whereas SVAR models are unrestricted except for an upper bound on the lag order. DSGE models in turn do not directly restrict the dynamics of the endogenous model variables, but restrict the dynamics of the underlying exogenous state variables.

While DSEMs allow the inclusion of a large number of variables, standard VAR models and to a lesser extent DSGE models are much more limited in scope. They also have to make do with a smaller number of shocks. Both DSGE models and SVAR models explicitly account for the trending behavior of many macroeconomic time series, whereas in DSEMs the trends are often implicit. Finally, only DSGE models require explicit assumptions about the model's market microstructure including the specification of the deep parameters governing the preferences of households and production possibilities of firms, the choice of functional forms for utility and production functions, and the specification of market structures (e.g., perfect competition, monopolistic competition, imperfect competition).

It is not surprising in light of this comparison that there have been attempts to combine DSEMs and SVAR models as well as SVAR models and DSGE models in an effort to improve their performance.

6.4.1 Combining DSEMs and SVAR Models

In the 1980s and 1990s, many central banks experimented with modernized versions of DSEMs. Drawing on insights from time series analysis, these

models paid more attention to the trend specification, relaxed the dynamic specification of the errors, embodied restrictions that preserve the long-run comovement of many variables, and reduced the number of variables and equations. These changes, however, also made it more difficult to interpret the estimates economically because of the uneasy coexistence of reduced-form and structural elements in the same model. These models appear to have been abandoned in favor of estimated DSGE models in recent years.

At the same time, VAR users remained cognizant of the fact that standard VAR models are not able to compete with DSEM models in providing detailed answers about the response of the economy to structural shocks. Leeper, Sims, and Zha (1996), among others, exploited advances in computational techniques to build SVAR models with up to 20 variables. They relied on informative Bayesian priors to deal with the curse of dimensionality (see Chapter 5). Nevertheless, they could not overcome skepticism about their models' ability to extract useful information from the data, and such medium-scale VAR models never gained traction in empirical macroeconomics. More recently, there have been other efforts to extend the VAR framework to even larger-dimensional data sets, however. Examples are large-scale Bayesian VAR models and factor augmented VAR (FAVAR) models. These developments are reviewed in Chapter 16.

6.4.2 Combining DSGE and SVAR Models

With the demise of DSEMs in academic research, the focus shifted in the direction of combining DSGE models and SVAR models. Assuming that a VAR approximation to the DSGE model exists, the key difference between VAR models and DSGE models is that the latter imply additional cross-equation restrictions on the parameters of the VAR model that standard VAR models ignore. Unrestricted VAR models have the advantage that they are potentially consistent with a whole class of DSGE models. Their disadvantage is that additional structure, if correct, can reduce the VAR estimation errors in small samples. Much of the literature on combining DSGE and SVAR models has focused on the potential benefits of imposing DSGE structure on VAR estimates.

Early efforts to estimate small-scale SVAR models subject to restrictions implied by rational expectations models were abandoned because these restrictions were routinely rejected by the data (see, e.g., Keating 1990). Subsequently, King, Plosser, Stock, and Watson (1991) argued for imposing on VAR models the cointegration structure implied by RBC models with random walk technology shocks, resulting in a vector error correction model. Interest in forging a closer relationship between VAR and DSGE models increased a decade later, after DSGE models had become larger and embedded enough shocks and frictions to generate time series that look like actual

data. Del Negro and Schorfheide (2004) in particular proposed priors for the estimation of VAR models that were derived from a New Keynesian DSGE model (see also Chapter 5). The extent to which such priors improve the accuracy of VAR estimates continues to be debated.

More recently, skepticism toward the structure of DSGE models has given rise to less parametric approaches of combining DSGE models and VAR models. For example, Giacomini and Ragusa (2014) advocate forcing the VAR model to satisfy the nonlinear equilibrium conditions in the DSGE model by exponential tilting. This approach is designed to impose only a subset of equations from the DSGE model, allowing us to discount DSGE model equations with less theoretical content or that we have less faith in. Similarly, Canova and Paustian (2011) rely on results from a broad range of DSGE models to derive identifying restrictions on the signs of structural impulse responses. This approach is discussed in more detail in Chapter 13.

7 A Historical Perspective on Causal Inference in Macroeconometrics

The central objective in structural VAR analysis is to quantify causal relationships in the data. Before discussing the identification of causal relationships in structural VAR models, it is useful to review the precursors to structural VAR analysis. Our discussion traces how the focus of the literature has evolved from documenting lead-lag patterns in the data, as discussed in Sections 7.2–7.4, to quantifying unanticipated shifts in the data reflecting exogenous events, as discussed in Section 7.5. There are several approaches to constructing such exogenous shocks. We review the narrative approach to measuring exogenous policy shocks, the derivation of exogenous shocks from data-based counterfactuals, the construction of news shocks from macroeconomic announcements, and the measurement of shocks to financial market expectations. The definition of exogenous shocks was generalized with the introduction of the structural VAR framework, as discussed in Section 7.6. The latter approach is based on decomposing fluctuations in the data that cannot be predicted based on past data into mutually uncorrelated exogenous shocks with economic interpretation that need not be directly observable. As we trace the evolution of this literature, we also formally introduce the concepts of predeterminedness, strict exogeneity, and Granger causality, highlighting the extent to which each approach relies on these concepts.

7.1 A Motivating Example

The need for structural models in studying causal relationships between economic time series is best illustrated by the debate about causality from monetary aggregates to national income in the 1960s and 1970s. It had long been observed that money growth and income growth in the United States were positively correlated. Based on a careful review of the historical evidence, Friedman and Schwartz (1963) in their *Monetary History of the United States* concluded that changes in money growth are causing changes in income growth (an obvious implication being that the Federal Reserve should pursue

a constant money growth rule to stabilize the business cycle). This position evolved into a school of thought known as monetarism. Monetarism emphasizes the relation of the level of the money stock to the level of aggregate real economic activity (see Sims 1980b).

The monetarist position contrasted with the prevailing Keynesian wisdom that monetary policy was not nearly as important as fiscal policy in explaining economic fluctuations. Keynesians responded to Friedman and Schwartz by making the case that monetary aggregates were passive and that changes in the growth of monetary aggregates were endogenous responses to changes in real economic activity. Because higher output requires more “grease” in the form of money to keep the economy going, it is changes in real activity that cause endogenous changes in monetary aggregates through the money multiplier of the banking system (see Tobin 1970).

In its simplest form, monetarism is a statement about contemporaneous correlations between changes in real money stocks and real income. Because this correlation, while consistent with monetarist theory, is easy to explain away as a passive response of the money stock to changes in real activity, Friedman and Schwartz stressed the historical tendency for movements in the money stock (or its rate of change) to precede movements in aggregate activity. This additional implication of monetarist theory is harder to explain as a passive response of the money stock to changes in real activity and, hence, is considered a more challenging test of monetarist theory.

Amid this sometimes heated debate between monetarists and Keynesians, a new time series methodology emerged in the early 1970s that promised an answer to questions of causality and soon enjoyed considerable popularity. This methodology was developed by Clive Granger, among others, and the statistical tests in question became widely known as Granger causality tests (see Granger 1969; Sims 1972). More precisely, Granger’s proposal was to test the null hypothesis of no (Granger) causality, with (Granger) causality being implied by the rejection of the null, as discussed in Chapter 2.

7.2 Granger Causality Tests for Covariance Stationary VAR Models

For expository purposes, consider the bivariate money-income autoregression

$$\begin{pmatrix} \Delta m_t \\ \Delta n_t \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \begin{bmatrix} a_{11,1} & a_{12,1} \\ a_{21,1} & a_{22,1} \end{bmatrix} \begin{pmatrix} \Delta m_{t-1} \\ \Delta n_{t-1} \end{pmatrix} + \dots \\ + \begin{bmatrix} a_{11,p} & a_{12,p} \\ a_{21,p} & a_{22,p} \end{bmatrix} \begin{pmatrix} \Delta m_{t-p} \\ \Delta n_{t-p} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix},$$

where Δm_t denotes money growth and Δn_t denotes growth in national income. Then $H_0 : a_{12,1} = \dots = a_{12,p} = 0$ means that Δn_t does not Granger cause Δm_t , whereas $H_0 : a_{21,1} = \dots = a_{21,p} = 0$ means that Δm_t does not Granger

cause Δn_t . A Granger causality test can be conducted as a Wald test of the null hypothesis of no Granger causality. The possible results are:

1. Granger noncausality cannot be rejected in either direction.
2. Unidirectional Granger causality.
3. Bidirectional Granger causality.

Evidence of unidirectional Granger causality from money growth to income growth by many economists at the time would have been taken as evidence of Friedman and Schwartz being right about a causal role for money growth. Upon closer inspection, it is clear that Granger causality actually means that, on average over the sample period, including lags of one variable helps reduce the squared error in predicting another variable. It does not imply causality, only precedence. In other words, movements in one variable on average predate (or lead) movements in the other. The key difference between predictability and causality is that statistical precedence alone is not exploitable by policymakers. Separating cause and effect instead requires a structural model. This point is best illustrated by examples.

Example 1. A recent study by Shimmura and Yoshimura (2013) shows that roosters have internal clocks that make them crow before the sun comes up, even when kept in the dark. Hence, the crowing of a rooster helps predict the sunrise. If this were also a causal relationship, then a deliberate intervention such as strangling the rooster should prevent the sun from coming up. We know this not to be the case, so the predictive relationship is not causal.

In this example the data would have suggested Granger causality (precedence), yet it is clear that there is no causality in the sense we usually have in mind. The example was chosen to make the point that there is an obvious difference between precedence and causality. No one would confuse the two concepts in this example. Now consider a similar example drawn from economics.

Example 2. Most oil price increases have been followed by U.S. recessions. Hence, oil price increases cause recessions.

The difference is that now it is not immediately obvious whether the conclusion that oil price increases cause recessions is true or false, but it is clear that the logic that this statement appeals to in deriving its conclusion is flawed. The logical flaw is known as the *post hoc, ergo propter hoc* fallacy. In other words, precedence does not necessarily imply causality.

Granger and Newbold (1977, p. 225) acknowledge this drawback in the concept of Granger causality, but dismiss it: “Possibly cause is too strong a term, or one too emotionally laden, to be used. A better term might be temporally related, but since cause is such a simple term we shall continue to use it”. Granger (1980, p. 333) elaborates that, “provided I define what I personally

mean by causation, I can use the term. I could, if I so wish, replace the word ‘cause’ throughout my lecture by some other words, such as ‘oshkosh’ or ‘snerd’, but what would be gained? It is like saying that whenever I use x , you would prefer me to use z .¹

This cavalier attitude has not gone unchallenged. For example, Leamer (1985, p. 284) in a discussion of Granger’s work strongly objects to Granger taking control of the English language for his own purposes: “If I were to continue in that tradition I would propose that we henceforth refer to this notion of precedence by the wordpair: ‘fool’s causation’. This substitutes a loaded word ‘fool’ for the neutral ‘Granger’, just as causation has replaced the neutral precedence. Moreover, ‘fool’ is decidedly simpler than ‘Granger’ – it contains only four letters, one of which is repeated – and like ‘cause’, it is rather difficult to define precisely. One man’s fool is another man’s genius. My definition of a ‘fool’ would be a friend of mine living in San Diego.” The last sentence is a reference to Granger who at the time was a professor at UC San Diego.

7.3 Granger Causality, Predeterminedness, and Exogeneity

The deeper reason why the distinction between predictability and causality matters is that predictive relationships need not be exploitable for policy purposes. It is useful to define more formally what we mean by a causal relationship and how causality relates to the concepts of Granger causality, predeterminedness, and strict exogeneity.¹

7.3.1 Basic Concepts

The following discussion builds on Cooley and Leroy (1985). As before, let Δm_t denote money growth and Δn_t growth in national income. For expository purposes, consider the structural model:

$$\begin{aligned}\Delta m_t &= \theta \Delta n_t + \beta_{11} \Delta m_{t-1} + \beta_{12} \Delta n_{t-1} + w_{1t}, \\ \Delta n_t &= \gamma \Delta m_t + \beta_{21} \Delta m_{t-1} + \beta_{22} \Delta n_{t-1} + w_{2t},\end{aligned}$$

where w_{1t} and w_{2t} refer to mutually uncorrelated white noise innovations. Given this structural model, we can define the notions of predeterminedness and strict exogeneity.

Definition 1. Δm_t is *predetermined* for Δn_t if $\theta = 0$.

¹ Engle, Hendry, and Richard (1983) generalized the notion of exogeneity to possibly nonlinear models. In doing so they changed the terminology. Their notion of *weak exogeneity* in linear models corresponds to predeterminedness. Their notion of *strong exogeneity* in linear models reduces to the concept of strict exogeneity. Given our focus on linear models, we use the traditional definitions.

Definition 2. Δm_t is (*strictly*) **exogenous** with respect to Δn_t if $\theta = \beta_{12} = 0$ (so there is neither contemporaneous nor lagged feedback from Δn_t to Δm_t).

The notion of strict exogeneity includes that of predeterminedness by construction. It is immediately apparent that the notion of predeterminedness captures what economists have in mind when referring to causality because a policymaker controlling Δm_t also controls Δn_t . Predeterminedness allows us to interpret a correlation between Δm_t and Δn_t as evidence of a causal effect of Δm_t on Δn_t , provided Δm_t is not correlated with any other predetermined variable that is not included in the bivariate model above. For example, if we think of the change in the price of oil as predetermined with respect to income growth and if this variable is correlated with Δm_t , the correlation between Δm_t and Δn_t can no longer be used to quantify the causal effect of Δm_t on Δn_t .

The notion of Granger causality may be defined based on the VAR(1) reduced-form representation of the structural model above:

$$\Delta m_t = a_{11,1} \Delta m_{t-1} + a_{12,1} \Delta n_{t-1} + u_{1t},$$

$$\Delta n_t = a_{21,1} \Delta m_{t-1} + a_{22,1} \Delta n_{t-1} + u_{2t}.$$

Definition 3. Δn_t fails to **Granger cause** Δm_t if $a_{12,1} = 0$.

The first question is what Granger noncausality tells us about predeterminedness and strict exogeneity.

Predeterminedness. Granger noncausality is neither necessary nor sufficient for predeterminedness because

$$a_{12,1} \equiv \frac{\theta \beta_{22} + \beta_{12}}{1 - \theta \gamma} = 0 \quad \not\Rightarrow \quad \theta = 0.$$

Clearly, $\theta \beta_{22} = -\beta_{12}$ can hold for $\theta \neq 0$. Moreover,

$$\theta = 0 \quad \not\Rightarrow \quad a_{12,1} \equiv \frac{\theta \beta_{22} + \beta_{12}}{1 - \theta \gamma} = 0.$$

For example, $a_{12,1} \neq 0$ for $\theta = 0$ and $\beta_{12} \neq 0$. We conclude that one variable being Granger noncausal for the other does not imply that the latter variable is predetermined with respect to the former. Conversely, one variable being predetermined with respect to the other does not necessarily imply Granger noncausality from the latter variable to the former variable.

Next we turn to the relationship between strict exogeneity and Granger causality.

Strict exogeneity. Similarly,

$$a_{12,1} \equiv \frac{\theta \beta_{22} + \beta_{12}}{1 - \theta \gamma} = 0 \quad \not\Rightarrow \quad \theta = \beta_{12} = 0$$

because $\theta\beta_{22} = -\beta_{12}$ can hold for $\theta \neq 0$ and $\beta_{12} \neq 0$. Intuitively, Granger noncausality cannot imply exogeneity, because predeterminedness is part of the definition of exogeneity. However, strict exogeneity implies Granger noncausality because

$$\theta = \beta_{12} = 0 \quad \Rightarrow \quad a_{12,1} \equiv \frac{\theta\beta_{22} + \beta_{12}}{1 - \theta\gamma} = 0.$$

This means that by testing Granger noncausality we can test one of the implications of exogeneity. If we reject Granger noncausality, we also reject strict exogeneity. If we do not reject Granger noncausality, we learn nothing about whether strict exogeneity holds.

Sometimes the objective of conducting a Granger causality test is to establish that a variable is exogenous. Unfortunately, this result is precisely what a Granger noncausality test is unable to establish. Moreover, establishing exogeneity is useful if we want to justify exclusion restrictions in a structural model, but is stronger than what is needed to study the effects of policy interventions. For the latter purpose, predeterminedness suffices, which, however, is inherently untestable within the VAR framework.²

In practical terms, we conclude that even if money unidirectionally Granger causes income in a VAR model, this does not necessarily mean that the economy will grow faster, if the money supply is increased. This fact has not prevented many researchers to this day from misinterpreting Granger causality tests as genuine tests of causality. While they have learned not to refer to “causality” explicitly, and often acknowledge at some point that Granger causality does not imply causality, when it comes to the substantive results of their work, they simply replace the statement “ X causes Y ” by substantively identical statements such as “ X is responsible for Y ,” “ X explains Y ,” “ X influences Y ,” “ X is the source of Y ,” or “ X has an effect on Y .” This practice is incorrect and misleading, but it illustrates the lure of answering the causality question.

7.3.2 Granger Causality and Forward-Looking Behavior

Questions of causality are just as interesting in the context of financial economics. An important result is that financial asset prices will tend to Granger cause macroeconomic aggregates (but not the other way around) even in the absence of a causal relationship. The following example from Hamilton (1994, Example 11.1, pp. 306–307) illustrates this point. Recall that an investor who buys one share of a stock for the price P_t at date t , and on date $t + 1$ receives a

² Kilian and Vega (2011) provide an example of how high-frequency data on U.S. macroeconomic news may be used to address the question of the predeterminedness of energy prices with respect to U.S. macroeconomic aggregates.

dividend D_{t+1} and sells the stock for P_{t+1} , receives an ex post rate of return of

$$r_{t+1} \equiv \frac{P_{t+1} + D_{t+1}}{P_t} - 1.$$

Hamilton postulates that the expected rate of return is a constant r such that

$$(1 + r)P_t = \mathbb{E}_t(P_{t+1} + D_{t+1}).$$

This first-order difference equation in P_t implies (after some algebra):

$$P_t = \mathbb{E}_t \left(\sum_{j=1}^{\infty} \left(\frac{1}{1+r} \right)^j D_{t+j} \right)$$

provided we rule out speculative bubbles in P_t . This means that the stock price today reflects the market's anticipation of future dividends. In other words, expectations of dividends drive the stock price in this model.

Next, Hamilton shows that P_t and D_t may be modeled as a bivariate VAR process. For concreteness, suppose that dividends follow the process

$$D_t = d + u_t + \delta u_{t-1} + v_t,$$

where d denotes a constant and u_t and v_t are mutually independent white noise processes. Hamilton observes that

$$\mathbb{E}_t(D_{t+j}) = \begin{cases} d + \delta u_t, & j = 1 \\ d, & j = 2, 3, \dots \end{cases}$$

He substitutes this expectation of future dividends into the expression for P_t above:

$$\begin{aligned} P_t &= \mathbb{E}_t \left(\sum_{j=1}^{\infty} \left(\frac{1}{1+r} \right)^j d \right) + \mathbb{E}_t \left(\frac{1}{1+r} \delta u_t \right) \\ &= \frac{d}{r} + \frac{\delta u_t}{1+r} \quad \forall t, \end{aligned}$$

where $\sum_{j=1}^{\infty} \left(\frac{1}{1+r} \right)^j = \sum_{j=0}^{\infty} \left(\frac{1}{1+r} \right)^j - 1 = 1/r$. From

$$P_t = \frac{d}{r} + \frac{\delta u_t}{1+r}$$

it follows that

$$P_{t-1} = \frac{d}{r} + \frac{\delta u_{t-1}}{1+r} \Leftrightarrow \delta u_{t-1} = (1+r)P_{t-1} - (1+r)\frac{d}{r}.$$

Finally, Hamilton substitutes for δu_{t-1} in the expression for D_t :

$$\begin{aligned} D_t &= d + u_t + \delta u_{t-1} + v_t \\ &= d + u_t + (1+r)P_{t-1} - (1+r)\frac{d}{r} + v_t. \end{aligned}$$

Hence, the expressions for P_t and D_t can be written in VAR(1) format as:

$$\begin{pmatrix} P_t \\ D_t \end{pmatrix} = \begin{bmatrix} d/r & 0 \\ -d/r & 1+r \end{bmatrix} \begin{pmatrix} P_{t-1} \\ D_{t-1} \end{pmatrix} + \begin{pmatrix} \delta u_t/(1+r) \\ u_t + v_t \end{pmatrix}.$$

The VAR representation shows that D_t fails to Granger cause P_t , but P_t Granger causes D_t . This direction of Granger causality may seem surprising given that stock prices in the model depend on the expected path of dividends. Certainly, higher stock prices do not actually cause dividends to increase. Rather, this pattern of Granger causality arises because stock prices are determined by forward-looking agents who anticipate future dividends. This example provides some intuition for the tendency of time series that reflect forward-looking behavior such as asset prices or interest rates to be excellent predictors of macroeconomic time series such as real GDP or inflation. This clearly does not mean that these time series cause GDP or inflation to move. Rather, asset prices incorporate the market's assessment of where real GDP and inflation are headed.

7.3.3 Strict Exogeneity in Modern Macroeconomic Models

As discussed earlier, one of the few practical uses of Granger causality tests is for rejecting the strict exogeneity of economic variables. With the demise of traditional DSEMs and the emergence of SVAR models with fully endogenous variables, this question lost much of its urgency. The profession quickly accepted the notion that macroeconomic aggregates are unlikely to be strictly exogenous. One exception is macroeconomic models for small open economies. It has remained common to treat macroeconomic aggregates abroad as exogenously determined from the point of view of small open economies (see, e.g., Cushman and Zha 1997).

For a long time, the most promising candidate for an exogenous variable in macroeconomic models had been the price of crude oil. For example, Hamilton (2003) made the case that fluctuations in the price of crude oil are exogenous with respect to the U.S. economy to the extent that they are driven by politically motivated disruptions of oil production in the Middle East. This interpretation was subsequently questioned by Barsky and Kilian (2001) who challenged the exogeneity of some of the political events considered by Hamilton, and highlighted instabilities in the relationship between arguably exogenous oil supply disruptions and oil price increases. Subsequently, Kilian (2008a; 2008b) established that exogenous oil supply shocks in the Middle

East lack predictive power for the price of oil. Finally, Alquist, Kilian, and Vigfusson (2013) demonstrated that after 1973 the nominal price of oil is Granger caused by both U.S. and global macroeconomic aggregates, while the real price of oil is Granger caused by measures of global macroeconomic aggregates. At this point, there is no doubt that even the price of oil is not strictly exogenous, and a literature has evolved on modeling the endogeneity of oil prices both empirically and theoretically.

While strictly exogenous variables are hard to find in macroeconomics, the strict exogeneity assumption has survived in DSGE models which postulate that variables such as technology growth, money supply growth, or government spending growth are strictly exogenous. These variables are treated as exogenous mainly because of our inability or unwillingness to model them, not because there is evidence supporting the strict exogeneity assumption. Indeed, the endogeneity of U.S. money growth with respect to the real economy had been stressed by Keynesians already in the 1960s, and the endogeneity of government spending with respect to the business cycle is self-evident. Even the exogeneity of technology shocks is not obvious. For example, the theory of economic growth is devoted to making technology shocks endogenous with respect to the economy.

7.4 The Demise of Granger Causality Tests in Macroeconomics

As the profession developed a better understanding of the meaning of Granger causality tests and their statistical properties, the debate over whether money Granger causes income evolved. Initially, economists thought that bivariate Granger causality tests established that money leads income, but that result weakened once more variables were included in the VAR model. It also proved highly sensitive to different forms of detrending and to changes in the model specification. Moreover, it was shown that apparent Granger noncausality from money to income may simply reflect the omission of a third variable, whereas a finding of bivariate Granger causality may likewise reflect the omission of a third variable, calling into question even further the usefulness of Granger causality tests (see Lütkepohl 1982b).

By the late 1980s, the consensus was that money probably Granger causes income, but that this result does not prove monetarists right. At best it can be considered an empirical regularity that a macroeconomic model should be able to explain, not unlike a correlation pattern. As a result, the profession lost interest in Granger causality tests. Granger causality tests were replaced by the new idea that we are not interested in the contemporaneous correlation between money and income growth or the lead-lag pattern, but in the question of how income responds to unanticipated changes in money growth (also known as innovations or shocks). The hope was that these shocks would be exogenous, even if the underlying money growth time series was not.

7.5 Responses to Unanticipated Changes in Money Growth

This new idea of focusing on unanticipated changes in money growth as measures of exogenous shocks emerged in four different guises: (1) an effort to identify exogenous changes in monetary policy based on the narrative approach to monetary policy; (2) efforts to measure the news (or surprise) component of monetary aggregates; (3) efforts to recover shocks to expectations from the prices of futures contracts; and (4) efforts to identify monetary policy shocks within the context of a structural VAR model.

7.5.1 *The Narrative Approach*

The narrative approach to identifying monetary policy shocks dates back to Friedman and Schwartz (1963). Not content to rely on statistical evidence only, Friedman and Schwartz used historical records to provide evidence for the existence of major swings in the real money stock which not only precede major swings in real activity but cannot themselves be explained as endogenous responses to changes in real activity. This approach was subsequently extended and formalized by Romer and Romer (1989), who introduced a dummy variable that takes on a value of 1 during periods when, according to the Romers' reading of the U.S. monetary policy records, the Federal Reserve exogenously tightened monetary policy, and zero otherwise. Their analysis involved a careful reading of statements of policymakers to determine whether a given change in monetary policy (say, an increase in the Federal Funds rate) represented an endogenous response to macroeconomic developments or an exogenous tightening of monetary policy.³

The original Romer dates included October 1947, September 1955, December 1968, April 1974, August 1978, and October 1979. Subsequently, they added December 1988 to this list. In related work, Kashyap, Stein, and Wilcox (1993) suggested February 1966 as another candidate and Oliner and Rudebusch (1995) proposed August 1988. To the extent that these Romer dummies, as they have come to be known, represent exogenous shocks, we may use distributed-lag models of the form

$$\Delta n_t = \nu + \sum_{i=0}^m \phi_i d_{t-i} + \varepsilon_t,$$

³ The Romer shock measures are distinct from measures of the monetary policy stance that measure whether the Federal Reserve is leaning toward a contractionary or an expansionary policy. Examples of the latter measures include Boschen and Mills' (1995) quantitative dummy measure of the intensity of monetary policy stance (expressed on a scale of $\{-2, -1, 0, +1, +2\}$) and Bernanke and Mihov's (1998b) VAR-based index which is measured on a continuous scale. These measures of stance are not measures of exogenous shocks, but of policy actions. In other words, they are endogenous with respect to the state of the macroeconomy.

where d_t denotes the dummy variable and Δn_t denotes income growth, to estimate the impulse responses directly by LS, where ε_t may be serially correlated and/or heteroskedastic, necessitating the use of robust standard errors in constructing t or Wald test statistics. The impulse response in the distributed-lag model simply is:

$$\frac{\partial \Delta n_t}{\partial d_{t-i}} = \frac{\partial \Delta n_{t+i}}{\partial d_t} = \phi_i$$

and confidence bands for each horizon i can be constructed based on each coefficient's robust standard errors. The number of lags determines the maximum horizon of the response functions. Note that the distributed-lag model may be viewed as a special case of the final-form representation of a dynamic simultaneous equations model (e.g., Lütkepohl 2005, section 10.2.2). For these impulse response estimates to capture the causal effects of policy interventions it must be the case that the dummies are truly exogenous, of course, and that they are uncorrelated with alternative sources of exogenous variation in Δn_t .

While the distributed-lag model is the simplest approach to estimating the responses in question, it is not the only one. In fact, Romer and Romer (1989), rather than estimating a distributed-lag model, report results based on the LS estimates of a single-equation model that includes additional lags of the dependent variable:

$$\Delta n_t = \nu + \sum_{i=0}^m \gamma_i d_{t-i} + \sum_{i=1}^p \beta_i \Delta n_{t-i} + \epsilon_t,$$

where ϵ_t may be heteroskedastic, but p has been chosen to render the error term serially uncorrelated. This alternative model can be motivated based on the premise that Δn_t and d_t are jointly determined by a structural VAR model of the form

$$\begin{aligned} d_t &= \theta \Delta n_t + \beta_{11} d_{t-1} + \beta_{12} \Delta n_{t-1} + w_{1t}, \\ \Delta n_t &= \gamma d_t + \beta_{21} d_{t-1} + \beta_{22} \Delta n_{t-1} + w_{2t}, \end{aligned}$$

where we have set $p = m = 1$ for expository purposes and suppressed all deterministic regressors. Romer and Romer (1989) postulate that (1) d_t is strictly exogenous with respect to Δn_t such that $\theta = 0$ and $\beta_{12} = 0$, and that (2) d_t is serially uncorrelated such that $\beta_{11} = 0$. Hence,

$$\begin{aligned} d_t &= w_{1t}, \\ \Delta n_t &= \gamma d_t + \beta_{21} d_{t-1} + \beta_{22} \Delta n_{t-1} + w_{2t}. \end{aligned}$$

One can compute $\partial \Delta n_{t+i} / \partial d_t$ from this restricted VAR model using the standard tools for impulse response analysis discussed in Chapter 4. This derivation

is not quite correct, however, in that d_t is a binary variable, which is at odds with the definition of the linear VAR model in Chapter 2.

One problem with the Romer dummies is that they are few in numbers, so the accuracy of the response estimates tends to be low. A second problem is that they make no allowance for the magnitude of the exogenous policy intervention. Third, they ignore episodes in which monetary policy expanded exogenously. The fourth and most important problem is that estimates of probit and logit models suggest that d_t is predictable based on past and/or expected macroeconomic aggregates and hence not exogenous with respect to the U.S. economy (see Shapiro 1994; Leeper 1997). This evidence suggests that Romer and Romer did not succeed in isolating the exogenous component of the change in monetary policy.

In response to this concern, Romer and Romer (2004) presented a refined measure of monetary policy shocks, which is constructed as the residual of a regression of the change in the intended federal funds rate around the dates of Federal Open Market Committee (FOMC) meetings on a set of predictors intended to purge the endogenous component of this series. If there are no meetings in a given month, the monetary policy shock is set to zero. The set of predictors used to distinguish exogenous shifts in monetary policy from policy responses to the expected state of the economy, includes the Federal Reserve's internal forecasts of inflation, real output growth, and the unemployment rate, as published in the Federal Reserve's Green Book. The resulting modified monetary policy shocks are available at monthly frequency for 1969m1-1996m12 and may take on positive, zero, or negative values. Constructing such modified monetary policy shocks, of course, creates a generated regressor problem that is typically ignored in applied work (see Pagan 1984). Quarterly measures of these monetary policy shocks are constructed by summing the monthly policy shocks by quarter.

The narrative approach has also been used for quantifying fiscal policy shocks. As far as government spending is concerned, Ramey and Shapiro (1998) introduce binary dummies for shocks to U.S. government spending associated with exogenous military buildups. Ramey (2011) refines this approach. Her work uses newspaper sources to quantify the changes in the expected present value of government spending associated with military buildups, resulting in a quantitative dummy variable that takes on nonzero values (measured on a continuous scale) on dates selected based on extraneous information and is zero otherwise.

On the government revenue side, Romer and Romer (2010) use the legislative records to construct dummies for tax changes. Their analysis allows them to separate legislated changes into those taken for reasons related to prospective economic conditions and those taken for exogenous reasons. Exogenous nominal tax changes are expressed as a fraction of nominal GDP in the quarter the change occurred, resulting in a quantitative dummy series.

Fiscal dummies can be used for econometric analysis much the same way as monetary dummies. It is also common to include quantitative fiscal dummies as exogenous variables in VAR models (e.g., Ramey 2011; Mertens and Ravn 2012, 2013), as discussed in Chapter 15.

This literature has evolved further in recent years. A number of authors have defined fiscal policy shocks as an event consisting of multiple components, some of which are implemented immediately and some of which are left to be implemented in the future. An example is Alesina, Favero, and Giavazzi (2015). These components must be viewed in conjunction when assessing the response of the variable of interest. In other words, the fiscal policy shock is multidimensional rather than a scalar. Responses to such fiscal shocks may be constructed as the average difference between the simulated path of the variable of interest with and without all components of the plan implemented (see also Chapter 18).

A third area in which the narrative approach has been popular is the literature on oil supply shocks. Dotsey and Reid (1992) and Hoover and Perez (1994) introduce binary dummies for the dates of exogenous shortfalls in the production of oil in member countries of the Organization of Petroleum Exporting Countries (OPEC). Hamilton (2003) proposes quantitative dummies for measuring exogenous shortfalls in OPEC oil production, defined to be negative on a continuous scale on dates of exogenous political events in the Middle East and zero otherwise.

7.5.2 Exogenous Shocks Derived from Data-Based Counterfactuals

Although the narrative approach to identifying shocks has gained some prominence in measuring monetary policy shocks, oil supply shocks, and, more recently, fiscal policy shocks, it is but one example of methodologies aimed at identifying exogenous variation in the data. For example, Kilian (2008b) designs a methodology for quantifying the extent of exogenous variation in OPEC oil production from explicit counterfactuals about how oil production would have evolved in the absence of exogenous political events such as wars or revolutions. Related work includes Bastianin and Manera (2017). This approach results in a time series of the evolution of the exogenous component of OPEC oil production rather than a quantitative dummy. No estimation is involved, so there is no generated regressor problem. Changes in this exogenous series are serially uncorrelated, allowing the estimation of impulse responses from distributed-lag models as well as restricted VAR models. Another prominent example is the measure of exogenous technology shocks proposed by Basu, Fernald, and Kimball (2006). Their analysis controls for variation in capacity utilization in an effort to isolate the exogenous component of aggregate productivity.

7.5.3 News Shocks

A closely related idea in the literature is that we can identify exogenous shocks by comparing the announcements of macroeconomic data releases with measures of what the market expected prior to the release. One advantage of this approach is that the news component of the announcement is unanticipated by construction. Another advantage is that we can obtain measures of news shocks at daily frequency, which makes this approach very useful when working with daily asset returns, r_t , for example. The news shock for a variable z_t after the announcement of its latest data release, denoted by $z_t^{\text{announcement}}$, is defined as

$$z_t^{\text{news}} \equiv z_t^{\text{announcement}} - \mathbb{E}_{t-1}(z_t^{\text{announcement}})$$

and may be standardized for better comparability across different types of news.

Of course, news shocks are only as good as the underlying measures of ex ante expectations. Typically, we obtain those expectations from market surveys taken the day before the announcement rather than from econometric models. In some cases, futures prices have been used as measures of expectations. There are data releases for about 30 different U.S. macroeconomic time series, including the latest industrial production data, housing starts, retail sales, construction spending, manufacturing orders, consumer confidence, initial unemployment claims, and other leading indicators of real activity. There also are news releases for monetary aggregates such as the federal funds target rate, allowing us to quantify monetary policy surprises. Given that news shocks are constructed at very high frequency, we can think of macroeconomic news shocks as plausibly exogenous shocks to the variables in question.

Again, the distributed-lag model

$$r_t = v + \sum_{i=0}^m \phi_i z_{t-i}^{\text{news}} + \varepsilon_t,$$

may be applied to estimate the impulse response

$$\frac{\partial r_{t+i}}{\partial z_t^{\text{news}}} = \phi_i.$$

Often the presumption is that asset prices should be informationally efficient in that news should be absorbed within the same day. That means that we do not need to worry about including lags of z_t^{news} :

$$r_t = v + \phi_0 z_t^{\text{news}} + \varepsilon_t.$$

When running regressions on news shocks, we sometimes stack the observations only for those time periods for which we have a news shock observation. This makes this type of model different from standard distributed-lag models. The sample size thus corresponds to the number of news shocks rather

than T . The model lends itself to t and Wald tests of whether news shocks matter. This distributed-lag model again may be viewed as a special case of the final form representation of a dynamic simultaneous equations model.

A good example for the literature on news shocks is Andersen, Bollerslev, Diebold, and Vega (2003). This paper studies the real-time price discovery in foreign exchange markets. The question is whether daily foreign exchange returns respond to the news component of macroeconomic data releases. The answer is yes. This finding is in striking contrast to the usual random walk result that monthly foreign exchange returns cannot be predicted based on monthly macroeconomic aggregates.⁴

Often it is difficult to find data on agents' expectations for the variable of interest. This fact has prompted the development of a closely related approach exploiting data from professional forecasters. For example, Kilian and Hicks (2013) define exogenous shocks to real GDP based on the revisions to professional real GDP forecasts. Similarly, Ramey (2011) proposes a measure of government spending shocks based on the errors made by professional forecasters.

It should be noted that news shock regressions tend to be based on daily (or even intra-daily) data, making it difficult to retain power when extrapolating to horizons longer than 20 business days. Thus, this framework is rarely used for characterizing the monthly or quarterly shocks of interest to macroeconomists. It has been used, however, to provide evidence for the assumption that monthly U.S. oil prices and gasoline prices are predetermined with respect to the U.S. macroeconomy (see Kilian and Vega 2011).

7.5.4 Shocks to Financial Market Expectations

An alternative approach to measuring exogenous shocks to expectations about policy variables or about other variables of interest is to rely on financial markets. In the absence of a risk premium, under standard assumptions the price of a futures contract of maturity h , F_t^h , in equilibrium equals the conditional expectation of the spot price, S_{t+h} . In practice, there is evidence of risk premia in most futures markets that drives a wedge between the futures price and the conditional expectation of the spot price. This fact suggests adjusting the futures price by an estimate of the risk premium, \widehat{RP}_t^h , to construct a time series of financial market expectations,

$$\mathbb{E}_t [S_{t+h}] = F_t^h - \widehat{RP}_t^h.$$

⁴ It should be added that news shocks as defined in this literature are distinct from the more recent use of this term in macroeconomics, originating with Beaudry and Portier (2006) (see Chapter 10). The latter literature is about shifts in expectations, not about news properly defined as in this section. Clearly, news need not be linearly related to shifts in expectations.

A generalization of this approach is discussed in Baumeister and Kilian (2016c). By comparing the market expectations for horizon h with the corresponding realizations, monthly time series of shocks to the market expectations may be obtained from financial market data (see Baumeister and Kilian 2016a).

One problem is that shocks to interest rate expectations over the course of a month, for example, are not necessarily tied to monetary policy decisions. They may occur for many reasons. Piazzesi and Swanson (2008) suggest that in measuring monetary policy shocks we may simply focus on changes in F_t^h from the day preceding a policy announcement to the following day. Under the premise that the risk premium evolves only slowly over time, changes in the risk premium from one day to the next are likely to be negligible. This allows the construction of a daily time series of monetary shocks without having to estimate the risk premium at daily frequency. For related discussion, see also Rudebusch (1998), Kuttner (2001), and Cochrane and Piazzesi (2002), among others. This approach, which may be extended to intra-daily frequencies, of course raises the question of how to recover the monthly and quarterly policy shocks of interest to macroeconomists. In practice, various proposals have been made to scale and aggregate high-frequency policy shocks for use in VAR models. These proposals also form the basis of the nonstandard VAR models of monetary policy discussed in Chapter 15.

7.5.5 Summary

This discussion shows that there are many creative ways of constructing measures of exogenous shocks that can be used to quantify causal relationships in the data. While most studies in this literature rely on least-squares estimates of the effects of these shocks from distributed-lag models or VARX models, some studies have used exogenous shocks as instruments. For example, Hamilton (2003) and Kilian (2008a, 2008b) report results for single-equation instrumental variable (IV) estimates based on exogenous OPEC oil supply shocks. The latter two studies draw attention to the problem that exogenous shocks may be weak instruments, invalidating conventional IV analysis. Similar issues may arise when including exogenous shock series in VAR models. A formal analysis of estimating VAR models with weak external instruments can be found in Montiel Olea, Stock, and Watson (2015a). This approach is discussed in Chapter 15.

7.6 Structural VAR Shocks

Although direct measures of policy shocks, news shocks based on macroeconomic announcements, shock measures based on financial market expectations, and estimates of shocks based on data-based counterfactuals have been

used extensively in applied work, the range of questions that can be analyzed with these tools is limited. By far the most common approach to estimating economically meaningful shocks has been to rely on structural vector autoregressions, building on the work of Sims (1980a).⁵ The premise of this approach is that the DGP can be approximated by a structural VAR model of the form

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t, \quad (7.6.1)$$

where the deterministic terms have been suppressed, y_t is a $K \times 1$ vector of model variables, the model coefficients B_i , $i = 0, \dots, p$, are $K \times K$ matrices, and the elements of the $K \times 1$ vector w_t are mutually uncorrelated white noise with nonsingular diagonal covariance matrix Σ_w , with one or more of the elements of w_t having a distinct economic interpretation. This model is recognizable as a special case of the DSEM model of Chapter 6, in which the coefficients of the lagged variables have been left unrestricted.

By pre-multiplying both sides of equation (7.6.1) by B_0^{-1} , we obtain the corresponding reduced-form VAR representation,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t, \quad (7.6.2)$$

where $A_i = B_0^{-1} B_i$, $i = 1, \dots, p$, and $u_t = B_0^{-1} w_t$, the estimation of which has been the subject of Chapters 2, 3, and 5.

7.6.1 The Identification Problem

It is readily apparent that knowledge of B_0^{-1} suffices to recover the parameters and shocks of the structural model (7.6.1), given a consistent estimate of the reduced-form model (7.6.2). The estimation of B_0^{-1} requires the user to impose additional identifying restrictions on B_0 or B_0^{-1} that can be motivated based on economic theory, institutional knowledge, or other external constraints on the structural model. Imposing these additional identifying restrictions allows one to decompose the reduced-form errors u_t into mutually uncorrelated structural shocks, w_t , with an economic interpretation.

It is important to keep in mind that without a clear economic interpretation of the elements of w_t , model (7.6.1) would not be structural. In particular, it is not enough for the elements of w_t to be mutually uncorrelated. Some early VAR applications overlooked this requirement and relied on ad hoc assumptions for identification that made no economic sense. Such atheoretical VAR models attracted strong criticism, spurring the development of more explicitly structural VAR models starting in the mid-1980s (see, e.g., Cooley and Leroy 1985).

⁵ Although structural VAR models were not used in applied work prior to the 1980s, their antecedents can be traced back to the pioneering work of Mann and Wald (1943).

In response to ongoing questions about the validity of commonly used identifying assumptions, the structural VAR model literature has continuously evolved since the 1980s. The next chapters trace the evolution of this literature. We focus on alternative approaches to the identification of structural shocks within the framework of a reduced-form VAR model, highlighting the conditions under which each approach is valid and discussing potential limitations of commonly employed methods. For example, Chapter 8 focuses on identification by short-run restrictions. Chapter 10 reviews identification by long-run restrictions. Identification by sign restrictions is discussed in Chapter 13. Chapters 14 and 15 summarize alternative approaches of achieving identification by exploiting heteroskedasticity in the data, by using external instruments, or based on high-frequency changes in futures prices. For now we set these issues aside and focus on the interpretation of structural VAR shocks, assuming that an estimate of B_0^{-1} can be obtained.

7.6.2 The Relationship between Structural VAR Shocks and Direct Shock Measures

Knowledge of the mapping from the reduced-form VAR representation to the structural VAR representation allows us to quantify the structural shocks, given that $w_t = B_0 u_t$. Compared with direct measures of exogenous shocks of the type discussed in Section 7.5, a key difference is that the structural shocks in model (7.6.1) are in general unobservable and need not be associated with any one VAR model variable in particular, which greatly increases the range of applications of the VAR approach. This point is best illustrated by example. Consider a prototypical microeconomic model of the price and quantity of a commodity, in which the observables price (p_t) and quantity (q_t) are driven by latent demand shocks (w_t^d) and supply shocks (w_t^s). This model may be written as a structural VAR model of the form

$$\begin{bmatrix} b_{11,0} & b_{12,0} \\ b_{21,0} & b_{22,0} \end{bmatrix} \begin{pmatrix} q_t \\ p_t \end{pmatrix} = \begin{bmatrix} b_{11,1} & b_{12,1} \\ b_{21,1} & b_{22,1} \end{bmatrix} \begin{pmatrix} q_{t-1} \\ p_{t-1} \end{pmatrix} + \dots \\ + \begin{bmatrix} b_{11,p} & b_{12,p} \\ b_{21,p} & b_{22,p} \end{bmatrix} \begin{pmatrix} q_{t-p} \\ p_{t-p} \end{pmatrix} + \begin{pmatrix} w_t^s \\ w_t^d \end{pmatrix},$$

or more compactly as

$$B_0 u_t = w_t \iff \begin{bmatrix} b_{11,0} & b_{12,0} \\ b_{21,0} & b_{22,0} \end{bmatrix} \begin{pmatrix} u_t^q \\ u_t^p \end{pmatrix} = \begin{pmatrix} w_t^s \\ w_t^d \end{pmatrix},$$

where u_t is the error of the reduced-form VAR representation, which also may be interpreted as the unanticipated change in p_t and q_t caused by the structural

shocks in period t . This expression may be rearranged as

$$u_t = B_0^{-1} w_t \iff \begin{pmatrix} u_t^q \\ u_t^p \end{pmatrix} = \begin{bmatrix} b_0^{11} & b_0^{12} \\ b_0^{21} & b_0^{22} \end{bmatrix} \begin{pmatrix} w_t^s \\ w_t^d \end{pmatrix}.$$

In this model, the observables price and quantity are simultaneously determined by demand and supply shocks, allowing us to express the unpredictable component of the data, u_t^q and u_t^p , as a weighted average of the w_t^s and w_t^d with the weights provided by the rows of B_0^{-1} , as illustrated by the last equation.

By construction, each structural shock affects the price and the quantity of the commodity in question simultaneously. Only if the model is recursive such that $b_0^{12} = b_{12,0} = 0$, rendering the quantity variable predetermined with respect to the price, is there a direct link between one of the structural shocks and one of the variables, as assumed in Section 7.5.

7.6.3 Causality in Structural VAR Models

The ultimate question of interest in structural VAR analysis is not the evolution of the structural shocks, but their causal effects on the model variables, as captured by impulse response functions (see Chapter 4). The nature of this link is best illustrated in the context of the model of demand and supply.

Structural shocks within this framework may be interpreted as exogenous shifts of the demand and supply curves in the underlying economic model. For example, a positive demand shock could be represented as an exogenous shift of the demand curve to the right along the supply curve. Thus, a demand shock traces out the slope of the supply curve, as captured by $b_{12,0}$. One of the requirements for a causal interpretation of the responses of future values of price and quantity to a demand or supply shock, respectively, in the current period, is that these shocks are mutually uncorrelated. Obviously, if w_t^d were correlated with w_t^s , both curves would shift, making it impossible to interpret the resulting changes in prices and quantities as the causal effect of either a demand shock or a supply shock.

Treating the demand and supply shocks in this example as exogenous makes sense, because they cannot be predicted based on past data and are not correlated with one another. Implicitly, we are assuming that the observed price and quantity data are generated by these two structural shocks only. This assumption is reasonable if the structural VAR model is correctly specified. There are counterexamples, however. For example, if the price and quantity data underlying this model were determined by the actions of economic agents whose expectations of future prices and quantities differ from the predictions implied by this VAR model, this particular structural VAR model would be invalid. This situation is discussed in Chapter 17. In some cases, this problem may be overcome by the use of suitable external instruments that embody the missing

information, as discussed in Chapter 15, or by specifying VAR models that allow for more information, as discussed in Chapters 16 and 17.

Structural VAR models tend to involve much richer structures than this stylized model of demand and supply, of course, but this example illustrates how structural VAR models may be used to quantify causal relationships in the data.

8 Identification by Short-Run Restrictions

8.1 Introduction

Consider a K -dimensional time series y_t , $t = 1, \dots, T$. We postulate that the DGP for y_t is well-approximated by a vector autoregression of finite order p . Our objective is to learn about the parameters of the structural vector autoregressive model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t, \quad (8.1.1)$$

where w_t denotes a mean zero serially uncorrelated error term, also referred to as a structural innovation or structural shock. The error term is assumed to be unconditionally homoskedastic, unless noted otherwise. The nonsingular matrix B_0 governs the contemporaneous interaction between the model variables. All deterministic regressors have been suppressed for notational convenience. Equivalently the model can be written more compactly as

$$B(L)y_t = w_t,$$

where $B(L) \equiv B_0 - B_1 L - B_2 L^2 - \cdots - B_p L^p$ is the autoregressive lag polynomial. The variance-covariance matrix of the structural error term is typically normalized such that

$$\mathbb{E}(w_t w_t') \equiv \Sigma_w = I_K.$$

This means, first, that there are as many structural shocks as variables in the model. Second, structural shocks by definition are mutually uncorrelated, which implies that Σ_w is diagonal. Third, we normalize the variance of all structural shocks to unity. The latter normalization does not involve a loss of generality as long as the diagonal elements of B_0 remain unrestricted. We defer a discussion of alternative normalizations until the end of this section.

For model (8.1.1) to be considered a structural VAR model it is not sufficient for the elements of w_t to be uncorrelated. In addition, these shocks must be

economically interpretable. It is worth noting that, in general, structural shocks do not correspond to particular model variables and, hence, have no natural unit of measurement. For example, in a VAR system consisting of only price and quantity, both the demand shock and the supply shock shift the price as well as the quantity. In fact, if price and quantity variables were mechanically associated with price and quantity shocks, this would be an indication that the proposed model is not truly structural.

Next we derive the reduced-form representation of this structural VAR model. This involves expressing y_t as a function of lags of y_t only. After pre-multiplying both sides of the structural VAR representation by B_0^{-1} ,

$$B_0^{-1}B_0y_t = B_0^{-1}B_1y_{t-1} + \cdots + B_0^{-1}B_py_{t-p} + B_0^{-1}w_t,$$

the model can be represented in reduced form as

$$y_t = A_1y_{t-1} + \cdots + A_py_{t-p} + u_t,$$

where $A_i = B_0^{-1}B_i$, $i = 1, \dots, p$, and $u_t = B_0^{-1}w_t$. The reduced-form innovations u_t are a weighted average of the structural shocks w_t . Equivalently this model can be written more compactly as

$$A(L)y_t = u_t,$$

where $A(L) \equiv I_K - A_1L - A_2L^2 - \cdots - A_pL^p$ denotes the autoregressive lag polynomial. Standard estimation methods allow us to obtain consistent estimates of the reduced-form parameters A_i , $i = 1, \dots, p$, the reduced-form innovations u_t , and their covariance matrix $\mathbb{E}(u_t u_t') \equiv \Sigma_u$ (see Chapter 2).

Having estimated the reduced-form model, the question arises of how to recover the structural representation of the VAR model. Knowledge of B_0 enables us to reconstruct w_t from $w_t = B_0u_t$ and B_i , $i = 1, \dots, p$, from $B_i = B_0A_i$. The central question therefore is how to recover the elements of B_0 (or its inverse) from consistent estimates of the reduced-form parameters.

By construction, $u_t = B_0^{-1}w_t$. Hence, the variance of u_t is

$$\Sigma_u = \mathbb{E}(u_t u_t') = B_0^{-1}\mathbb{E}(w_t w_t')B_0^{-1'} = B_0^{-1}\Sigma_w B_0^{-1'} = B_0^{-1}B_0^{-1'}, \quad (8.1.2)$$

where we made use of $\Sigma_w = I_K$. We can think of $\Sigma_u = B_0^{-1}B_0^{-1'}$ as a system of nonlinear equations in the unknown parameters of B_0^{-1} . Observe that due to the symmetry of Σ_u , (8.1.2) represents a system of $K(K+1)/2$ independent equations only. The reduced-form error covariance matrix Σ_u can be estimated consistently and hence is treated as known for the time being. The system of nonlinear equations (8.1.2) can be solved for the unknown parameters in B_0^{-1} using numerical methods, provided the number of unknown parameters in B_0^{-1} does not exceed the number of independent equations in (8.1.2). This

involves imposing additional restrictions on selected elements of B_0^{-1} . Such restrictions may take the form of exclusion restrictions, proportionality restrictions, or other equality restrictions. The most common approach is to impose exclusion restrictions on selected elements of B_0^{-1} by forcing these elements to be zero.

To verify that all of the elements of the unknown matrix B_0^{-1} are uniquely identified, observe that Σ_u has $K(K + 1)/2$ free parameters. This follows from the fact that any covariance matrix is symmetric about the main diagonal. Hence, $K(K + 1)/2$ by construction is the maximum number of parameters in B_0^{-1} that one can uniquely identify. This order condition for just identification is easily checked in practice by counting the unrestricted elements of B_0^{-1} . It is a necessary condition for identification only, however.

Even if the order condition is satisfied, the system of equations may fail to have a unique solution. To ensure identification of the structural shocks, the system, in addition, has to satisfy the rank condition for identification. Rubio-Ramírez, Waggoner, and Zha (2010) discuss a general approach to evaluating the rank condition for global identification in structural VAR models. For further discussion of the identification issue, see also Christiano, Eichenbaum, and Evans (1999) and Taylor (2004).

It should be noted that even if all elements of w_t are uniquely identified in a statistical sense, they need not be uniquely identified in the economic sense. For example, there are structural VAR models in which all elements of w_t are statistically uniquely identified, but only some of the elements of w_t are also economically identified.

The earlier discussion alluded to the existence of alternative normalization assumptions in structural VAR analysis. There are three equivalent representations of structural VAR models that differ only in how the model is normalized. All three representations have been used in applied work. In the discussion thus far we made the standard normalizing assumption that $\Sigma_w = I_K$, while leaving the diagonal elements of B_0 unrestricted. Identification was achieved by imposing identifying restrictions on B_0^{-1} . By construction, a unit innovation in the structural shocks in this representation is an innovation of a magnitude of one standard deviation, so structural impulse responses are responses to one-standard deviation shocks.

Equivalently, one could have left the diagonal elements of Σ_w unconstrained and set the diagonal elements of B_0 to unity in $w_t = B_0 u_t$ (see, e.g., Keating 1992). However, the variance of the structural errors will no longer be unity under these assumptions, so the implied responses to shocks must be rescaled by one standard deviation of the structural residual to ensure that the implied structural impulse responses represent responses to one-standard deviation shocks. Under this alternative normalization, the order condition for just identification requires B_0 to have $K(K - 1)/2$ restrictions, not counting the

restrictions on the diagonal elements. Which convention one uses is usually dictated by the nature of the economic identifying assumptions. In some cases it is more natural to restrict B_0^{-1} and in others B_0 .

Finally, these two approaches may be combined by changing notation and writing the model equivalently as

$$B_0 u_t = C w_t$$

with $\Sigma_w = I_K$ such that $\Sigma_u = B_0^{-1}CC'B_0^{-1'}$. The two representations above emerge as special cases of this representation with the alternative normalizations of $B_0 = I_K$ or $C = I_K$. The advantage of this third representation is that it allows one to relax the assumption that either $C = I_K$ or $B_0 = I_K$, which sometimes facilitates the exposition of the identifying assumptions.¹ If the diagonal elements of B_0 are normalized to unity and $\Sigma_w = I_K$, the order condition requires $K^2 + K(K - 1)/2$ restrictions on B_0 and C combined, including the restrictions on the diagonal elements of B_0 . Likewise, if the diagonal elements of both B_0 and C are restricted to unity instead and the diagonal elements of Σ_w are left unrestricted, the order condition calls for $K^2 + K(K + 1)/2$ restrictions, including the normalizing restrictions on the diagonals. For example, Blanchard and Perotti (2002) use this representation with the diagonal elements of both B_0 and C normalized to unity, but neither C nor B_0 diagonal.

8.2 Recursively Identified Models

One popular way of disentangling the structural innovations w_t from the reduced-form innovations u_t is to “orthogonalize” the reduced-form errors. Orthogonalization here means making the errors mutually uncorrelated. Mechanically, this can be accomplished as follows. Define the lower-triangular $K \times K$ matrix P with positive main diagonal such that $PP' = \Sigma_u$. The matrix P is known as the lower-triangular Cholesky decomposition of Σ_u .² It follows immediately from the condition $\Sigma_u = B_0^{-1}B_0^{-1'}$ that $B_0^{-1} = P$ is one possible solution to the problem of how to recover w_t . Since P is lower triangular, it has $K(K - 1)/2$ zero parameters. As a result, the order condition for the exact identification of the unknown parameters in B_0^{-1} is satisfied. A useful result in this context is that B_0^{-1} being lower triangular implies that B_0 is lower triangular as well.

¹ The latter representation is sometimes referred to as the AB-model with $Au_t = Bw_t$, the model with $B = I$ as the A-model, and the model with $A = I$ as the B-model (see, e.g., Lütkepohl 2005, section 9.1). There is no universally accepted notation, however. Amisano and Giannini (1997), for example, refer to the AB-model, K-model and C-model.

² Standard software provides built-in functions for generating the Cholesky decomposition of Σ_u .

It is important to keep in mind that the orthogonalization of the reduced-form residuals by applying a Cholesky decomposition is appropriate only if the recursive structure embodied in P can be justified on economic grounds. The distinguishing feature of orthogonalization by Cholesky decomposition is that the resulting structural model is recursive (conditional on lagged variables). This means that we impose a particular causal chain rather than learning about the nature of the causal relationships from the data. In essence, we solve the problem of which structural shock causes the variation in u_t by imposing a particular solution. This mechanical solution does not make economic sense, however, without a plausible economic interpretation for the recursive ordering.

The neutral and scientific-sounding term “orthogonalization” hides the fact that we are making strong identifying assumptions about the error term of the VAR model. In the early 1980s, many users of VAR models did not appreciate this point and thought the data alone would speak for themselves. Such “atheoretical” VAR models soon were severely criticized (see, e.g., Cooley and Leroy 1985). This critique prompted more careful attention to the economic underpinnings of recursive models. It was shown that in special cases the recursive model can be given a structural or semistructural interpretation (e.g., Bernanke and Blinder 1992). This critique also spurred the development of structural VAR models that impose nonrecursive identifying restrictions (e.g., Sims 1986; Bernanke 1986; Blanchard and Watson 1986). It became widely recognized that the structural VAR model is simply a special case of the DSEM, the main distinguishing feature of the structural VAR model being the nature of its identifying restrictions.

In practice, there is a different solution P for each ordering of the K variables in the VAR model. It is sometimes argued that one should conduct sensitivity analysis based on alternative orderings of the K variables. This proposal is problematic for three reasons:

1. On the one hand, we claim to be sure that the ordering is recursive, yet on the other hand we claim to have no clue in what order the variables are recursive. This approach is not credible.
2. For a small VAR model with $K = 4$, for example, there are $4 \cdot 3 \cdot 2 \cdot 1 = 24$ permutations of the ordering. Few researchers would be willing to try out this many model specifications, nor would there be much hope that the results would be the same in each case unless the reduced-form errors are uncorrelated, which can be checked by inspecting the off-diagonal elements of Σ_u .
3. Even if there were no difference across these 24 specifications, this would only prove that the results are robust across all recursive orderings, but there is no reason for the true model to be recursive in the first place. This point is best illustrated by example. Let p_t denote the

price and q_t the quantity of a good. Price and quantity are driven by structural demand shocks w_t^d and supply shocks w_t^s . All dynamics are suppressed for expository purposes such that $y_t = u_t$:

$$\underbrace{\begin{pmatrix} p_t \\ q_t \end{pmatrix}}_{u_t} = \underbrace{\begin{bmatrix} 1 & -0.5 \\ 0.5 & 1 \end{bmatrix}}_{B_0^{-1}} \underbrace{\begin{pmatrix} w_t^d \\ w_t^s \end{pmatrix}}_{w_t}.$$

In this example, by construction, Σ_u is diagonal and the observable data are uncorrelated such that all recursive orderings are identical. This outcome obviously does not imply that any of the recursive orderings are valid. In fact, B_0^{-1} differs from

$$P = \text{chol}(\Sigma_u) = \text{chol}\left(\begin{bmatrix} 1.25 & 0 \\ 0 & 1.25 \end{bmatrix}\right) = \begin{bmatrix} 1.118 & 0 \\ 0 & 1.118 \end{bmatrix}$$

by construction. This point holds more generally. Let $u_t = B_0^{-1}w_t$ denote the true structural relationship and $u_t = Pw_t^{chol}$ be the Cholesky relationship. Then

$$w_t^{chol} = P^{-1}u_t = P^{-1}(B_0^{-1}w_t) \neq w_t,$$

so the Cholesky decomposition will fail to identify the true structural shocks unless $B_0^{-1} = P$.

8.3 Sources of Identifying Restrictions

In the preceding subsection we stressed that unless we can come up with a convincing rationale for a particular recursive ordering, the resulting VAR impulse responses, forecast error variance decompositions, and historical decompositions are economically meaningless. This raises the question of where the economic rationale of the identifying restrictions on B_0^{-1} or B_0 comes from. There are a number of potential sources.

One is economic theory. In some cases, we may wish to impose the structure provided by a specific economic model. Of course, in that case the empirical results will only be as credible as the underlying model. A case in point is Blanchard's (1989) structural VAR analysis of the traditional Keynesian model involving an aggregate demand equation, Okun's law, a price-setting equation, the Phillips curve, and a monetary policy rule. Another example is Sims and Zha (2006a) who propose a VAR model with identifying assumptions that are approximately satisfied in a fully specified DSGE model. A third example is Fisher (2006) who derives his VAR identifying assumptions from a specific real business cycle model.

Another strategy is to specify an encompassing model that includes as special cases various alternative structural models implied by different economic

models, allowing tests for overidentifying restrictions. The advantage of this approach is that it avoids conditioning on one specific model that may be incorrect. Of course, this type of structural VAR model no longer admits a Cholesky representation and must be estimated by alternative numerical methods (see Chapter 9). This strategy has been used, for example, by Bernanke and Mihov (1998b) who model the market for bank reserves as part of a study of U.S. monetary policy.

Often there is no fully developed theoretical model available, in which case identification may be achieved by using extraneous information or by using selective insights from economic theory:

Information delays. Information may not be available instantaneously because data are released only infrequently, allowing us to rule out instantaneous feedback. This approach has been employed by Inoue, Kilian, and Kiraz (2009), for example, who exploit differences in the timing of data releases to motivate the imposition of short-run exclusion restrictions. They observe that Survey of Professional Forecasters (SPF) inflation forecasts for the current quarter are formed before the interest rate for the current quarter is set (reflecting their release in the middle of the preceding quarter), while household survey expectations of the real interest rate are formed after observing the nominal interest rate at the beginning of the quarter. This reasoning suggests a recursive structure with the SPF forecast ordered first, followed by the interest rate and households' real interest rate expectations. This structure allows us to assess the impact of inflation news contained in the SPF forecasts on household expectations of the inflation rate. A similar argument is invoked by Sims (1998) who suggests that monetary policymakers react immediately only to variables that they can observe without delay (such as commodity prices, monetary aggregates, and financial variables), but not to variables that they can observe only with a delay (such as real GDP or the GDP deflator).

Physical constraints. For example, it takes time for a firm's investment decision to be made and for the new equipment to be installed, so measured physical investment responds only with a delay to investment decisions. This delay motivates an exclusion restriction of the impact response of investment to new information about the economy. This type of friction is consistent with the use of time-to-build technologies in many DSGE models. More generally, it is common to postulate that variables such as real GDP, industrial production, and real investment are inherently sluggish and hence do not respond contemporaneously to monetary policy shocks (see Sims 1998).

Institutional knowledge. For example, we may have information about the inability of suppliers to respond to demand shocks in the short run due

to adjustment costs, which amounts to imposing a vertical slope on the short-run supply curve (see, e.g., Kilian 2009). Similarly, Davis and Kilian (2011) exploit the fact that gasoline taxes (excluding ad valorem taxes) do not respond instantaneously to the state of the economy because lawmakers move at a slow pace.

Assumptions about market structure. Another common identifying assumption in empirical work is that there is no feedback from a small open economy to the rest of the world. This identifying assumption has been used, for example, to motivate treating U.S. interest rates as contemporaneously exogenous with respect to the macroeconomic aggregates of small open economies such as Canada (see, e.g., Cushman and Zha 1997).

Homogeneity of demand functions. Another possible source of identifying information are homogeneity restrictions on demand functions. For example, Galí (1992) imposes short-run homogeneity on the demand for money when assuming that the demand for real balances is not affected by contemporaneous changes in prices (given the nominal rate and output). This assumption amounts to assuming away costs of adjusting nominal money holdings. Similar homogeneity restrictions have also been used in Bernanke (1986). Likewise, Keating (1992) postulates that money holdings rise in proportion to nominal income such that the response to a change in real income is the same as the response to a change in the price level.

Extraneous parameter estimates. When the elements of B_0 can be viewed as elasticities, it may be possible to impose values for those elasticities based on extraneous information from other studies. This approach has been used by Blanchard and Perotti (2002), for example. Similarly, Blanchard and Watson (1986) impose nonzero values for some structural parameters in B_0 based on extraneous information. If the parameter value cannot be pinned down with any degree of reliability, yet another possibility is to explore a grid of possible structural parameter values, as in Abraham and Haltiwanger (1995) and Blanchard (1989). A similar approach has also been used in Kilian (2010) and Davis and Kilian (2011) in an effort to assess the robustness of their baseline results. In a different context, Todd (1990) interprets Sims' (1980b) recursive VAR model of monetary policy in terms of alternative assumptions about the slopes of money demand and money supply curves.

High-frequency data. In rare cases, it may be possible to motivate exclusion restrictions more directly. For example, Kilian and Vega (2011) use daily data on U.S. macroeconomic news to formally test the identifying assumption of no feedback within the month from U.S. macroeconomic aggregates to the dollar price of oil and the U.S. price of gasoline. They demonstrate that this test is unable to reject the null of no feedback for

these energy prices, while having enough power to reject the same null for a large number of asset prices. This result lends credence to exclusion restrictions in monthly VAR models ruling out instantaneous feedback from U.S. macroeconomic aggregates to the U.S. prices of oil and gasoline.

It is fair to say that coming up with a set of credible short-run identifying restrictions is difficult. Whether a particular exclusion restriction is convincing often depends on the data frequency, and in many cases there are not enough credible exclusion restrictions to achieve full identification. This fact has stimulated interest in the alternative identification methods discussed in subsequent chapters.

8.4 Examples of Recursively Identified Models

8.4.1 A Simple Macroeconomic Model

Let $y_t = (p_t, gdp_t, m_t, i_t)'$, where p_t is the log price level, gdp_t is log real GDP, m_t the log of a monetary aggregate such as M1, and i_t the federal funds rate. The data are quarterly and the proposed identification is recursive such that

$$\begin{pmatrix} u_t^p \\ u_t^{gdp} \\ u_t^m \\ u_t^i \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} & 0 \\ b_0^{41} & b_0^{42} & b_0^{43} & b_0^{44} \end{bmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \\ w_{4t} \end{pmatrix}. \quad (8.4.1)$$

Note that each line can be viewed as an equation. Each reduced-form shock is a weighted average of selected structural shocks. The letters b_0^{ij} represent the weights attached to the structural shocks. For example, the first equation is $u_t^p = b_0^{11}w_{1t} + 0 + 0 + 0$, the second reads $u_t^{gdp} = b_0^{21}w_{1t} + b_0^{22}w_{2t} + 0 + 0$, etc.

One way of rationalizing this identification would be to interpret the first two equations as an aggregate supply (AS) and aggregate demand (AD) model with a horizontal AS curve and downward-sloping AD curve. w_{1t} moves the price level and real output, so it must be a shift of the AS curve. w_{2t} moves real output only, so it must represent a shift of the AD curve. The third equation could be interpreted as a money demand equation derived from the quantity equation: $MV = PY$, where V stands for velocity and Y for real income. Hence, w_{3t} can be interpreted as a velocity shock or money demand shock, if we take real GDP to represent real income. The last equation could represent a monetary policy reaction function. The Federal Reserve systematically responds to u_t^p , u_t^{gdp} , and u_t^m (as well as lags of all variables). Any change in the interest rate not accounted for by this response would be an exogenous monetary policy (or money supply) shock. Such policy shocks could arise from

changes in the composition of the Federal Open Market Committee, for example, or may reflect reactions to idiosyncratic events such as the 9/11 terrorist attack on the United States or the housing crisis that are not captured by standard policy rules.

It is easy to spot the limitations of this model. For example, why does money demand not respond to the interest rate within a quarter? How plausible is the assumption of a horizontal short-run supply curve? These are the types of questions that one must ask when assessing the plausibility of a structural VAR model. This example also illustrates that theory typically is not sufficient for identification, even if we are willing to condition on a particular theoretical model. For example, if the AS curve were vertical, but the AD curve horizontal by assumption, the first two equations of the structural model (8.4.1) would have to be modified. More generally, no recursive structure would be able to accommodate a theoretical model in which the AS and AD curves are neither horizontal nor vertical, but upward and downward sloping. This point highlights the difficulty of specifying fully structural models of the macroeconomy in recursive form and explains why such models have declined in importance in macroeconomic analysis.

8.4.2 A Model of the Global Market for Crude Oil

The second example is a structural VAR model of the global market for crude oil based on Kilian (2009). Let $y_t = (\Delta prod_t, rea_t, rpoil_t)'$, where $\Delta prod_t$ denotes the percent change in world crude oil production, rea_t is a business cycle index measuring global real economic activity, and $rpoil_t$ is the log of the real price of oil. The data are monthly. Let

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} \end{bmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix}.$$

This model of the global market for crude oil embodies a vertical short-run oil supply curve and a downward-sloping short-run oil demand curve (conditional on lags of all variables). There are two demand shocks that are separately identified by the delay restriction that oil-specific demand shocks raise the real price of oil, but without affecting global real economic activity within the same month, while shocks to the aggregate demand for all industrial commodities affect both real activity and the real price of oil on impact.

One might question whether one could have imposed an overidentifying restriction of the form $b_0^{21} = 0$. In other words, one would expect that higher oil prices triggered by unanticipated oil supply disruptions would not slow down global real activity within the month any more or less than oil-specific demand shocks. It turns out that the estimate of b_0^{21} is essentially zero, even without

imposing that restriction, making this point moot.³ One could also question whether the short-run supply curve is truly vertical. Defending this assumption requires institutional knowledge of oil markets or extraneous econometric evidence. Additional theoretical support for this assumption is provided in Anderson, Kellogg, and Salant (2016).

Kilian (2009) constructs historical decompositions based on this VAR model that establish that oil demand shocks rather than oil supply shocks explain most fluctuations in the real price of oil, not just in recent years, but as far back as in the late 1970s. Moreover, structural impulse response analysis reveals that no two oil price surges are alike. Depending on the composition of the underlying oil demand and oil supply shocks, the persistence, shape, and magnitude of the response of the real price of oil may differ substantially. This result has important implications for studies of the transmission of oil price shocks.

8.4.3 Oil Price Shocks and Stock Returns

Kilian and Park (2009) use a generalization of the model in Kilian (2009) to study the dynamic response of U.S. real stock prices to oil supply shocks, aggregate demand shocks, and oil-specific demand shocks in the global market for crude oil:

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \\ u_t^{ret} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} & 0 \\ b_0^{41} & b_0^{42} & b_0^{43} & b_0^{44} \end{bmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \\ w_t^{\text{other}} \end{pmatrix},$$

where ret_t denotes real U.S. stock returns and u_t^{ret} is the reduced-form error of the corresponding VAR equation. In this model, U.S. real stock returns may respond instantaneously to any of the shocks that determine the real price of oil, as one would expect of an asset price. The residual shock w_t^{other} collectively captures all shocks that affect U.S. real stock returns without affecting the real price of oil within the same month.

The key additional identifying assumption is that the real global price of oil does not respond within the current month to changes in U.S. real stock returns that are not already explained by the first three shocks. This assumption allows for expectations of a booming world economy to drive up both the global real price of oil and U.S. real stock returns. It does not allow for an unrelated U.S. stock market correction to affect the real price of oil within the same month. This identifying assumption is implied by the conventional assumption that the

³ Further discussion of this point can be found in Chapters 9 and 12.

real price of oil is predetermined with respect to U.S. macroeconomic aggregates because

$$u_t^{rpoil} = b_0^{31} w_t^{\text{oil supply}} + b_0^{32} w_t^{\text{aggregate demand}} + b_0^{33} w_t^{\text{oil-specific demand}}.$$

By cumulating the return responses, one can infer the response of U.S. real stock prices to each oil demand and oil supply shock. Kilian and Park illustrate that the dynamic correlation between oil returns and stock returns evolves with changes in the composition of oil demand and oil supply shocks and may be positive, negative, or zero. This finding explains the seeming instability in the reduced-form relationship between oil prices and stock prices documented in the finance literature.

8.4.4 Models of the Transmission of Energy Price Shocks

Recursively identified VAR models such as the model in Kilian (2009) are fully identified in that each structural shock is uniquely identified. Often we do not have enough restrictions to fully identify a VAR model. This has prompted the development of semistructural or partially identified VAR models. The idea of semistructural models is that in some cases we may be satisfied if we can identify a subset of the structural shocks. Often we are interested in one structural shock only. An example is models of the transmission of energy price shocks in which the price of energy is predetermined with respect to all domestic macroeconomic aggregates, consistent with the empirical evidence provided in Kilian and Vega (2011). For example, Edelstein and Kilian (2009) utilized a recursively identified monthly bivariate model similar to the model:

$$\begin{pmatrix} u_t^{\Delta p} \\ u_t^{\Delta c} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 \\ b_0^{21} & b_0^{22} \end{bmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix},$$

where Δp_t denotes the percent change in U.S. energy prices (possibly weighted by the energy share in expenditures) and Δc_t denotes the percent growth in real U.S. consumption. The model is semistructural in that only the innovation in the price of energy, w_{1t} , is economically identified. The w_{2t} term, in contrast, is a conglomerate of other structural shocks that are not individually identified. Hence, w_{2t} does not have an economic interpretation. Put differently, both elements of w_t are statistically identified, but only w_{1t} is economically identified. The identification scheme postulates that Δp_t is predetermined with respect to Δc_t .

It is important to keep in mind that responses to energy price innovations from models such as this one are designed to tell us about the average response to an energy price shock over the sample. The actual response at a given point in time may be different from the average response, depending on the determinants of the energy price shock in question. Nevertheless, models of this type

have been used extensively in the literature, given the challenges of identifying energy demand and supply shocks (see Rotemberg and Woodford 1996).

8.4.5 Semistructural Models of Monetary Policy

A large literature is devoted to identifying monetary policy shocks using structural VAR models that are only partially identified. Under suitable conditions such models may be estimated using standard techniques for recursively identified VAR models with the monetary policy instrument ordered last among the VAR variables. This idea was pioneered by Bernanke and Blinder (1992).

A Stylized Model of U.S. Monetary Policy. The simplest example is a quarterly model for $y_t = (\Delta gdp_t, \pi_t, i_t)'$, where Δgdp_t denotes U.S. real GDP growth, π_t the inflation rate, and i_t the federal funds rate. We use the lower-triangular Cholesky decomposition to compute

$$\begin{pmatrix} u_t^{\Delta gdp} \\ u_t^\pi \\ u_t^i \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} \end{bmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \end{pmatrix}.$$

The last equation of the model is interpreted as a linear monetary policy reaction function. The interest rate is the policy instrument. In setting u_t^i , the Federal Reserve responds endogenously to contemporaneous variation in Δgdp_t and π_t and to variation in the lagged observables. This part of the model captures the systematic response of the central bank to the economy. The residual left after accounting for all endogenous variation in the interest rate, w_{3t} , is interpreted as an exogenous monetary policy shock. This policy shock may reflect deviations from the expected (or average) policy response that may arise, for example, from changes in the composition of the Federal Open Market Committee (FOMC), from shifts in the weights that the FOMC attaches to different objectives, or from discretionary policy decisions in response to extraordinary events. It may also arise from self-fulfilling shocks to market expectations about Fed policy, or from errors the Federal Reserve makes in forecasting the extent to which Treasury operations will add or drain reserves available to private banks (see, e.g., Chari, Christiano, and Eichenbaum 1998; Hamilton 1997). The policy shock, w_{3t} , is the only structural shock of interest in this model. No attempt is made to identify economically the structural shocks w_{1t} and w_{2t} .

It is easy to see that alternative orderings of $u_t^{\Delta gdp}$ and u_t^π will leave w_{3t} unaffected because the last column of B_0^{-1} is unaffected by the elements in the first two columns of this matrix. This result even allows for a nonrecusive relationship between $u_t^{\Delta gdp}$ and u_t^π . A related result about the invariance of the monetary policy shock to alternative block-triangular orderings of B_0 can be found in Christiano, Eichenbaum, and Evans (1999).

Models of this type have been commonly used in empirical work. The policy variable in semistructural VAR models need not be the short-term interest rate. A similar approach to identification may be followed with alternative policy indicators such as nonborrowed reserves (see, e.g., Strongin 1995). Regardless of the details of the specification, this identification scheme requires that the shock of interest be ordered at (or near) the bottom of the recursive ordering, given that the central bank tends to respond to most (if not all) of the other model variables.

Semistructural VAR models of monetary policy have five important weaknesses that are apparent even in the trivariate setting. First, this model does not allow for feedback within a given quarter from w_{3t} to Δgdp_t and π_t . This restriction seems implausible at least at quarterly frequency. Because Δgdp_t is not available at higher frequency, there is little we can do about this problem.⁴ It might seem that the same identification scheme would be more credible if we replaced Δgdp_t by the growth rate of industrial production and estimated the model at monthly frequency. This is not the case. One problem is that industrial output accounts for only a fraction of total output and that fraction is unstable over time. Moreover, real GDP is a measure of value added, whereas industrial output is a gross output measure. Finally, it is well known that the Federal Reserve is concerned with broader measures of real activity, making a policy reaction function based on industrial production growth economically less plausible and hence less interesting. In this regard, a better measure of monthly U.S. real activity would be the Chicago Fed's monthly principal components index of U.S. real activity (CFNAI). Yet another approach in the literature has been to interpolate quarterly real GDP data based on the fluctuations in monthly industrial production data and other monthly indicators. Such ad hoc methods not only suffer from the same deficiencies as the use of industrial production data, but they are also likely to distort the structural impulse responses to be estimated.

Second, the Federal Reserve may respond systematically to more variables than just Δgdp_t and π_t . To the extent that these variables are omitted from the model, we obtain inconsistent estimates of b_0^{31} and b_0^{32} , and incorrect measures of the monetary policy shock w_{3t} . In essence, the problem is that the policy shocks must be exogenous to allow us to learn about the effects of monetary policy shocks.

One potential solution to this problem is to enrich the set of variables ordered above the interest rate relative to the simple benchmark model and to estimate much larger VAR systems (see, e.g., Bernanke and Blinder 1992;

⁴ The Bureau of Economic Analysis does not release monthly U.S. real GDP data. Unofficial measures of monthly U.S. real GDP constructed similarly to the official quarterly data have recently been provided by Macroeconomic Advisers, LLC. These time series currently are not long enough for estimating VAR models of monetary policy, however.

Sims 1992; Christiano, Eichenbaum, and Evans 1999). Adding more variables, however, invites overfitting and undermines the credibility of the VAR estimates. Standard VAR models cannot handle more than about half a dozen variables, given typical sample sizes.

One potential remedy of this problem is to work with factor-augmented VAR (FAVAR) models, as in Bernanke and Boivin (2003), Bernanke, Boivin, and Eliasz (2005), Stock and Watson (2005), or Forni, Giannone, Lippi, and Reichlin (2009). Alternatively, one can work with large-scale Bayesian VAR models in which the cross-sectional dimension K is allowed to be larger than the time dimension T , as in Bańbura, Giannone, and Reichlin (2010). These large-scale models are designed to incorporate a much richer information structure than conventional semistructural VAR models of monetary policy. FAVAR models and large-scale BVAR models have three distinct advantages over conventional small to medium-sized VAR models: (1) They allow for the fact that central bankers form expectations about domestic real activity and inflation based on hundreds of economic and financial time series rather than a handful of time series; (2) they allow for the fact that economic concepts such as domestic economic activity and inflation may not be well represented by a single observable time series; (3) they allow the user to construct the responses of many variables not included in conventional VAR models. There is evidence that allowing for richer information sets in specifying VAR models may improve the plausibility of the estimated responses. A detailed review of structural FAVAR models and large-scale structural BVAR models can be found in Chapter 16.

An alternative strategy is to add to the model selectively one or more variables that carry information about future economic activity and inflation and hence help address the informational deficiencies of low-dimensional VAR models (see Sims and Zha 2006a). We already noted in Chapter 7 that asset prices in particular quickly respond to new information about future economic conditions. Examples are housing prices, stock prices, and industrial commodity prices. Of particular interest in this context is the inclusion of an index of commodity prices.

A common problem in monetary policy VAR models has been the presence of a so-called price puzzle, which refers to the finding of an increase in the price level in response to an unanticipated monetary tightening. This finding suggests that the VAR model is unable to capture expected inflation. Sims (1992) suggested that this puzzle can be resolved by including global commodity prices as an indicator of expected inflation in the model.⁵

⁵ Sims' premise was that the Federal Reserve considers global commodity prices as a predictor of expected inflation. Subsequently, Hanson (2004) showed that there is no systematic relationship between the ability of alternative measures of global commodity prices to predict U.S. inflation and their ability to resolve the price puzzle, casting doubt on this premise.

Following Sims, it has been customary to order commodity prices above the interest rate in the monetary policy VAR model, which allows the central bank to respond to movements in commodity prices. This assumption, however, prevents commodity prices from responding within the current period to policy changes in the interest rate. The latter restriction does not seem economically plausible. Ordering the commodity price index last addresses the latter concern, but is inconsistent with Sims' premise that the central bank responds contemporaneously to commodity price fluctuations. Thus, neither specification is economically appealing. Indeed, subsequent research has shown that the price puzzle more often than not persists even after including global commodity prices in the VAR model, suggesting that the model remains misspecified.

Third, the identification of the VAR model hinges on the monetary policy reaction function being stable over time. To the extent that policymakers have at times drastically changed the weights attached to their inflation and output objectives or have even changed the policy instrument, it becomes essential to split the sample in estimating the VAR model. The resulting shorter sample in turn makes it more difficult to include many variables in the model due to the lack of degrees of freedom. It also complicates statistical inference.

Fourth, the VAR model is linear. It does not allow for a lower bound on the interest rate, for example, making this model unsuitable for studying the quantitative easing of the Federal Reserve Board in recent years. Lower bounds on the interest rate imply a nonlinearity in the model. Nonlinear VAR models are discussed in Chapter 18.

Fifth, most VAR models of monetary policy ignore the real-time nature of the policy decision problem. Not all data relevant to policymakers are available without delay, and when data become available, they tend to be preliminary and subject to further revisions. To the extent that monetary policy shocks are defined as the residual of the policy reaction function, a misspecification of the policymaker's information set will cause biases in the estimated policy shocks. Bernanke and Boivin (2003) is an example of a study that explores the role of real-time data limitations in semistructural VAR models. Their conclusion is that – at least for their sample period and model – the distinction between real-time data and ex-post revised data is of limited importance. A similar conclusion is reached by Croushore and Evans (2006).

Finally, it is useful to reiterate that the thought experiment contemplated in structural VAR models is an unanticipated monetary policy shock within an existing monetary policy rule (also referred to as a policy regime at times). This exercise is distinct from that of changing the monetary policy rule (as happened in 1979 under Paul Volcker or in 2008 following the quantitative easing of the Federal Reserve Board). The question of what the effects of a shift in the monetary policy rule are on the economy is of independent interest, but much harder to answer. The role of systematic monetary policy has been stressed in Leeper, Sims, and Zha (1996), Bernanke, Gertler, and Watson

(1997), Hamilton and Herrera (2004), and Kilian and Lewis (2011), for example. The econometric evaluation of the role of systematic monetary policy in linear VAR models, however, remains controversial and may easily run afoul of the Lucas critique, as discussed in Chapter 4. One potential alternative is the use of nonlinear structural VAR models that allow for time variation, as discussed in Chapter 18.

Three Benchmark Models of U.S. Monetary Policy. Recursively identified models of monetary policy in practice tend to include many more variables than the stylized three-variable model we considered so far. For example, Christiano, Eichenbaum, and Evans (1999) propose three benchmark specifications for semistructural VAR models of monetary policy, which differ depending on whether we view the federal funds rate (i_t), nonborrowed reserves (nbr_t) or the ratio of nonborrowed to total reserves (nbr_t/tr_t) as the policy instrument. The models are quarterly and involve seven variables each. The first specification is

$$y_t = \begin{pmatrix} y_{1t} \\ i_t \\ y_{2t} \end{pmatrix},$$

where y_{1t} is a vector of macroeconomic aggregates that the Federal Reserve responds to contemporaneously in setting the policy instrument, and y_{2t} is a vector of macroeconomic aggregates that the Federal Reserve does not respond to within the current quarter. Whereas y_{1t} is assumed not to respond to policy surprises within the first quarter, y_{2t} is allowed to respond within the first quarter. Christiano et al. define $y_{1t} = (gdp_t, p_t, pcom_t)'$, where gdp_t denotes U.S. real GDP, p_t the GDP deflator, and $pcom_t$ a commodity price index, and $y_{2t} = (tr_t, nbr_t, m_t)'$, where tr_t and nbr_t denote total reserves and nonborrowed reserves, respectively, and m_t refers to a suitably defined monetary aggregate. All variables but the interest rate are expressed in logs.

The second specification is

$$y_t = \begin{pmatrix} y_{1t} \\ nbr_t \\ y_{2t} \end{pmatrix},$$

where the policy instrument nbr_t is nonborrowed reserves, $y_{1t} = (gdp_t, p_t, pcom_t)'$, and $y_{2t} = (i_t, tr_t, m_t)'$. The third specification is

$$y_t = \begin{pmatrix} y_{1t} \\ nbr_t \\ y_{2t} \end{pmatrix},$$

where again the policy instrument nbr_t is nonborrowed reserves, $y_{1t} = (gdp_t, p_t, pcom_t, tr_t)'$, and $y_{2t} = (i_t, m_t)'$. Christiano et al. refer to the resulting policy

shock as an nbr_t/tr_t shock, building on Strongin (1995) who proposed the ratio of nonborrowed reserves to total reserves as an alternative monetary policy instrument.

There is a large literature on the theoretical merits of each of these specifications and on how to model the relationship between the federal funds rate, nonborrowed reserves, and total reserves. A review of this literature is provided in Christiano, Eichenbaum, and Evans (1999), who also demonstrate that, in all three specifications, the federal funds rate increases, monetary aggregates decline over time, the price level initially is sluggish, real output temporarily declines, and commodity prices fall in response to an unanticipated monetary contraction.

A Model of the Macroeconomic Effects of Monetary Policy. More recent models in this literature have tended to ignore data on nonborrowed reserves and total reserves and have largely focused on the federal funds rate as the policy instrument. For example, Christiano, Eichenbaum, and Evans (2005) analyze a nine-variable version of the recursive VAR model of monetary policy. The data are quarterly. As before,

$$y_t = \begin{pmatrix} y_{1t} \\ i_t \\ y_{2t} \end{pmatrix},$$

where i_t is the federal funds rate. Christiano et al. define

$$y_{1t} = (gdp_t, c_t, p_t, inv_t, wage_t, prod_t)',$$

where gdp_t denotes U.S. real GDP, p_t the GDP deflator, c_t real consumption, inv_t real investment, $wage_t$ the real wage, and $prod_t$ labor productivity. They define $y_{2t} = (\Delta m_t, rp_t)'$, where Δm_t refers to M2 growth and rp_t stands for real profits. All variables but the interest rate and Δm_t are expressed in logs. The monetary policy shock is identified by imposing a recursive structure on B_0^{-1} (and hence on B_0) such that the seventh row of B_0 may be interpreted as a policy reaction function conditional on past data of all model variables. None of the other structural shocks are identified from an economic point of view.

The Effect of Monetary Policy Shocks on the Exchange Rate. Eichenbaum and Evans (2005) extend the closed-economy semistructural VAR framework to include the bilateral U.S. dollar exchange rate with respect to selected other countries. The data are monthly. Let

$$y_t = \begin{pmatrix} y_{1t} \\ i_t^{\text{US}} \\ y_{2t} \end{pmatrix},$$

where i_t^{US} is the U.S. federal funds rate. Eichenbaum and Evans define

$$y_{1t} = (ip_t^{\text{US}}, p_t^{\text{US}})',$$

where ip_t^{US} is the log of U.S. industrial production and p_t^{US} denotes the log of the U.S. consumer price index, and

$$y_{2t} = \begin{pmatrix} i_t^{\text{Foreign}} - i_t^{\text{US}} \\ er_t^{\text{US}/\text{Foreign}} \end{pmatrix},$$

where $i_t^{\text{Foreign}} - i_t^{\text{US}}$ is the short-term nominal interest rate differential and $er_t^{\text{US}/\text{Foreign}}$ denotes, alternatively, the U.S. nominal exchange rate or the U.S. real exchange rate. The foreign countries (considered one at a time) include Japan, Germany, Italy, France, and the United Kingdom. One identifying assumption is that the Federal Reserve responds to the current level of U.S. real output and U.S. prices but disregards the current foreign interest rate and the current exchange rate in setting policy. The other identifying assumption is that U.S. real output and the U.S. price level do not respond within the month to monetary policy disturbances. The model is again partially identified in that only the monetary policy shock is economically identified.

The study focuses on two questions. One is the exchange rate responses to a monetary policy shock. Eichenbaum and Evans document a persistent appreciation of nominal and real exchange rates in response to a contractionary monetary policy shock. The other question is whether uncovered interest rate parity applies to the responses to monetary policy shocks. Based on the estimated model, Eichenbaum and Evans compute the implied excess returns defined as the ex post difference in the return between investing \$1 in one-period foreign bonds and investing \$1 in one-period domestic bonds, expressed in U.S. dollars. Under uncovered interest rate parity, one would not expect a monetary policy shock to be associated with excess returns. Eichenbaum and Evans find persistent and statistically significant departures from uncovered interest rate parity.

8.4.6 The Permanent Income Model of Consumption

Cochrane (1994) proposes another application of the recursive model. His interest is not in identifying demand or supply shocks, but in decomposing permanent and transitory shocks within the framework of the permanent income model of consumption. The standard permanent income model implies that log real consumption (c_t) and log real income (gnp_t) are cointegrated such that the consumption-income ratio is stationary. Cochrane imposes this cointegration restriction on the reduced-form VAR model for (c_t, gnp_t) . The permanent income model also predicts that if income changes unexpectedly without a corresponding change in consumption, then consumers will regard the shock

to income as having purely transitory effects on income. Cochrane identifies such a shock by recursively ordering innovations to consumption first in the Cholesky decomposition of the reduced-form error covariance matrix. This decomposition allows him to separate permanent from transitory shocks and to quantify their importance for the variability of consumption and income:

$$\begin{pmatrix} u_t^c \\ u_t^{gnp} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 \\ b_0^{21} & b_0^{22} \end{bmatrix} \begin{pmatrix} w_t^{\text{permanent}} \\ w_t^{\text{transitory}} \end{pmatrix}.$$

Note that, by construction, consumption only depends on the permanent shock, whereas income in addition depends on the transitory shock.⁶ Cochrane verifies that the response of income to the transitory shock is indeed rapidly mean-reverting, whereas the response of income to a shock that moves both consumption and income on impact has long-lasting effects on income, as expected from a permanent shock. Moreover, much of the consumption response to a permanent shock is immediate, whereas the response of consumption to a transitory shock is close to zero at all horizons.⁷

Unlike in our earlier examples, this methodology is silent about the economic interpretation of permanent and transitory shocks. It is common in the literature, however, to equate the permanent shock in this model with a supply or productivity shock and the transitory shock with a demand shock. In general, the transitory and permanent shocks will be a mixture of these deeper economic shocks.

8.5 Examples of Nonrecursively Identified Models

Not all structural VAR models have a recursive structure. Increasing skepticism toward atheoretical recursively identified models in the mid-1980s stimulated a series of studies proposing explicitly structural models identified by nonrecursive short-run restrictions (see, e.g., Bernanke 1986; Sims 1986; Blanchard and Watson 1986). As in the recursive model, the identifying restrictions on B_0 or B_0^{-1} generate moment conditions that can be used to estimate the unknown coefficients in B_0 . In general, solving the moment conditions for the unknown structural parameters will require iteration, but in some cases the GMM estimator can be constructed using traditional instrumental-variable techniques. An alternative commonly used approach is to model the error distribution as Gaussian and to estimate the structural model by full information maximum

⁶ The terminology of transitory shocks and permanent shocks is somewhat misleading in that any shock by construction involves a one-time disturbance only. A transitory shock, more precisely, is defined as a shock with purely transitory effects on the observables, whereas a permanent shock refers to a shock with permanent (or long-run) effects on the observables.

⁷ It can be shown that the results of Cochrane's model would be identical to the results from a model in which the transitory shock has no long-run effect on the level of income and consumption. Such long-run restrictions are discussed in Chapter 10.

likelihood methods. This approach involves the maximization of the concentrated likelihood with respect to the structural model parameters subject to the identifying restrictions. A detailed discussion of these estimation issues is provided in Chapter 9. Our focus in this section is on providing examples of identifying assumptions used in this literature.

8.5.1 Fiscal Policy Shocks

Blanchard and Perotti (2002) introduce a model of U.S. fiscal policy that deviates from the usual recursive structure. They propose a quarterly model of the U.S. economy for $y_t = (tax_t, gov_t, gdp_t)$, where tax_t refers to real taxes, gov_t to real government spending, and gdp_t to real GDP. All variables are in logs. Ignoring lags, the structural relations can be written as

$$\begin{bmatrix} 1 & 0 & b_{13,0} \\ 0 & 1 & b_{23,0} \\ b_{31,0} & b_{32,0} & 1 \end{bmatrix} \begin{pmatrix} u_t^{tax} \\ u_t^{gov} \\ u_t^{gdp} \end{pmatrix} = \begin{bmatrix} 1 & c_{12} & 0 \\ c_{21} & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{pmatrix} w_t^{tax} \\ w_t^{gov} \\ w_t^{gdp} \end{pmatrix}$$

such that

$$\begin{aligned} u_t^{tax} &= -b_{13,0}u_t^{gdp} + c_{12}w_t^{gov} + w_t^{tax}, \\ u_t^{gov} &= -b_{23,0}u_t^{gdp} + c_{21}w_t^{tax} + w_t^{gov}, \\ u_t^{gdp} &= -b_{31,0}u_t^{tax} - b_{32,0}u_t^{gov} + w_t^{gdp}. \end{aligned}$$

Blanchard and Perotti first provide institutional arguments for the delay restriction $b_{23,0} = 0$ which rules out automatic feedback from economic activity to government spending within the quarter. They then show that the within-quarter response of taxes to economic activity, $-b_{13,0}$, can be derived on the basis of extraneous tax elasticity estimates and is $b_{13,0} = -2.08$. The parameters $b_{31,0}$ and $b_{32,0}$ are left unrestricted. The potential endogeneity between taxes and spending is dealt with by imposing either $c_{21} = 0$ or $c_{12} = 0$. In the latter case, for example, we obtain

$$\begin{aligned} u_t^{tax} &= 2.08u_t^{gdp} + w_t^{tax}, \\ u_t^{gov} &= c_{21}w_t^{tax} + w_t^{gov}, \\ u_t^{gdp} &= -b_{31,0}u_t^{tax} - b_{32,0}u_t^{gov} + w_t^{gdp}. \end{aligned}$$

This system can be estimated numerically, as discussed in Chapter 9. Note that Blanchard and Perotti effectively treat the first two innovations as mutually exogenous without imposing the overidentifying restriction on c_{21} . An obvious concern is that the model does not allow for the anticipation of fiscal shocks. Blanchard and Perotti discuss how this concern may be addressed by changing the timing assumptions and adding further identifying restrictions, if we are willing to postulate a specific form of foresight. Another concern is that the

model does not condition on the debt structure (see, e.g., Chung and Leeper 2007). Allowing for the debt structure to matter would result in a nonlinear dynamic model.

8.5.2 An Alternative Simple Macroeconomic Model

Keating (1992) discusses a variation of the simple macroeconomic model we discussed earlier that does not impose a recursive structure and involves a different economic interpretation. His structural equations are

$$\begin{aligned} u_t^p &= w_t^{AS}, \\ u_t^{gdp} &= -b_{21,0}u_t^p - b_{23,0}u_t^i - b_{24,0}u_t^m + w_t^{IS}, \\ u_t^i &= -b_{34,0}u_t^m + w_t^{MS}, \\ u_t^m &= -b_{42,0}(u_t^{gdp} + u_t^p) - b_{43,0}u_t^i + w_t^{MD}. \end{aligned}$$

The first equation again represents a horizontal AS curve, but the second equation now can be interpreted as an IS curve, allowing real output to respond to all other model variables. The third equation represents a simple money supply function, according to which the central bank adjusts the rate of interest in relation to the money stock, and the fourth equation is a money demand function in which short-run money holdings rise in proportion to nominal income, yielding the final restriction required for exact identification. Unlike in the earlier example, money holdings are allowed to depend on the interest rate as well. Clearly, this model specification embodies a very different view of what monetary policymakers do than more recently developed structural VAR models motivated by the literature on Taylor rules (see Taylor 1993).

8.5.3 Discussion

Nonrecursively identified VAR models more closely resemble traditional simultaneous equations models. This means that they also are susceptible to the weaknesses of such models, including the difficulty of finding strong instruments for identifying causal effects in the data (see Chapter 9). A case in point is the literature on the liquidity effect. The liquidity effect refers to the short-run negative response of interest rates to an unanticipated monetary expansion. Although the presence of such an effect has been suspected for a long time, it has only been in the 1990s that structural VAR studies emerged concluding that there is a liquidity effect. Whereas the evidence of a liquidity effect is at best mixed in recursively identified models of monetary policy, empirical VAR studies based on nonrecursive simultaneous equations systems have reliably produced a strong liquidity effect. This evidence might seem to suggest that more explicitly structural models are inherently superior to earlier semistructural models of monetary policy. However, Pagan and Robertson (1998) show

that the instruments underlying the three most important nonrecursive studies of the liquidity effect appear weak in the econometric sense, calling into question any inferences made about the magnitude of the liquidity effect. We defer the discussion of alternative estimation methods for recursively and nonrecursively identified structural VAR models to Chapter 9.

8.5.4 The Graph-Theoretic Approach

Another strand of the literature uses graph-theoretic tools and terminology for representing causal relations between variables (see Pearl 2000; Spirtes, Glymour, and Scheines 2000; Eichler 2012). The idea is to represent causal relations between the variables of interest by directed graphs and to use statistical methods to narrow down the admissible causal structures that are compatible with the data. As a result, this approach has become known as the graph-theoretic approach to identification. The graph-theoretic approach was first used for structural VAR analysis by Swanson and Granger (1997) and was later applied, extended, and refined by Bessler and Lee (2002), Demiralp and Hoover (2003), Hoover (2005), Demiralp, Hoover, and Perez (2008), Hoover, Demiralp, and Perez (2009), Moneta (2008), Moneta, Entner, Hoyer, and Coad (2013), and Heinlein and Krolzig (2013), among others.

The graph-theoretic approach explores the partial correlations between the reduced-form VAR model errors. For example, in a three-dimensional Gaussian model, if the partial correlation between u_{1t} and u_{3t} , given u_{2t} , is zero, u_{1t} and u_{3t} are independent conditionally on u_{2t} . Hence, the coefficient of u_{3t} in a regression

$$u_{1t} = -b_{12,0}u_{2t} - b_{13,0}u_{3t} + w_{1t}$$

is zero. In practice, if the null hypothesis of a zero partial correlation between u_{1t} and u_{3t} cannot be rejected, an exclusion restriction on the corresponding element of B_0 is imposed, and we set $b_{13,0} = 0$.

Similarly, the result of a partial correlation analysis in a four-dimensional VAR system may be

$$\begin{bmatrix} * & 0 & 0 & 0 \\ * & * & 0 & 0 \\ * & 0 & * & 0 \\ 0 & * & * & * \end{bmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{pmatrix} = \begin{pmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \\ w_{4t} \end{pmatrix},$$

where asterisks denote unrestricted elements. Such a result would suggest that the DGP is compatible with a recursive structure and that there are in fact over-identifying zero restrictions. This information is obviously helpful in identifying mutually uncorrelated shocks.

Of course, there is no guarantee that a fully identified B_0 matrix is obtained using this approach. For example, an outcome such as

$$\begin{bmatrix} * & * & 0 & 0 \\ * & * & 0 & 0 \\ 0 & 0 & * & * \\ 0 & 0 & * & * \end{bmatrix} \begin{pmatrix} u_{1t} \\ u_{2t} \\ u_{3t} \\ u_{4t} \end{pmatrix} = \begin{pmatrix} w_{1t} \\ w_{2t} \\ w_{3t} \\ w_{4t} \end{pmatrix},$$

where again * denotes elements that may be nonzero, is possible as well. In the latter case, the B_0 matrix is not fully identified, although there are more zero elements than required by the order condition. The problem is that the upper left-hand and lower right-hand submatrices are not identified. Full identification requires additional identifying restrictions from other sources.

Although this data-based approach to identification has some appeal, it also has some obvious drawbacks. First, the fact that the data admit a set of zero elements for B_0 does not mean that the B_0 matrix obtained by the graph-theoretic approach reflects the actual economic structure, as we already illustrated by example in Section 8.2. Put differently, even a finding of uncorrelated reduced-form residuals with partial correlations of zero does not rule out that the variables are related by an instantaneous economic causal structure. In fact, it is not clear how to interpret statements about one variable causing another from an economic point of view, because structural shocks in simultaneous equations models in general are not associated with specific observables.

Second, in practice, the graph-theoretic approach relies on statistical tests to determine whether the partial correlations in the data are zero or not. An obvious concern is that a failure to reject the null hypothesis does not necessarily imply the validity of the null model. Especially in small samples, statistical tests are likely not to reject an invalid null hypothesis of zero partial correlation simply due to low power. Moreover, the full set of zero restrictions is the result of a multiple testing procedure with unknown overall size.

In short, this data-driven identification approach is not well-suited for uncovering economically meaningful structures. This fact helps explain why, as of now, there are not many structural VAR applications based on this approach.

8.6 Summary

The central question in structural VAR analysis is how to recover the unknown parameters of the structural model from the parameters of the reduced-form representation of this model. Knowledge of these structural parameters allows us to characterize the responses of the model variables to the structural shocks and related statistics of interest (see Chapter 4). The problem of solving for the unknown structural parameters as functions of the parameters of the reduced

form is known as the identification problem. Solving this identification problem requires the user to impose additional identifying restrictions. As long as we impose enough identifying restrictions, all parameters of the structural model may be uniquely solved for. The choice of these restrictions determines the numerical values of the structural model parameters. It is essential that these identifying restrictions be economically meaningful. Imposing ad hoc restrictions without economic justification will fail to identify correctly the underlying structural model parameters. The challenge for applied users thus is to come up with identifying restrictions that are economically plausible.

In this chapter, we examined the conditions required for the unique identification of the parameters of the structural VAR model from the reduced-form representation. We discussed the role of normalizing assumptions for identification and the challenges in coming up with economically credible identifying restrictions. There are many different types of identifying restrictions, but the most common approach in applied work has been to impose exclusion restrictions on selected elements of the structural impact multiplier matrix (or its inverse). Because such restrictions only affect the contemporaneous interaction among the model variables, conditional on their past values, they are also known as short-run identifying restrictions.

We illustrated the use of short-run identifying restrictions by example. One common assumption has been that there is a causal ordering among the model variables. Such recursively identified structural models are usually difficult to motivate from an economic point of view. Many applications of this approach have been rightly criticized as atheoretical in that the implied shocks are mutually uncorrelated but devoid of economic content and hence not truly structural. If we are willing to restrict attention to the identification of only one structural shock, recursive models may sometimes be given a semistructural interpretation. The latter class of models has played an important role in monetary economics, for example, but is not without serious limitations. A less common alternative approach has been to allow the structural model to be non-recursive, conditional on past data. Such restrictions can sometimes be derived from an explicit economic model structure.

In response to the scarcity of credible short-run exclusion restrictions in applied work, a large number of alternative (or sometimes complementary) types of restrictions has been developed in the structural VAR literature since the 1990s. Given the widespread use of short-run identifying restrictions in applied work, it is useful to examine in detail the estimation of models identified by short-run restrictions in Chapter 9, before discussing these alternative approaches.

9 Estimation Subject to Short-Run Restrictions

This chapter illustrates the estimation of structural VAR models subject to short-run identifying restrictions. A variety of estimation methods have been proposed to estimate these models, including the method of moments, instrumental variable (IV) estimation, full information maximum likelihood (FIML) estimation, and Bayesian estimation.

9.1 Model Setup

We consider the K -dimensional structural-form VAR(p) model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t = B Y_{t-1} + w_t, \quad (9.1.1)$$

where $Y'_{t-1} \equiv (y'_{t-1}, \dots, y'_{t-p})$, $B \equiv [B_1, \dots, B_p]$, and $w_t \sim (0, \Sigma_w)$. All deterministic terms are neglected because they are of no importance for the following discussion. It is straightforward to add these terms, as needed, in the derivations of this chapter. The structural errors have a diagonal covariance matrix Σ_w and are serially uncorrelated.

The corresponding reduced form is

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t = A Y_{t-1} + u_t, \quad (9.1.2)$$

where $A \equiv [A_1, \dots, A_p] = B_0^{-1} B$ and $u_t = B_0^{-1} w_t \sim (0, \Sigma_u)$ is a white noise error term with positive definite covariance matrix $\Sigma_u = B_0^{-1} \Sigma_w B_0^{-1'}$.

For the purpose of this chapter it is sometimes useful to assume in addition that $\Sigma_w = I_K$. Alternatively, it is sometimes assumed that the diagonal elements of B_0 are standardized to one and that the covariance matrix of w_t is a diagonal matrix such that

$$B_0 = \begin{bmatrix} 1 & b_{12,0} & \cdots & b_{1K,0} \\ b_{21,0} & 1 & \ddots & \vdots \\ \vdots & \ddots & \ddots & b_{K-1,K,0} \\ b_{K1,0} & \cdots & b_{K,K-1,0} & 1 \end{bmatrix}$$

and

$$\Sigma_w = \begin{bmatrix} \sigma_{w_1}^2 & & 0 \\ & \ddots & \vdots \\ 0 & \dots & \sigma_{w_K}^2 \end{bmatrix}.$$

Which assumption is preferred depends on the nature of the identifying assumptions, which in turn restricts the choice of the estimation method. As we illustrate in this chapter, some estimation methods are designed for models with identifying restrictions on B_0 , whereas others are designed for estimating models with identifying restrictions on B_0^{-1} . There also are estimation methods that allow for combinations of restrictions on B_0 and B_0^{-1} .

9.2 Method-of-Moments Estimation

Many structural VAR models in the empirical literature are just identified. For such models there are a number of two-step estimation methods that may be viewed as alternative implementations of the method of moments. The method-of-moments approach to estimating B_0^{-1} is based on the second-moment matrix of the VAR innovations, Σ_u , which may be expressed in terms of the structural model parameters as

$$\Sigma_u = B_0^{-1} \Sigma_w B_0^{-1'}, \quad (9.2.1)$$

where Σ_w is diagonal by assumption. Because there are $K(K + 1)/2$ unique elements in Σ_u , expression (9.2.1) may be viewed as defining $K(K + 1)/2$ moment conditions involving the unknown parameters of B_0^{-1} and Σ_w . After replacing Σ_u by a consistent estimate $\widehat{\Sigma}_u$ and imposing suitable normalizing and identifying restrictions on the elements of Σ_w and B_0^{-1} (or alternatively on B_0), one can uniquely solve for the remaining unknown structural parameters.

In Section 9.2.1 we focus on recursively identified structural models with a lower-triangular structure for B_0^{-1} , followed by structural models identified based on nonrecursive short-run identifying restrictions in Section 9.2.2. Extensions to overidentified models, which may be estimated by the generalized method of moments (GMM), are discussed in Section 9.2.3.

9.2.1 Recursively Identified Models

The simplest approach to estimating a recursively identified model employs the Cholesky decomposition of $\widehat{\Sigma}_u$. Consider the following example. A popular argument in macroeconomics has been that oil price shocks in particular may act as domestic supply shocks for the U.S. economy. Thus, the question of how oil price shocks affect U.S. real GDP and inflation has a long tradition in

macroeconomics (see, e.g., Blanchard 2002; Barsky and Kilian 2004; Kilian 2008c). We address this question by postulating a VAR(4) model with intercept for the percent changes in the real WTI price of crude oil ($\Delta rpoil_t$), the U.S. GDP deflator inflation rate (Δp_t), and U.S. real GDP growth (Δgdp_t). The data are quarterly and the estimation period is 1987q1-2013q2. The model is identified recursively with the real price of oil ordered first such that the real price of oil is predetermined with respect to the U.S. economy. We focus on the effect of an unanticipated increase in the real price of oil on the real price of oil, on the change in inflation, and on U.S. real GDP growth. The model is partially identified in that only the oil price shock can be given an economic interpretation.

Let $y_t = (\Delta rpoil_t, \Delta p_t, \Delta gdp_t)'$. The LS estimates of the reduced-form slope parameters are

$$\begin{aligned}\hat{A}_1 &= \begin{bmatrix} -0.0064 & 0.9365 & 1.7510 \\ 0.0018 & 0.5991 & 0.0282 \\ -0.0024 & -0.2187 & 0.3324 \end{bmatrix}, \\ \hat{A}_2 &= \begin{bmatrix} -0.1822 & 12.6454 & -3.4391 \\ 0.0043 & 0.1276 & -0.0388 \\ -0.0063 & 0.3877 & 0.1459 \end{bmatrix}, \\ \hat{A}_3 &= \begin{bmatrix} -0.0073 & -3.5986 & 1.7337 \\ 0.0021 & 0.0477 & 0.0234 \\ -0.0002 & -0.1413 & -0.0439 \end{bmatrix}, \\ \hat{A}_4 &= \begin{bmatrix} -0.1194 & -8.1807 & 0.9309 \\ 0.0007 & 0.1488 & 0.0706 \\ -0.0081 & -0.0812 & 0.0116 \end{bmatrix},\end{aligned}$$

and the estimate of the reduced-form error covariance matrix is

$$\hat{\Sigma}_u = \begin{bmatrix} 312.5246 & 0.7736 & 0.9193 \\ 0.7736 & 0.0515 & 0.0149 \\ 0.9193 & 0.0149 & 0.5570 \end{bmatrix}.$$

Estimating the Model Using the Cholesky Decomposition. If B_0^{-1} is lower triangular, as in this empirical example, the easiest way of estimating the structural VAR model involves two steps. First, we estimate the reduced-form VAR parameters and compute the residual variance-covariance matrix $\hat{\Sigma}_u$. Then we estimate the structural impact multiplier matrix B_0^{-1} based on a lower-triangular Cholesky decomposition of the residual variance-covariance matrix $\hat{\Sigma}_u$. Given the recursive structure of B_0^{-1} and the normalizing assumption that $\Sigma_w = I_K$, the system of nonlinear equations, $\Sigma_u = B_0^{-1}B_0^{-1}'$, with Σ_u replaced by $\hat{\Sigma}_u$, implicitly defines the unknown elements of B_0^{-1} .

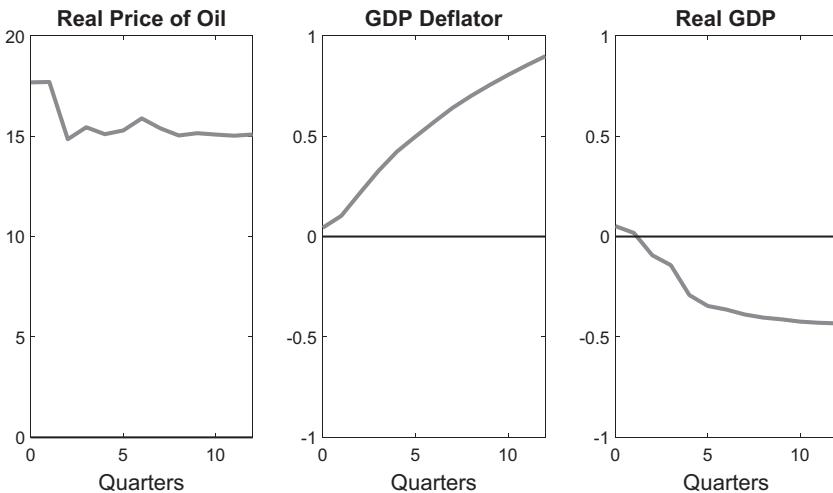


Figure 9.1. Responses of the U.S. economy to an unexpected increase in the real price of oil.

In the example above, this approach yields the estimate

$$\widehat{B}_0^{-1} = \text{chol}(\widehat{\Sigma}_u) = \begin{bmatrix} 17.6784 & 0 & 0 \\ 0.0438 & 0.2227 & 0 \\ 0.0520 & 0.0566 & 0.7424 \end{bmatrix},$$

where chol denotes a function that returns the lower-triangular Cholesky decomposition such that $\widehat{B}_0^{-1}\widehat{B}_0^{-1'} = \widehat{\Sigma}_u$. Such functions are available in commonly used software. Note that the first column of \widehat{B}_0^{-1} contains the responses of the three model variables in the impact period to an unexpected change in the real price of oil.

Given an estimate of B_0^{-1} and of the reduced-form parameters, it is straightforward to compute estimates of the structural impulse responses (see Chapter 4). Under the maintained assumption of predetermined oil prices, only the responses to the oil price innovation are economically identified. Figure 9.1 shows that an unexpected increase in the real price of crude oil is associated with an increase in the U.S. GDP deflator and a decline in U.S. real GDP.

For the construction of the structural impulse responses we require an estimate of B_0^{-1} rather than B_0 . If an estimate of B_0 is of interest, such an estimate may be generated by inverting \widehat{B}_0^{-1} . It should be noted that B_0^{-1} being lower triangular implies that B_0 is lower triangular (and vice versa). Hence the approach described above may also be used without loss of generality when estimating models that impose a recursive structure on B_0 .

Estimating the Model Using a Nonlinear Equation Solver. An alternative approach is to solve the system of nonlinear equations that implicitly defines the elements of $B_0^{-1} = [b_0^{ij}]$ using a nonlinear equation solver that finds the vector x such that $F(x) = 0$, where $F(x)$ denotes a system of nonlinear equations in x .¹ After normalizing $\Sigma_w = I_3$, while leaving the diagonal elements of B_0 unrestricted, we vectorize the system of equations

$$B_0^{-1}B_0^{-1'} - \widehat{\Sigma}_u = 0.$$

The objective is to find the unknown elements of B_0^{-1} such that

$$\begin{bmatrix} \text{vech}(B_0^{-1}B_0^{-1'} - \widehat{\Sigma}_u) \\ b_0^{12} \\ b_0^{13} \\ b_0^{23} \end{bmatrix} = 0, \quad (9.2.2)$$

where the vech operator is used to select the set of unique elements of $B_0^{-1}B_0^{-1'} - \widehat{\Sigma}_u$. The nonlinear equation solver iterates on expression (9.2.2) until convergence, given an initial guess for B_0^{-1} . It should be noted that the sign of the columns in \widehat{B}_0^{-1} is not unique and may flip, depending on how this numerical procedure is initialized. In practice, an additional normalization may be required to match the results from the Cholesky decomposition, which imposes by default that the diagonal elements of \widehat{B}_0^{-1} are positive (see also Taylor 2004).

For example, we may obtain the solution

$$\widehat{B}_0^{-1} = \begin{bmatrix} 17.6784 & 0 & 0 \\ 0.0438 & -0.2227 & 0 \\ 0.0520 & -0.0566 & 0.7424 \end{bmatrix},$$

which exactly matches that based on the Cholesky decomposition only after flipping the signs of the second column to ensure that all diagonal elements are positive. Indeed, in this simple example, there is no advantage to using a nonlinear equation solver, given the recursive structure of B_0^{-1} . The advantage of the nonlinear equation solver is that it can also accommodate nonrecursive structures, as discussed in Section 9.2.2.

Under the alternative normalization that Σ_w is diagonal with positive elements on the diagonal, while the diagonal of B_0 is restricted to a vector of ones, the structural impact multiplier matrix is $B_0^{-1}\Sigma_w^{1/2}$, where $\Sigma_w^{1/2}$ is obtained by

¹ An example of such a nonlinear equation solver is *fso* in the optimization toolbox of MATLAB.

taking the square root of the diagonal elements of Σ_w . We solve

$$\begin{bmatrix} \text{vech}(B_0^{-1}\Sigma_w B_0^{-1'} - \widehat{\Sigma}_w) \\ b_{11,0} - 1 \\ b_{22,0} - 1 \\ b_{33,0} - 1 \\ b_0^{12} \\ b_0^{13} \\ b_0^{23} \\ \sigma_{21,w} \\ \sigma_{31,w} \\ \sigma_{32,w} \end{bmatrix} = 0 \quad (9.2.3)$$

for the unknown elements of B_0 and Σ_w , where $b_{ij,0}$ and b_0^{ij} denote the ij^{th} elements of B_0 and B_0^{-1} , respectively, and $\sigma_{ij,w}$ denotes the ij^{th} element of Σ_w . Expression (9.2.3) may be solved by iterating until convergence using a nonlinear equation solver, given initial guesses for B_0 and for Σ_w . The estimate of the structural impact multiplier matrix is the same as when using the Cholesky decomposition. One advantage of this alternative normalization is that the signs of the implied structural impact multiplier matrix $\widehat{B}_0^{-1}\widehat{\Sigma}_w^{1/2}$ automatically match those in the lower triangular Cholesky decomposition.

If the identifying restrictions are imposed on the elements of B_0 rather than B_0^{-1} , it is simpler and more computationally attractive to solve

$$\begin{bmatrix} \text{vech}(B_0\widehat{\Sigma}_w B_0' - \Sigma_w) \\ b_{11,0} - 1 \\ b_{22,0} - 1 \\ b_{33,0} - 1 \\ b_{12,0} \\ b_{13,0} \\ b_{23,0} \\ \sigma_{21,w} \\ \sigma_{31,w} \\ \sigma_{32,w} \end{bmatrix} = 0. \quad (9.2.4)$$

Because B_0^{-1} being lower triangular implies that B_0 is lower triangular, the numerical solution will be identical to that from expression (9.2.3).

The computational efficiency may be increased by directly imposing the restrictions that the off-diagonal elements of Σ_w are zero and that the diagonal elements of B_0 are unity, and only solving for the remaining elements in (9.2.3) and (9.2.4). In this case we only require an initial guess for the off-diagonal elements of B_0 and for the diagonal elements of Σ_w .

Estimating the Model Using the Algorithm of Rubio-Ramirez et al. (2010). Rubio-Ramírez, Waggoner, and Zha (2010) propose an alternative algorithm

for VAR models with possibly nonrecursive structure that are exactly identified by short-run exclusion restrictions on the structural impulse responses. This algorithm avoids having to numerically solve a system of nonlinear equations by brute force.

Let L_0 denote an initial guess for the structural impact multiplier matrix B_0^{-1} such that $L_0 L_0' = \widehat{\Sigma}_u$. Possible choices include $L_0 = \widehat{\Sigma}_u^{1/2}$ or $L_0 = \text{chol}(\widehat{\Sigma}_u)$. Zero restrictions on B_0^{-1} can be represented by matrices Z_j for $1 \leq j \leq K$. The dimension of Z_j equals that of L_0 . The $K \times K$ matrix Z_j summarizes the exclusion restrictions on the impact effect of structural shock j on variable $i = 1, \dots, K$. The presence of an exclusion restriction is indicated by a 1 in Z_j and its absence by a zero. If there are no restrictions for a given structural shock j , $Z_j = 0_{K \times K}$. Exclusion restrictions on the impact effect of the i^{th} variable correspond to a 1 in the i^{th} column of a given row. There is at most one restriction per row. Restrictions are imposed starting with the first row of Z_j .

If shock j implies multiple restrictions, each of these exclusion restrictions is imposed in a different row, starting with the first variable to be restricted in the first row of Z_j , then the second variable to be restricted in the second row of Z_j , etc., in ascending order. A shock j that restricts the impact effect of the second and third variable, but not of the first variable, for example, would result in

$$Z_j = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}.$$

If the rank of Z_j is z_j , then z_j is the number of zero restrictions associated with the j^{th} shock. The total number of zero restrictions on B_0^{-1} is $z = \sum_{j=1}^K z_j$. The structural parameters satisfy the zero restrictions if and only if $Z_j L_0 e_j = 0$ for $1 \leq j \leq K$. Here e_j denotes the j^{th} column of the identity matrix I_K . In what follows, let Z_j represent the zero restrictions with the equations of the VAR model ordered such that $z_j \leq K - j$, where $1 \leq j \leq K$. Observe that $Z_j L_0 e_j = 0$ and $\bar{Z}_j L_0 e_j = 0$ are equivalent statements, provided \bar{Z}_j exists, where \bar{Z}_j is defined by deleting all rows of zeros in Z_j .

An Algorithm for Solving Exactly Identified Models Based on Short-Run Restrictions.

1. Let $j = 1$.
2. If $j = 1$, then $Q'_j = \bar{Z}_1 L_0$. For $j > 1$, form the matrix

$$Q'_j = \begin{bmatrix} \bar{Z}_j L_0 \\ q'_1 \\ \vdots \\ q'_{j-1} \end{bmatrix}.$$

3. There exists a vector q_j of unit length such that $Q'_j q_j = 0$. To find q_j such that $Q'_j q_j = 0$, use the QR decomposition $Q_j = \tilde{Q}R$, where \tilde{Q} is orthogonal and R is upper triangular. Choose q_j to be the last column of \tilde{Q} .
4. If $j = K$, stop. Otherwise set $j = j + 1$ and go to step 2.

If the model is exactly identified, this algorithm produces a $K \times K$ orthogonal rotation matrix

$$Q' = [q_1 \dots q_K]$$

such that $L_0 Q'$ represents a unique solution for B_0^{-1} , given the estimates of the reduced-form VAR model and the identifying restrictions. As in the case of estimates based on nonlinear equation solvers, it may be necessary to normalize the sign of the columns of this solution. It should be noted that Rubio-Ramírez, Waggoner, and Zha (2010) also propose a computationally more efficient algorithm designed specifically for triangular systems. We focus on the more general algorithm above because it also accommodates models with non-recursive B_0^{-1} .

Applying this algorithm to our empirical example, consider the initial guess

$$L_0 = \widehat{\Sigma}_u^{1/2} = \begin{bmatrix} 17.6784 & 0.0432 & 0.0499 \\ 0.0432 & 0.2225 & 0.0132 \\ 0.0499 & 0.0132 & 0.7446 \end{bmatrix},$$

where $\widehat{\Sigma}_u^{1/2}$ denotes the matrix square root of $\widehat{\Sigma}_u$. We define the restrictions to be imposed on L_0 , starting with the structural shock that implies the most exclusion restrictions, which is the third structural shock in the original specification, followed by the structural shock that implies the second-largest number of exclusion restrictions, which is the second structural shock. Given the order of the variables in $y_t = (\Delta rpoil_t, \Delta p_t, \Delta gdp_t)'$, the first column of Z_j , $j = 1, 2, 3$, corresponds to the variable $\Delta rpoil_t$, the second column to Δp_t , and the third column to Δgdp_t . Hence,

$$Z_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Z_2 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad Z_3 = 0_{3 \times 3}.$$

By deleting the rows of zeros from the Z_j s, we obtain

$$\bar{Z}_1 = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \end{bmatrix},$$

$$\bar{Z}_2 = (1, 0, 0).$$

Since all rows of Z_3 are zeros, there is no \bar{Z}_3 .

For $j = 1$, we obtain

$$\mathcal{Q}'_1 = \tilde{Z}_1 L_0 = \begin{bmatrix} 17.6784 & 0.0432 & 0.0499 \\ 0.0432 & 0.2225 & 0.0132 \end{bmatrix}.$$

The QR decomposition for \mathcal{Q}_1 yields

$$\tilde{\mathcal{Q}} = \begin{bmatrix} -1.0000 & 0.0026 & -0.0027 \\ -0.0024 & -0.9983 & -0.0585 \\ -0.0028 & -0.0585 & 0.9983 \end{bmatrix},$$

so $q'_1 = [-0.0027 \quad -0.0585 \quad 0.9983]$, where q_1 is the last column of $\tilde{\mathcal{Q}}$. For $j = 2$, we obtain

$$\mathcal{Q}'_2 = \begin{bmatrix} \tilde{Z}_2 L_0 \\ q'_1 \end{bmatrix} = \begin{bmatrix} 17.6784 & 0.0432 & 0.0499 \\ -0.0027 & -0.0585 & 0.9983 \end{bmatrix}.$$

The QR decomposition of \mathcal{Q}_2 yields

$$\tilde{\mathcal{Q}} = \begin{bmatrix} -1.0000 & -0.0027 & -0.0026 \\ -0.0024 & -0.0585 & 0.9983 \\ -0.0028 & 0.9983 & 0.0585 \end{bmatrix},$$

and $q'_2 = [-0.0026 \quad 0.9983 \quad 0.0585]$ is obtained from the last column of $\tilde{\mathcal{Q}}$. For $j = 3$, we obtain

$$\mathcal{Q}'_3 = \begin{bmatrix} q'_1 \\ q'_2 \end{bmatrix} = \begin{bmatrix} -0.0027 & -0.0585 & 0.9983 \\ -0.0026 & 0.9983 & 0.0585 \end{bmatrix},$$

because $\tilde{Z}_3 L_0$ is the empty matrix. Finally, the QR decomposition of \mathcal{Q}_3 yields

$$\tilde{\mathcal{Q}} = \begin{bmatrix} -0.0027 & 0.0026 & 1.0000 \\ -0.0585 & -0.9983 & 0.0024 \\ 0.9983 & -0.0585 & 0.0028 \end{bmatrix},$$

and $q'_3 = [1.0000 \quad 0.0024 \quad 0.0028]$ is obtained from the last column of $\tilde{\mathcal{Q}}$.

Upon completing this subroutine, we obtain the following solution for the restricted rotation matrix:

$$\mathcal{Q}' = [q_3 \quad q_2 \quad q_1] = \begin{bmatrix} 1.0000 & -0.0026 & -0.0027 \\ 0.0024 & 0.9983 & -0.0585 \\ 0.0028 & 0.0585 & 0.9983 \end{bmatrix},$$

where the ordering of q_1 , q_2 , and q_3 has been reversed to match the ordering of the structural shocks in the original recursive model specification. The implied solution for the structural impact multiplier matrix is

$$L_0 \mathcal{Q}' = \begin{bmatrix} 17.6784 & 0 & 0 \\ 0.0438 & 0.2227 & 0 \\ 0.0520 & 0.0566 & 0.7424 \end{bmatrix}.$$

In this example there is no need to flip the signs of any of the columns to ensure that the diagonal elements of \widehat{B}_0^{-1} are positive. We obtain the estimate

$$\widehat{B}_0^{-1} = \begin{bmatrix} 17.6784 & 0 & 0 \\ 0.0438 & 0.2227 & 0 \\ 0.0520 & 0.0566 & 0.7424 \end{bmatrix},$$

which matches the lower-triangular Cholesky decomposition. Of course, there is no advantage over the use of the Cholesky decomposition in the recursive setting considered here, but this example is nevertheless instructive and helps prepare for more general applications of this algorithm to nonrecursive structural models, which cannot be estimated by applying a Cholesky decomposition.

9.2.2 Nonrecursively Identified Models

For expository purposes, we focus on a quarterly model of U.S. monetary policy due to Keating (1992) that has already been reviewed in Chapter 8. Let $y_t = (\Delta p_t, \Delta gnp_t, i_t, \Delta m_t)'$, where p_t refers to the log of the GNP deflator, gnp_t to the log of real GNP, i_t to the federal funds rate, averaged by quarter, and m_t to the log of M1. The VAR model includes four lags and an intercept. The estimation period is restricted to 1959q2-2007q4 in order to exclude the recent period of unconventional monetary policy measures. The unrestricted reduced-form estimates are

$$\begin{aligned}\widehat{A}_1 &= \begin{bmatrix} 0.4885 & -0.0240 & 0.0693 & 0.0290 \\ 0.1695 & 0.1826 & 0.0124 & 0.0013 \\ 0.5739 & 0.3156 & 1.1325 & 0.0656 \\ 0.0003 & -0.0671 & -0.2889 & 0.1904 \end{bmatrix}, \\ \widehat{A}_2 &= \begin{bmatrix} 0.1255 & -0.0187 & -0.0543 & -0.0134 \\ 0.0885 & 0.2456 & -0.3570 & 0.0859 \\ 0.5174 & 0.2097 & -0.5362 & -0.0638 \\ 0.2851 & -0.0118 & 0.3221 & 0.3370 \end{bmatrix}, \\ \widehat{A}_3 &= \begin{bmatrix} 0.1491 & 0.0187 & -0.0092 & 0.0173 \\ -0.3633 & -0.0011 & 0.3065 & -0.0240 \\ -0.3880 & 0.0201 & 0.5237 & 0.0620 \\ -0.3367 & 0.0909 & -0.1357 & 0.0702 \end{bmatrix}, \\ \widehat{A}_4 &= \begin{bmatrix} 0.1962 & 0.0607 & -0.0160 & 0.0041 \\ 0.1320 & 0.0341 & -0.0126 & -0.0135 \\ -0.2784 & -0.0327 & -0.1876 & -0.0102 \\ 0.1903 & 0.0846 & 0.1359 & -0.0013 \end{bmatrix},\end{aligned}$$

and

$$\widehat{\Sigma}_w = \begin{bmatrix} 0.0611 & -0.0153 & 0.0424 & 0.0038 \\ -0.0153 & 0.5230 & 0.0797 & 0.0306 \\ 0.0424 & 0.0797 & 0.7169 & -0.2451 \\ 0.0038 & 0.0306 & -0.2451 & 1.1093 \end{bmatrix}.$$

The structural shock vector $w_t = (w_t^{AS}, w_t^{IS}, w_t^{MS}, w_t^{MD})'$ includes an aggregate supply shock, an IS shock, a money supply shock, and a money demand shock. The structural model can be written as

$$\begin{pmatrix} u_t^P \\ u_t^{gnp} \\ u_t^i \\ u_t^m \end{pmatrix} = \begin{pmatrix} w_t^{AS} \\ -b_{21,0}u_t^P - b_{23,0}u_t^i - b_{24,0}u_t^m + w_t^{IS} \\ -b_{34,0}u_t^m + w_t^{MS} \\ -b_{41,0}(u_t^{gnp} + u_t^P) - b_{43,0}u_t^i + w_t^{MD} \end{pmatrix}.$$

As discussed earlier, the first equation represents a horizontal AS curve. The second equation can be interpreted as an IS curve, allowing real output to respond to all other model variables. The third equation represents a simple money supply function, according to which the central bank adjusts the rate of interest in relation to the money stock, and the fourth equation is a money demand function in which short-run money holdings rise in proportion to nominal income, yielding the final restriction required for exact identification.

The identifying restrictions may be summarized as

$$B_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ b_{21,0} & 1 & b_{23,0} & b_{24,0} \\ 0 & 0 & 1 & b_{34,0} \\ b_{41,0} & b_{41,0} & b_{43,0} & 1 \end{bmatrix}.$$

Note that the model imposes the normalizing assumption that the diagonal elements of B_0 equal unity and leaves the diagonal elements of Σ_w unrestricted, except for their positive sign. All identifying restrictions are imposed on B_0 rather than B_0^{-1} . Clearly, B_0 is not recursive, preventing the use of a Cholesky decomposition.

Of the method-of-moments estimation methods we discussed so far only the use of a nonlinear equation solver is practical in this nonrecursive example.

Estimating the Model Using a Nonlinear Equation Solver. The objective is to find the unknown elements of B_0 and Σ_w such that

$$\left[\begin{array}{l} \text{vech} (B_0 \widehat{\Sigma}_w B_0' - \Sigma_w) \\ b_{12,0} \\ b_{13,0} \\ b_{14,0} \\ b_{31,0} \\ b_{32,0} \\ b_{41,0} - b_{42,0} \\ b_{11,0} - 1 \\ b_{22,0} - 1 \\ b_{33,0} - 1 \\ b_{44,0} - 1 \\ \sigma_{21,w} \\ \sigma_{31,w} \\ \sigma_{32,w} \\ \sigma_{41,w} \\ \sigma_{42,w} \\ \sigma_{43,w} \end{array} \right] = 0, \quad (9.2.5)$$

where the elements on the diagonal of Σ_w are restricted to be positive. The non-linear equation solver iterates on expression (9.2.5) until convergence, given initial guesses for B_0 and Σ_w . As noted earlier, the restrictions on the off-diagonal elements of Σ_w and on the diagonal elements of B_0 should be imposed directly to increase computational efficiency. The solution for B_0 is

$$\widehat{B}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -0.2669 & 1 & 0.7288 & 0.1784 \\ 0 & 0 & 1 & -11.2057 \\ -3.2443 & -3.2443 & 3.4133 & 1 \end{bmatrix},$$

$$\widehat{\Sigma}_w = \begin{bmatrix} 0.0611 & 0 & 0 & 0 \\ 0 & 0.9981 & 0 & 0 \\ 0 & 0 & 145.4997 & 0 \\ 0 & 0 & 0 & 10.6879 \end{bmatrix},$$

and the implied structural impact multiplier matrix is

$$\widehat{B}_0^{-1} \widehat{\Sigma}_w^{1/2} = \begin{bmatrix} 0.2471 & 0 & 0 & 0 \\ -0.0618 & 0.5912 & -0.0218 & -0.4114 \\ 0.1716 & 0.5476 & 0.2871 & 0.5524 \\ 0.0153 & 0.0489 & -1.0508 & 0.0493 \end{bmatrix}.$$

Figure 9.2 shows the implied responses to an unexpected upward shift of the aggregate supply curve. Given that the aggregate supply equation is normalized on the price level, the aggregate supply shock raises the price deflator and lowers real GNP. The effect on the federal funds rate is positive and the effect on M1 is ultimately positive. Of course, this example is just an illustration.

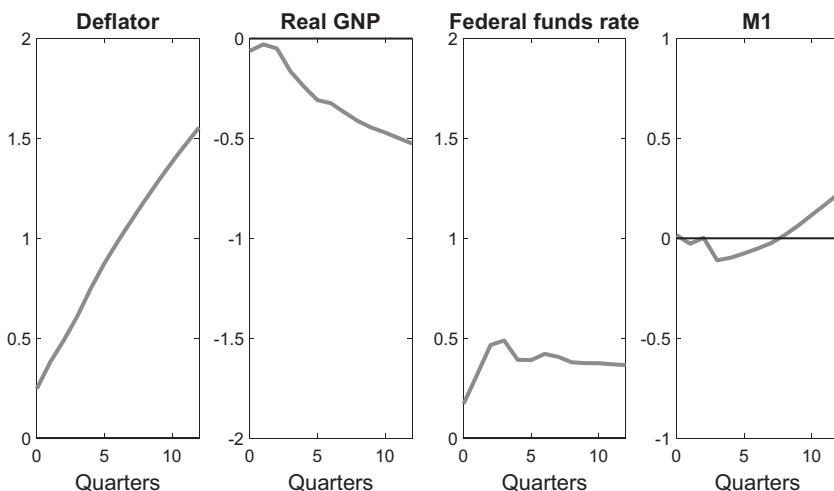


Figure 9.2. Responses of the U.S. economy to an aggregate supply shock in the Keating (1992) model.

It may not be appropriate for the U.S. economy during the sample period in question.

9.2.3 GMM Estimation of Overidentified Models

So far we have focused on models that are just identified. There are situations, however, in which there are more identifying restrictions on B_0^{-1} (or B_0) than are necessary to achieve identification. A common approach to estimating such models is to rely on the generalized method-of-moments (GMM) estimator originally proposed by Hansen (1982). Because there are more moment conditions than unknown parameters, the method-of-moments estimator cannot satisfy all moment conditions at the same time. Instead, the objective is to search for the structural parameter values that minimize a weighted average of the moment conditions where the asymptotically optimal set of weights corresponds to the inverse of the variance-covariance matrix of the sample moment conditions. This GMM estimator may be implemented in two steps, much like the method-of-moments estimator discussed earlier. The first step is to obtain consistent estimates of the reduced-form VAR model. The second step is to equate the sample second moment of the reduced-form VAR innovations with the same second moment written as function of the structural parameters in B_0^{-1} . Building on Hansen (1982), Bernanke and Mihov (1995) show that this GMM estimator is consistent and asymptotically normal. It is also efficient even without accounting for the uncertainty about the estimates

of the variance-covariance matrix of the VAR residuals underlying the weighting matrix (see also Watson 1994).

Restrictions on B_0^{-1} . The relevant moment condition for estimating B_0^{-1} is

$$\mathbb{E}(\text{vech}(u_t u_t') - \text{vech}(B_0^{-1} B_0^{-1'})) = 0.$$

where the normalizing assumption $\Sigma_w = I_K$ is imposed. Using the empirical counterpart of the expectation, we have

$$\text{vech}\left(\frac{1}{T} \sum_{t=1}^T \widehat{u}_t \widehat{u}_t' - B_0^{-1} B_0^{-1'}\right) = \text{vech}(\widetilde{\Sigma}_u) - \text{vech}(B_0^{-1} B_0^{-1'}).$$

This means that we are searching for B_0^{-1} such that this expression is as close to zero as possible. We therefore choose the unknown elements of B_0^{-1} to minimize

$$J = T(\text{vech}(\widetilde{\Sigma}_u) - \text{vech}(B_0^{-1} B_0^{-1'}))' \widehat{W} (\text{vech}(\widetilde{\Sigma}_u) - \text{vech}(B_0^{-1} B_0^{-1'})),$$

where all identifying and overidentifying restrictions have been imposed on B_0^{-1} and \widehat{W} is the inverse of the estimate of the variance-covariance matrix of the sample moment restrictions. More specifically,

$$\widehat{W} = \left[\frac{1}{T} \sum_{t=1}^T \left(\text{vech}(\widehat{u}_t \widehat{u}_t') - \overline{\text{vech}(\widehat{u} \widehat{u}')}\right) \left(\text{vech}(\widehat{u}_t \widehat{u}_t') - \overline{\text{vech}(\widehat{u} \widehat{u}')}\right)' \right]^{-1}$$

where

$$\overline{\text{vech}(\widehat{u} \widehat{u}')} = \frac{1}{T} \sum_{t=1}^T \text{vech}(\widehat{u}_t \widehat{u}_t').$$

Implementing the GMM estimator for overidentified models thus necessarily requires iteration. An empirical example for this GMM estimator can be found in Chapter 12.

Hansen (1982) shows that under the null hypothesis that the overidentifying restrictions are correct, under general conditions,

$$\hat{J} \xrightarrow{d} \chi_n^2,$$

where n is the number of overidentifying restrictions and \hat{J} is the J -statistic evaluated at \widehat{B}_0^{-1} . This result allows a formal test of overidentifying restrictions conditional on the premise that the remaining restrictions are correct.

Restrictions on B_0 . If identifying restrictions are imposed on B_0 rather than B_0^{-1} , it is possible to employ an alternative GMM estimation algorithm. This alternative estimator requires the user to take a stand on which restrictions are exactly identifying the structural model and which are overidentifying. We

start with the exactly identifying restrictions and write the method-of-moments estimator in terms of the structural VAR representation.

For expository purposes, let us for the moment assume that we are dealing with a VAR(0) model such that $B_0 y_t = w_t$, and that the model can be exactly identified using only exclusion restrictions.² Then the k^{th} equation of the structural-form VAR model can be written as

$$y_{kt} = \mathbf{y}'_{kt} b_k + w_{kt},$$

where \mathbf{y}_{kt} is the column vector of the elements of y_t that appear on the right-hand side of the k^{th} equation, where y_{kt} is the left-hand side variable in the k^{th} equation, and where b_k is the vector of associated structural parameters. For example, if B_0 is a lower-triangular matrix, there are no parameters to estimate in the first equation. For the remaining equations, $k = 2, \dots, K$, we have $\mathbf{y}'_{kt} = (y_{1t}, \dots, y_{k-1,t})'$ and $b_k = (-b_{k1,0}, \dots, -b_{k,k-1,0})'$.

For $t = 1, \dots, T$, we obtain in matrix notation

$$\mathbf{y}_k = Y_k b_k + \mathbf{w}_k, \quad (9.2.6)$$

where $\mathbf{y}_k \equiv (y_{k1}, \dots, y_{kT})'$, $Y_k \equiv [\mathbf{y}_{k1}, \dots, \mathbf{y}_{kT}]'$ and $\mathbf{w}_k \equiv (w_{k1}, \dots, w_{kT})'$. Assuming that the k^{th} equation is identified such that all right-hand side regressors are uncorrelated with w_{kt} , we can use the moment conditions

$$\mathbb{E}(\mathbf{y}_{kt} w_{kt}) = 0$$

to estimate b_k . In other words, \hat{b}_k minimizes the objective function

$$J(b_k) = (\mathbf{y}_k - Y_k b_k)'(\mathbf{y}_k - Y_k b_k).$$

To incorporate the overidentifying restrictions into the estimation requires a modification of this objective function. If the overidentifying restrictions are exclusion restrictions, we can express the linear restrictions on the parameters in the k^{th} equation as

$$b_k = R_k \gamma_k, \quad (9.2.7)$$

where R_k is a fixed given matrix that specifies the restrictions and γ_k is the vector of unrestricted parameters. For example, if the overidentifying restriction involves a zero restriction on the first parameter in the third equation of B_0 in a lower-triangular model, we have

$$R_3 = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad \text{and} \quad \gamma_3 = -b_{32,0}.$$

² The estimation method must be modified if the exactly identifying restrictions involve nonexclusion restrictions. This complicates the exposition. For illustrative purposes we focus on the more common case of exclusion restrictions.

As another example, if the overidentifying restriction forces to zero the first parameter in the second equation of the lower-triangular model, there are no parameters left to be estimated in that equation.

Overidentifying restrictions on B_0 other than exclusion restrictions may be imposed as well. For expository purposes, suppose that there are further linear restrictions on the parameters in the k^{th} equation. Such restrictions can also be expressed in the form of equation (9.2.7). For example, if we wish to impose in a four-dimensional model that $b_{41,0} = b_{42,0}$,

$$R_4 = \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \quad \text{and} \quad \gamma_4 = \begin{pmatrix} -b_{41,0} \\ -b_{43,0} \end{pmatrix}.$$

Having defined the matrix R_k , let $Y_k^\dagger = Y_k R_k$ and estimate γ_k in the regression equation

$$\mathbf{y}_k = Y_k^\dagger \gamma_k + \mathbf{w}_k.$$

Because we have more moment conditions than parameters in the k^{th} equation, we estimate γ_k by minimizing the GMM objective function

$$\begin{aligned} J(\gamma_k) &= \left(\frac{1}{T} \mathbf{w}'_k Y_k \right) \left(\frac{1}{T} Y'_k Y_k \right)^{-1} \left(\frac{1}{T} Y'_k \mathbf{w}_k \right) \\ &= \left(\frac{1}{T} (\mathbf{y}_k - Y_k^\dagger \gamma_k)' Y_k \right) \left(\frac{1}{T} Y'_k Y_k \right)^{-1} \left(\frac{1}{T} Y'_k (\mathbf{y}_k - Y_k^\dagger \gamma_k) \right). \end{aligned}$$

This GMM estimator has the closed-form solution

$$\hat{\gamma}_k^{GMM} = (\hat{Y}_k^\dagger)' Y_k^{-1} \hat{Y}_k^\dagger \mathbf{y}_k, \quad (9.2.8)$$

where $\hat{Y}_k^\dagger \equiv Y_k (Y'_k Y_k)^{-1} Y'_k Y_k^\dagger$. As is well known, if the errors are iid, this GMM estimator is efficient.

So far we have only discussed the estimation of the k^{th} equation of the structural VAR model. Since the components of w_t are instantaneously uncorrelated, single-equation GMM is identical to estimating the full system

$$\begin{pmatrix} \mathbf{y}_1 \\ \vdots \\ \mathbf{y}_K \end{pmatrix} = \begin{bmatrix} Y_1^\dagger & & 0 \\ & \ddots & \\ 0 & & Y_K^\dagger \end{bmatrix} \begin{pmatrix} \gamma_1 \\ \vdots \\ \gamma_K \end{pmatrix} + \begin{pmatrix} \mathbf{w}_1 \\ \vdots \\ \mathbf{w}_K \end{pmatrix}$$

by GMM. Note that there may be equations without any parameters to estimate. Such equations are dropped from the system.

In practice, we typically do not deal with VAR(0) processes, and this procedure must be adapted to allow for lagged variables. In that case the k^{th} equation becomes

$$y_{kt} = \mathbf{y}'_{kt} R \gamma_k + Y'_{t-1} \mathbf{b}_k + w_{kt}, \quad (9.2.9)$$

where $Y'_{t-1} \equiv (y'_{t-1}, \dots, y'_{t-p})$, as before, and \mathbf{b}'_k is the k^{th} row of $B = [B_1, \dots, B_p]$. Defining $Z \equiv [Y_0, \dots, Y_{T-1}]$, the same analysis as for the VAR(0) case can be applied to

$$\mathbf{y}_k = [Y_k^\dagger, Z'] \begin{pmatrix} \gamma_k \\ \mathbf{b}_k \end{pmatrix} + \mathbf{w}_k. \quad (9.2.10)$$

The GMM estimator of the parameters then becomes

$$\begin{pmatrix} \widehat{\gamma}_k^{GMM} \\ \widehat{\mathbf{b}}_k^{GMM} \end{pmatrix} = \left(\widehat{Z}_k [Y_k^\dagger, Z'] \right)^{-1} \widehat{Z}_k \mathbf{y}_k,$$

where

$$\widehat{Z}'_k = [Y_k, Z'] \left(\begin{bmatrix} Y'_k \\ Z \end{bmatrix} [Y_k, Z'] \right)^{-1} \begin{bmatrix} Y'_k \\ Z \end{bmatrix} [Y_k^\dagger, Z'].$$

This setup facilitates not only the computation of the estimates, but also the derivation of the asymptotic distribution. Under conditions that are satisfied for stationary VAR processes with martingale difference innovations, these estimators are known to be consistent and asymptotically normally distributed. The joint distribution of the GMM estimator of all parameters of the system is also asymptotically normal and can be obtained by considering the system of all K equations jointly. In practice, the GMM estimator may also be bootstrapped, as discussed in Chapter 12.

Equivalently the estimator for γ_k may be computed by concentrating out the lags and replacing the observed variables by the corresponding residuals. In other words, we estimate the reduced-form model

$$y_t = [A_1, \dots, A_p] Y_{t-1} + u_t$$

by LS and use the residuals \widehat{u}_t instead of the y_t in the expressions for the VAR(0) model. It is important to note, however, that standard estimates of the covariance matrix of this estimator will be invalid because replacing the observations in the VAR(0) model by residuals obtained from a VAR(p) model introduces a generated-regressor problem, which invalidates conventional estimates of the standard errors of the structural parameters (see Pagan 1984).

Having obtained estimates of γ_k , we construct the corresponding estimates of the elements of B_0 from $b_k = R_k \gamma_k$. As discussed earlier, the construction of the structural impact multiplier matrix requires not only an estimate of B_0 but also an estimate of the diagonal matrix Σ_w . The diagonal elements of this matrix are estimated as

$$\hat{\sigma}_{w_k}^2 = \frac{1}{T} \sum_{t=1}^T \hat{w}_{kt}^2.$$

Restrictions on B_0^{-1} and B_0 . If there are identifying restrictions on both B_0^{-1} and B_0 in an overidentified structural VAR model, one minimizes the same objective function as in the case when there are only restrictions on B_0^{-1} ,

$$J = T(\text{vech}(\widehat{\Sigma}_u) - \text{vech}(B_0^{-1}B_0^{-1'}))' \widehat{W} (\text{vech}(\widehat{\Sigma}_u) - \text{vech}(B_0^{-1}B_0^{-1'})).$$

The difference is that the objective function must be minimized subject to the additional identifying restrictions on B_0 rather than by unconstrained minimization of J . This involves nonlinear restrictions on B_0^{-1} . Unlike in the case of restrictions on B_0 only, there is no closed-form solution in this case. Under the conditions invoked in the discussion of the GMM estimator without restrictions on B_0 , this procedure generates a consistent and asymptotically normal estimator of B_0^{-1} and hence B_0 .

9.3 Instrumental Variable Estimation

If there are identifying restrictions on B_0^{-1} that cannot be expressed as exclusion restrictions on B_0 , the GMM estimator must be constructed by iterative techniques. In contrast, if the identifying restrictions can be imposed on B_0 , GMM estimates may also be constructed using traditional IV estimators based on linear regressions.

IV Analysis of the Structural VAR(0) Model. For expository purposes we again assume for the moment that we are dealing with a VAR(0) model. Then, as in the previous section, the k^{th} equation of the structural-form VAR model can be written as

$$y_{kt} = \mathbf{y}'_{kt} b_k + w_{kt},$$

where \mathbf{y}_{kt} is the column vector of the elements of y_t that appear on the right-hand side of the k^{th} equation, where y_{kt} is the left-hand side variable in the k^{th} equation, and where b_k is the vector of associated structural parameters. Equivalently, in matrix notation

$$\mathbf{y}_k = Y_k b_k + \mathbf{w}_k, \tag{9.3.1}$$

where all notation is defined as in equation (9.2.6). Assuming that the k^{th} equation is identified such that all right-hand side variables are uncorrelated with w_{kt} , we can use Y_k as the instrument, which is equivalent to estimating equation (9.3.1) by LS:

$$\hat{b}_k^{IV} = \hat{b}_k^{LS} = (Y'_k Y_k)^{-1} Y'_k \mathbf{y}_k. \tag{9.3.2}$$

Note that this is also the method-of-moments estimator given moment conditions $\mathbb{E}(\mathbf{y}'_{kt} w_{kt}) = 0$. Thus, for just-identified models, we have

$$\hat{b}_k^{LS} = \hat{b}_k^{IV} = \hat{b}_k^{GMM}.$$

This IV estimation method also works if there are overidentifying restrictions. Suppose that there are further linear restrictions on the parameters in the k^{th} equation such that

$$b_k = R_k \gamma_k, \quad (9.3.3)$$

where R_k is a matrix that relates the unrestricted parameters γ_k to the restricted parameter vector b_k .

Having defined R_k , we recover estimates of γ_k from the regression

$$\mathbf{y}_k = Y_k^\dagger \gamma_k + \mathbf{w}_k,$$

where $Y_k^\dagger = Y_k R_k$. Because we have more instruments in Y_k than parameters in the k^{th} equation, we can estimate this regression model by two-stage LS (2SLS) using the instruments $\widehat{Y}_k^\dagger \equiv Y_k (Y_k' Y_k)^{-1} Y_k'$:

$$\widehat{\gamma}_k^{IV} = \widehat{\gamma}_k^{2SLS} = (\widehat{Y}_k^\dagger Y_k^\dagger)^{-1} \widehat{Y}_k^\dagger \mathbf{y}_k. \quad (9.3.4)$$

As is well known, if the errors are iid, this is also the GMM estimator based on the moment conditions $\mathbb{E}(\mathbf{y}_{kt}' w_{kt}) = 0$ and the optimal weighting matrix, because

$$\widehat{\gamma}_k^{GMM} = \widehat{\gamma}_k^{2SLS}$$

is precisely the solution that minimizes the GMM objective function.

IV Analysis of the Structural VAR(p) Model. The discussion so far focused on the VAR(0) model. There are two ways of generalizing this method to the VAR(p) model, $p > 0$. The first approach is to adapt equation (9.3.1) to include p lags of the model variables and to augment the set of instruments accordingly. In that case, the k^{th} equation becomes

$$y_{kt} = \mathbf{y}_{kt}' R_k \gamma_k + Y_{t-1}' \mathbf{b}_k + w_{kt}, \quad (9.3.5)$$

where $Y_{t-1}' \equiv (y_{t-1}', \dots, y_{t-p}')$ and \mathbf{b}_k' is the k^{th} row of $B = [B_1, \dots, B_p]$, as before. Defining $Z \equiv [Y_0, \dots, Y_{T-1}]$, the same analysis as for the VAR(0) case can be applied to

$$\mathbf{y}_k = [Y_k^\dagger, Z'] \begin{pmatrix} \gamma_k \\ \mathbf{b}_k \end{pmatrix} + \mathbf{w}_k. \quad (9.3.6)$$

The IV estimator of the parameters then becomes

$$\begin{pmatrix} \widehat{\gamma}_k^{IV} \\ \widehat{\mathbf{b}}_k^{IV} \end{pmatrix} = \left(\widehat{Z}_k [Y_k^\dagger, Z'] \right)^{-1} \widehat{Z}_k \mathbf{y}_k,$$

where

$$\widehat{Z}_k' = [Y_k, Z'] \left(\begin{bmatrix} Y_k' \\ Z \end{bmatrix} [Y_k, Z'] \right)^{-1} \begin{bmatrix} Y_k' \\ Z \end{bmatrix} [Y_k^\dagger, Z']. \quad (9.3.7)$$

An alternative approach to constructing the IV estimator of the VAR(p), $p > 0$, model is to concentrate out the lagged regressors and to replace the observed variables in the VAR(0) model by the residuals obtained from the VAR(p) model. Although this residual-based IV approach is simple, such regressions do not produce valid estimates of the covariance matrix of the IV estimator. Using residuals as instruments introduces a generated regressor problem (see Pagan 1984). Corrections of the covariance estimator are discussed in King and Watson (1997). An alternative is the use of the IV estimator as the starting value in constructing a single-iteration GMM estimator, which allows standard software to be used to consistently estimate the covariance of the IV/GMM estimator. The easiest solution, however, would be to include the lagged observables in IV estimation, as discussed earlier.

As in the discussion of single-equation GMM, having obtained estimates of γ_k , we construct the corresponding estimates of the elements of B_0 from $b_k = R_k \gamma_k$, and estimates of the diagonal elements of the matrix Σ_w are obtained as

$$\hat{\sigma}_{w_k}^2 = \frac{1}{T} \sum_{t=1}^T \hat{w}_{kt}^2.$$

To conclude, the advantage of the GMM approach is that it can handle restrictions on B_0^{-1} , whereas the IV approach cannot. The advantage of the IV approach is that it does not require iteration.

As in the GMM framework, the asymptotic properties of the IV estimator follow from standard theoretical arguments for IV estimators (see, e.g., Judge, Griffiths, Hill, Lütkepohl, and Lee 1985). If the instruments are valid and strong, the IV estimator under our assumptions is consistent and asymptotically normally distributed. One advantage of the IV estimator is that the strength of the instrument in the first stage may be evaluated empirically (see, e.g., Pagan and Robertson 1998).

The strength of the instrument increases with the correlation between the instruments and the regressors. For example, in model (9.3.6) the contemporaneous regressors are contained in the matrix Y_k^\dagger and the lagged regressors are in Z' . Both regressors are instrumented by \widehat{Z}_k , as defined in equation (9.3.7). For the structural VAR model to be identified, the matrix of correlations between the regressors and the instruments must be of full rank. The strength of the identification is measured by the conditioning number of this matrix. Evidence of this matrix being ill-conditioned would be indicative of the instruments being weak and would cast doubt on the reliability of the empirical estimates. Similar problems of imprecise estimates may, of course, also arise when estimating structural VAR models by the method of moments or by ML. Exploiting the equivalence between IV estimators and many non-IV estimators can be helpful in diagnosing such problems.

An Empirical Illustration. For the example of the three-dimensional recursive model used in Section 9.2.1, the IV estimator is identical to the GMM estimator. We obtain the following estimates:

$$\begin{aligned}
y_{1t} &= \widehat{v}_1^{IV} + \sum_{i=1}^4 (\widehat{b}_{11,i}^{IV} y_{1,t-i} + \widehat{b}_{12,i}^{IV} y_{2,t-i} + \widehat{b}_{13,i}^{IV} y_{3,t-i}) + \widehat{w}_{1t} \\
&= -0.6807 - 0.0064 y_{1,t-1} + 0.9365 y_{2,t-1} + 1.7510 y_{3,t-1} \\
&\quad - 0.1822 y_{1,t-2} + 12.6454 y_{2,t-2} - 3.4391 y_{3,t-2} \\
&\quad - 0.0073 y_{1,t-3} - 3.5986 y_{2,t-3} + 1.7337 y_{3,t-3} \\
&\quad - 0.1194 y_{1,t-4} - 8.1807 y_{2,t-4} + 0.9309 y_{3,t-4} + \widehat{w}_{1t}, \\
\widehat{\sigma}_{w_1}^2 &= \frac{1}{T} \sum_{t=1}^T \widehat{w}_{1t}^2 = 286.8106, \\
y_{2t} &= \widehat{v}_2^{IV} - \widehat{b}_{21,0}^{IV} y_{1t} \\
&\quad + \sum_{i=1}^4 (\widehat{b}_{21,i}^{IV} y_{1,t-i} + \widehat{b}_{22,i}^{IV} y_{2,t-i} + \widehat{b}_{23,i}^{IV} y_{3,t-i}) + \widehat{w}_{2t} \\
&= -0.0135 + 0.0025 y_{1t} \\
&\quad + 0.0018 y_{1,t-1} + 0.5968 y_{2,t-1} + 0.0239 y_{3,t-1} \\
&\quad + 0.0047 y_{1,t-2} + 0.0963 y_{2,t-2} - 0.0303 y_{3,t-2} \\
&\quad + 0.0022 y_{1,t-3} + 0.0566 y_{2,t-3} + 0.0191 y_{3,t-3} \\
&\quad + 0.0010 y_{1,t-4} + 0.1690 y_{2,t-4} + 0.0683 y_{3,t-4} + \widehat{w}_{2t}, \\
\widehat{\sigma}_{w_2}^2 &= \frac{1}{T} \sum_{t=1}^T \widehat{w}_{2t}^2 = 0.0455, \\
y_{3t} &= \widehat{v}_3^{IV} - \widehat{b}_{31,0}^{IV} y_{1t} - \widehat{b}_{32,0}^{IV} y_{2t} \\
&\quad + \sum_{i=1}^4 (\widehat{b}_{31,i}^{IV} y_{1,t-i} + \widehat{b}_{32,i}^{IV} y_{2,t-i} + \widehat{b}_{33,i}^{IV} y_{3,t-i}) + \widehat{w}_{3t} \\
&= 0.4418 + 0.0023 y_{1t} + 0.2540 y_{2t} \\
&\quad - 0.0028 y_{1,t-1} - 0.3731 y_{2,t-1} + 0.3211 y_{3,t-1} \\
&\quad - 0.0070 y_{1,t-2} + 0.3261 y_{2,t-2} + 0.1637 y_{3,t-2} \\
&\quad - 0.0007 y_{1,t-3} - 0.1451 y_{2,t-3} - 0.0538 y_{3,t-3} \\
&\quad - 0.0080 y_{1,t-4} - 0.1001 y_{2,t-4} - 0.0085 y_{3,t-4} + \widehat{w}_{3t}, \\
\widehat{\sigma}_{w_3}^2 &= \frac{1}{T} \sum_{t=1}^T \widehat{w}_{3t}^2 = 0.5058.
\end{aligned}$$

Since the model is recursive and hence just identified, the estimates are identical to the LS estimates for the three equations. For example, the estimates in the first equation are the same as those in the first rows of the estimated \widehat{A}_j reduced-form coefficient matrices in Section 9.2.1. The estimates of the variances of the structural errors are computed without degrees-of-freedom adjustment.

Collecting the estimates associated with unlagged regressors in the second and third equations, we obtain an estimate for B_0 ,

$$\widehat{B}_0^{IV} = \begin{bmatrix} 1 & 0 & 0 \\ -0.0025 & 1 & 0 \\ -0.0023 & -0.2540 & 1 \end{bmatrix},$$

with diagonal elements standardized to 1. Defining the IV estimator

$$\widehat{\Sigma}_w = \begin{bmatrix} 286.8106 & 0 & 0 \\ 0 & 0.0455 & 0 \\ 0 & 0 & 0.5058 \end{bmatrix},$$

the quantity $(\widehat{B}_0^{IV})^{-1}\widehat{\Sigma}_w^{1/2}$ produces the same structural impact multiplier matrix as the Cholesky decomposition apart from the degrees-of-freedom adjustment we used in earlier sections.

9.4 Full Information Maximum Likelihood Estimation

The IV estimator may also be viewed as an alternative to constructing the Gaussian ML estimator of just-identified structural VAR models. The advantage of the full information ML (FIML) estimator is that (like some GMM approaches) it allows for identifying restrictions on both B_0 and B_0^{-1} . Given the structural VAR model

$$B_0 y_t = B_0 A Y_{t-1} + w_t, \quad (9.4.1)$$

where $Y'_{t-1} \equiv (y'_{t-1}, \dots, y'_{t-p})$ and $A \equiv [A_1, \dots, A_p]$ denotes the reduced-form slope VAR parameters, as before. Assuming that w_t is Gaussian white noise with diagonal covariance matrix Σ_w and $w_t \sim \mathcal{N}(0, \Sigma_w)$, the associated reduced-form residuals are $u_t = B_0^{-1} w_t \sim \mathcal{N}(0, \Sigma_u = B_0^{-1} \Sigma_w B_0^{-1})$.

The corresponding log-likelihood function for a sample $Y \equiv [y_1, \dots, y_T]$ is

$$\begin{aligned} \log l(A, B_0, \Sigma_w) &= -\frac{KT}{2} \log(2\pi) - \frac{T}{2} \log(\det(B_0^{-1} \Sigma_w B_0'^{-1})) \\ &\quad - \frac{1}{2} \text{tr}\{(Y - AZ)'[B_0^{-1} \Sigma_w B_0'^{-1}]^{-1}(Y - AZ)\} \\ &= \text{constant} + \frac{T}{2} \log(\det(B_0)^2) - \frac{T}{2} \log(\det(\Sigma_w)) \\ &\quad - \frac{1}{2} \text{tr}\{B_0' \Sigma_w^{-1} B_0(Y - AZ)(Y - AZ)'\}, \end{aligned} \quad (9.4.2)$$

where $Z \equiv [Y_0, \dots, Y_{T-1}]$, as before (see Lütkepohl 2005, chapter 9).

If there are no restrictions on the reduced-form parameters, then, for given B_0 and Σ_w , the log-likelihood function $\log l(A, B_0, \Sigma_w)$ is maximized with respect to A by

$$\widehat{A} = YZ'(ZZ')^{-1} = \left(\sum_{t=1}^T y_t Y'_{t-1} \right) \left(\sum_{t=1}^T Y_{t-1} Y'_{t-1} \right)^{-1}$$

(see Section 2.3). Replacing A with this estimator results in the concentrated log-likelihood function

$$\begin{aligned} \log l_c(B_0, \Sigma_w) &= \text{constant} + \frac{T}{2} \log(\det(B_0)^2) \\ &\quad - \frac{T}{2} \log(\det(\Sigma_w)) - \frac{T}{2} \text{tr}(B'_0 \Sigma_w^{-1} B_0 \widetilde{\Sigma}_u), \end{aligned} \quad (9.4.3)$$

where $\widetilde{\Sigma}_u = T^{-1} \sum_{t=1}^T \widehat{u}_t \widehat{u}'_t$ and $\widehat{u}_t = y_t - \widehat{A}Y_{t-1}$ is the LS residual. In general, the concentrated log-likelihood function can be maximized by numerical methods with respect to B_0 and Σ_w , subject to the identifying restrictions. For just-identified models it can be shown that the maximum is obtained when

$$B'_0 \Sigma_w^{-1} B_0 = \widetilde{\Sigma}_u^{-1}$$

(Lütkepohl 2005, chapter 9). If the variances of the structural errors are normalized to one so that $\Sigma_w = I_K$, then the ML estimator \widetilde{B}_0 of B_0 satisfies

$$\widetilde{B}'_0 \widetilde{B}_0 = \widetilde{\Sigma}_u^{-1} \quad \text{or} \quad \widetilde{B}_0^{-1} \widetilde{B}'_0 = \widetilde{\Sigma}_u.$$

For the VAR example from Section 9.2.1, where B_0^{-1} is assumed to be lower triangular, we obtain

$$\widetilde{B}_0^{-1} = \begin{bmatrix} 16.9355 & 0 & 0 \\ 0.0419 & 0.2134 & 0 \\ 0.0498 & 0.0542 & 0.7112 \end{bmatrix}$$

which is identical to the Cholesky decomposition of $\widehat{\Sigma}_u$, as in Section 9.2.1, after accounting for the differences in the degrees-of-freedom adjustment. The corresponding ML estimate of B_0 is

$$\widetilde{B}_0 = \begin{bmatrix} 0.0590 & 0 & 0 \\ -0.0116 & 4.6864 & 0 \\ -0.0033 & -0.3572 & 1.4061 \end{bmatrix}.$$

If instead the diagonal elements of B_0 are normalized to one and the variances of the structural shocks are unrestricted, we obtain the ML estimates

$$\widetilde{B}_0 = \begin{bmatrix} 1 & 0 & 0 \\ -0.0025 & 1 & 0 \\ -0.0023 & -0.2540 & 1 \end{bmatrix}$$

and

$$\tilde{\Sigma}_w = \begin{bmatrix} 286.8106 & 0 & 0 \\ 0 & 0.0455 & 0 \\ 0 & 0 & 0.5058 \end{bmatrix}.$$

such that $\tilde{B}_0^{-1} \tilde{\Sigma}_w^{1/2}$ produces the structural impact multiplier matrix.

Of course, ML estimates can also be computed in the same manner if there are nonrecursive identifying restrictions as in the Keating example model from Section 9.2.2. The estimates for this example are identical to the solutions in Section 9.2.2 after adjusting for the degrees of freedom. More precisely, if the diagonal elements of B_0 are normalized to unity, we have

$$\tilde{B}_0 = \begin{bmatrix} 1 & 0 & 0 & 0 \\ -0.2669 & 1 & 0.7288 & 0.1784 \\ 0 & 0 & 1 & -11.2056 \\ -3.2443 & -3.2443 & 3.4133 & 1 \end{bmatrix}$$

and

$$\tilde{\Sigma}_w = \begin{bmatrix} 0.0556 & 0 & 0 & 0 \\ 0 & 0.9092 & 0 & 0 \\ 0 & 0 & 132.5486 & 0 \\ 0 & 0 & 0 & 9.7366 \end{bmatrix}.$$

In practice, if the model is just-identified, this ML estimator often is implemented by first estimating the reduced form and then applying numerical solution methods to recover \tilde{B}_0^{-1} from $\tilde{\Sigma}_w$. An alternative is to rely on iterative methods to maximize the concentrated likelihood subject to constraints on B_0 and/or B_0^{-1} . A detailed review of these numerical methods is beyond the scope of this book (see, however, Judge, Griffiths, Hill, Lütkepohl, and Lee 1985, appendix B). The same iterative algorithms may also be used when there are overidentifying restrictions.

If y_t is a stationary Gaussian $\text{VAR}(p)$ process, the ML estimation framework facilitates the derivation of the asymptotic properties of the estimator. Standard ML theory implies that the unrestricted parameters are consistent and asymptotically normal. Suppose that the unrestricted parameters are collected in the vector η which contains the reduced-form parameters A , the unrestricted elements of B_0 , and the diagonal elements of Σ_w if the latter are not standardized to one. Denoting the ML estimator of η by $\tilde{\eta}$, under general assumptions we obtain

$$\sqrt{T}(\tilde{\eta} - \eta) \xrightarrow{d} \mathcal{N}(0, \mathcal{I}_a^{-1}(\eta)),$$

where $\mathcal{I}_a(\eta)$ denotes the asymptotic information matrix. Note that the diagonal elements of $\tilde{\Sigma}_w$ also have an asymptotic normal distribution.

If there are overidentifying restrictions on B_0 , the restricted ML estimator of Σ_u may not equal the unrestricted ML estimator of the reduced-form covariance matrix anymore. In that case the reduced-form residual covariance matrix Σ_u is estimated as

$$\tilde{\Sigma}_u^r \equiv \tilde{B}_0^{-1} \tilde{\Sigma}_w \tilde{B}_0^{-1'}, \quad (9.4.4)$$

where \tilde{B}_0^{-1} and $\tilde{\Sigma}_w$ are the restricted ML estimators, and the LR statistic,

$$LR = T (\log(\det(\tilde{\Sigma}_u^r)) - \log(\det(\tilde{\Sigma}_u))), \quad (9.4.5)$$

can be used to test the overidentifying restrictions. Under general conditions, this LR statistic has an asymptotic χ^2 distribution with degrees of freedom equal to the number of overidentifying restrictions, provided the restrictions are correct.

An alternative approach to estimating structural VAR models by ML that allows for a combination of restrictions on B_0 and B_0^{-1} (as is the case, for example, in the Blanchard and Perotti (2002) model of fiscal policy discussed in Section 8.5.1), sometimes facilitates the analysis. Recall that in this class of models

$$B_0 u_t = C w_t$$

with $\Sigma_w = I_K$ such that $\Sigma_u = B_0^{-1} C C' B_0^{-1'}$. The implied concentrated log likelihood of this model is

$$\begin{aligned} \log l_c(B_0, C) &= \text{constant} + \frac{T}{2} \log(\det(B_0)^2) - \frac{T}{2} \log(\det(C)^2) \\ &\quad - \frac{T}{2} \text{tr}(B_0' C^{-1'} C^{-1} B_0 \tilde{\Sigma}_u). \end{aligned}$$

The advantage of this setup is that one can impose the identifying restrictions directly on B_0 and C allowing the unconstrained maximization of the log-likelihood with respect to the unrestricted elements of B_0 and C . In addition, overidentifying restrictions may also be imposed directly. The resulting Gaussian ML estimator is consistent and asymptotically normal (see Lütkepohl 2005, section 9.3.1). Tests for overidentifying restrictions may be conducted as discussed earlier with $\tilde{\Sigma}_u^r \equiv \tilde{B}_0^{-1} \tilde{C} \tilde{C}' \tilde{B}_0^{-1'}$.

9.5 Bayesian Estimation

By definition, Bayesian estimators are obtained by evaluating the posterior distribution of the statistic of interest. They minimize the posterior expected value of a user-defined loss function. Under quadratic loss, for example, the Bayesian estimator of a scalar parameter would be the mean of the posterior

distribution of this parameter, and, under absolute loss, it would be the posterior median.

As discussed in Chapter 5, if the posterior is Gaussian, minimizing a symmetric loss function such as the quadratic or absolute loss function produces a Bayesian estimator that asymptotically coincides with the usual LS/ML estimator. More generally, under alternative loss functions or when the posterior distribution is not Gaussian, the Bayesian estimator may differ from conventional frequentist estimators even asymptotically.

The construction of the posterior distribution of the structural model parameters starts with the specification of a prior. The prior distributions discussed in Chapter 5 were for reduced-form VAR model parameters. For structural VAR analysis there are two common approaches. One is to rely on a conventional reduced-form prior such as the Gaussian-inverse Wishart prior and to generate posterior draws for the structural model parameters by suitably transforming each reduced-form posterior draw, as outlined in Chapter 5 (see also, e.g., Canova 1991; Gordon and Leeper 1994). In other words, if $\Sigma_u^{(i)}$ is the i^{th} draw from the posterior distribution of Σ_u , then the corresponding posterior draw for B_0 (or, equivalently, the corresponding draws for B_0^{-1} and Σ_w) may be constructed from $\Sigma_u^{(i)}$ by any of the method-of-moments methods already discussed in this chapter. For example, if a recursive identification scheme is used, a lower-triangular Cholesky decomposition of $\Sigma_u^{(i)}$ yields the i^{th} posterior draw for B_0^{-1} .

This widely used approach, however, lacks a theoretical foundation, as stressed by Sims and Zha (1998). One concern is that standard reduced-form priors make a distinction between own lags and other lags in each VAR equation that does not exist in the structural VAR representation. Moreover, the approach employed by Canova (1991) and Gordon and Leeper (1994), among many others, is practically feasible only when the structural VAR model is just identified. When B_0^{-1} (or B_0) is overidentified, it is necessary to impose the prior directly on the structural VAR representation, because the overidentifying restrictions impose restrictions on the reduced-form model. An alternative approach that allows one to impose the priors directly on the structural VAR model parameters is developed in Sims and Zha (1998, 1999). It may be applied regardless of whether the structural model is just identified or overidentified.

Recall that the Gaussian VAR(p) model may be expressed in structural form as

$$B_0 y_t = B Y_{t-1} + w_t, \quad w_t \sim \mathcal{N}(0, I_K),$$

where $Y'_{t-1} \equiv (y'_{t-1}, \dots, y'_{t-p})$ and $B \equiv [B_1, \dots, B_p]$. Stacking the observations for $t = 1, \dots, T$, yields

$$B_0 Y - B Z = \mathbf{B} X = W,$$

where $Y \equiv [y_1, \dots, y_T]$ is $K \times T$, $W \equiv [w_1, \dots, w_T]$ is $K \times T$, $Z \equiv [Y_0, \dots, Y_{T-1}]$ is $Kp \times T$, and

$$\mathbf{B} \equiv [B_0, B], \quad \text{and} \quad X \equiv \begin{bmatrix} Y \\ -Z \end{bmatrix}$$

are of dimension $K \times K(p+1)$ and $K(p+1) \times T$, respectively. Let $\mathbf{b} = \text{vec}(\mathbf{B})$, $b_0 = \text{vec}(B_0)$, and $b = \text{vec}(B)$. Then the likelihood function is

$$\begin{aligned} l(\mathbf{B}|Y) &\propto \det(B_0)^T \exp\left(-\frac{1}{2}\text{tr}(\mathbf{B}X)(\mathbf{B}X)'\right) \\ &= \det(B_0)^T \exp\left(-\frac{1}{2}\mathbf{b}'(XX' \otimes I_K)\mathbf{b}\right). \end{aligned}$$

The prior distribution for b_0 may have singularities caused by identifying restrictions. Let $g(b_0)$ be the prior for the parameters in B_0 and let $g(b|b_0)$ denote the density of $\mathcal{N}(\bar{b}(b_0), \bar{V}_b(b_0))$, where $\bar{b}(b_0)$ and $\bar{V}_b(b_0)$ denote the mean and the covariance matrix of the prior distribution of b conditional on b_0 . Then $g(\mathbf{b}) = g(b_0)g(b|b_0)$ and the posterior density of \mathbf{b} is

$$g(\mathbf{b}|Y) \propto g(\mathbf{b})l(\mathbf{B}|Y). \quad (9.5.1)$$

This posterior density is nonstandard in general. It can be shown that for fixed b_0 , the conditional posterior distribution of b given b_0 is Gaussian. Sims and Zha (1998) also derive the marginal posterior density of b_0 . An important question is how to specify the marginal prior of b_0 in this model. Sims and Zha (1998) recommend specifying a joint normal prior on the nonzero elements of B_0 in the case of a lower triangular normalization of B_0 .³

As discussed in Canova (2007, chapter 10.3, pp. 391–392), it is instructive to compare the structural priors discussed in Sims and Zha (1998) with more traditional Minnesota priors on the reduced-form parameters. There are three key differences. First, since $A = [A_1, \dots, A_p] = B_0^{-1}B$, prior beliefs about A may be correlated across equations if the prior beliefs about B_0 are. Second, there is no distinction between own lags and lags of other variables in a given equation, because in simultaneous equation models there is no unique normalization of the left-hand side variables. Third, the scale factors for the prior variances of the lag coefficients differ from those in the Minnesota prior.

It should be noted that the approach of Sims and Zha (1998) is specifically designed for structural VAR models with linear restrictions on B_0 under the assumption that the diagonal elements of B_0 are unity and that Σ_w is diagonal. Put differently, this approach cannot be used for models imposing identifying or overidentifying restrictions on B_0^{-1} . Canova and Pérez Forero

³ MATLAB code for the implementation for this method is provided at www.econ.yale.edu/~sims.

(2015) recently have generalized the approach of Sims and Zha by allowing for certain nonlinear identifying restrictions on B_0 . This allows them, for example, to accommodate restrictions on long-run responses of the type discussed in Chapters 10 and 11. It is not clear, however, whether their approach also handles direct restrictions on the elements of B_0^{-1} .

A generalization of the approach of Sims and Zha (1998) has recently been proposed by Baumeister and Hamilton (2015c). They allow for prior uncertainty about exclusion restrictions on B_0 . Rather than imposing that certain elements of B_0 are exactly zero, they postulate independent Student- t distributions for these elements that are centered on zero, allowing the user to account for identification uncertainty in conducting Bayesian analysis. Of course, this approach presumes that the users have extraneous information allowing them to parameterize the identification uncertainty in the form of a known density. Such information may not be readily available in practice.

9.6 Summary

This chapter reviewed estimation methods for structural VAR models with short-run restrictions on the matrix B_0 or B_0^{-1} . We stressed that there is a range of alternative estimation methods, some of which are designed for models with restrictions on B_0 , whereas others are designed for restrictions on B_0^{-1} . The most general estimators for models with identifying or overidentifying restrictions on both B_0 or B_0^{-1} are the Gaussian ML estimator and the GMM estimator. If the identifying or overidentifying restrictions are imposed on B_0 only, the GMM estimator may be computed in closed form. If there are restrictions on B_0^{-1} or on both B_0 and B_0^{-1} , then the GMM estimator must be computed in two steps. We first estimate Σ_u from the reduced-form representation of the VAR model, before expressing Σ_u in terms of the structural parameters of the model and solving for B_0 or B_0^{-1} numerically. If there are restrictions on B_0 only, the IV estimator for just- and overidentified models is another estimator that yields a closed-form solution. For just-identified models the method of moments can be used alternatively. Finally, for the special case of recursively identified models it is possible to estimate B_0^{-1} based on the lower-triangular Cholesky decomposition of the estimate of Σ_u .

We observed that it is also possible to estimate structural VAR models by Bayesian methods. The conventional approach of applying standard estimation methods for B_0 and/or B_0^{-1} to the posterior draws of Σ_u does not allow for overidentification. In contrast, Bayesian methods that impose priors directly on the structural form allow for both just-identifying and overidentifying restrictions, but only on B_0 . At present, they do not appear to allow for restrictions on B_0^{-1} except in the recursive case, because the triangularity of B_0^{-1} implies that B_0 is triangular.

10 Identification by Long-Run Restrictions

Finding enough short-run identifying restrictions can be a challenge in practice. One alternative idea has been to impose restrictions on the long-run response of variables to shocks. In the presence of unit roots in some variables, but not in others, this approach may allow us to identify at least some structural shocks. The promise of this alternative approach to identification is that it allows us to dispense with the controversy about what the right short-run restrictions are and to focus on long-run properties of models that most economists can more easily agree on. For example, it has been observed that most economists agree that demand shocks such as monetary policy shocks are neutral in the long run, whereas productivity shocks are not.

10.1 The Traditional Framework for Imposing Long-Run Restrictions

The idea of imposing long-run restrictions on structural VAR models was first proposed by Blanchard and Quah (1989) in the context of a bivariate model of the U.S. economy. It is useful to review their model for expository purposes. Blanchard and Quah's model attributes variation in U.S. real GDP and unemployment to an aggregate supply shock, w_t^{AS} , and an aggregate demand shock, w_t^{AD} . These structural shocks are identified by imposing that w_t^{AD} has no long-run effect on the level of real GDP. Let ur_t denote the U.S. unemployment rate and gdp_t the log of U.S. real GDP. Define

$$z_t = \begin{pmatrix} \Delta gdp_t \\ ur_t \end{pmatrix} \sim I(0),$$

where by assumption $z_t \sim I(0)$, but $gdp_t \sim I(1)$. The vector z_t is assumed to be generated by a reduced-form VAR process

$$A(L)z_t = u_t,$$

where $A(L) = I_2 - A_1L - \cdots - A_pL^p$, such that $A(1) = I_2 - A_1 - \cdots - A_p$, and $u_t \sim (0, \Sigma_u)$ is white noise. The corresponding structural form is

$$B(L)z_t = w_t,$$

where $B(L) = B_0 - B_1L - \cdots - B_pL^p = B_0A(L)$ and, hence, $B(1) = B_0 - B_1 - \cdots - B_p = B_0A(1)$. We impose the normalization $w_t = (w_t^{AS}, w_t^{AD})'$ $\sim (0, I_2)$. Given $w_t = B_0u_t$, it follows that $u_t = B_0^{-1}w_t$ and, hence, $\Sigma_u = B_0^{-1}B_0^{-1}$.

The effect of structural shocks, w_t , on the observed variables is obtained from the structural MA representation

$$z_t = B(L)^{-1}w_t = \Theta(L)w_t.$$

Because z_t is $I(0)$, the effect of any one structural shock on z_t will approach zero as the horizon increases. In other words, both Δgdp_t and ur_t by construction will return to their initial values eventually. This does not mean, however, that the level of real GDP will necessarily return to its initial value. The effect of a given structural shock on gdp_t is the cumulative sum of its effects on Δgdp_t . The long-run cumulative effects are summarized by the matrix $\Theta(1) = \sum_{i=0}^{\infty} \Theta_i = B(1)^{-1}$.

Requiring gdp_t to return to its initial level in the long run in response to an aggregate demand shock imposes an exclusion restriction on the upper right element of $\Theta(1)$ such that

$$\Theta(1) = \begin{bmatrix} \theta_{11}(1) & 0 \\ \theta_{21}(1) & \theta_{22}(1) \end{bmatrix}.$$

In contrast, $\theta_{11}(1)$ remains unrestricted, because aggregate supply shocks affect the level of real GDP in the long run. Moreover, there are no restrictions on the second row of $\Theta(1)$ because the cumulative responses of a stationary variable such as ur_t are clearly different from zero in general.

This example illustrates that if one structural shock is subject to a long-run restriction and the other one is not, it becomes possible to distinguish between these structural shocks. Using the relationship

$$\Theta(1) = B(1)^{-1} = A(1)^{-1}B_0^{-1},$$

it is seen that the exclusion restriction on $\Theta(1)$ is effectively an implicit restriction on B_0 because the reduced-form parameters $A(L)$ are given by the DGP. Thus, long-run restrictions provide an additional source of identifying restrictions for structural VAR models. Moreover, unlike in many of the models we reviewed in Chapter 8, B_0^{-1} (or equivalently B_0) is not restricted to be recursive. Although the choice of real GDP growth as the first variable in z_t is crucial for the interpretation of the Blanchard-Quah model, the choice of the second variable is not. In principle, any other stationary U.S. macroeconomic variable such as the U.S. capacity utilization rate would have done just as well from an econometric point of view.

Given knowledge of the reduced-form VAR model parameters, the unknown elements of $\Theta(1)$ may be recovered from

$$\begin{aligned}\Sigma_u &= B_0^{-1} B_0^{-1'} \\ &= [A(1)B(1)^{-1}] \underbrace{[A(1)B(1)^{-1}']}_{[B(1)^{-1}]' A(1')}.\end{aligned}$$

Pre-multiplying both sides by $A(1)^{-1}$ and post-multiplying both sides by $[A(1)^{-1}]' = [A(1)']^{-1}$ we obtain

$$\begin{aligned}A(1)^{-1} \Sigma_u [A(1)^{-1}]' &= A(1)^{-1} A(1) B(1)^{-1} [B(1)^{-1}]' A(1)' [A(1)']^{-1} \\ &= [B(1)^{-1}] [B(1)^{-1}]' \\ &= \Theta(1) \Theta(1)',\end{aligned}$$

where the left-hand side depends only on reduced-form parameters. Given the symmetry of Σ_u about its main diagonal, we need $K(K - 1)/2$ restrictions on $\Theta(1)$, where K is the number of variables in the model, to satisfy the order condition for exact identification of the parameters in $\Theta(1)$. This condition is satisfied by the exclusion restriction imposed by Blanchard and Quah (1989), allowing us to solve for the remaining elements of $\Theta(1)$. If the structure of $\Theta(1)$ is lower triangular, as in the Blanchard-Quah example, this may be accomplished by applying a lower-triangular Cholesky decomposition to $A(1)^{-1} \Sigma_u [A(1)^{-1}]' = \Phi(1) \Sigma_u \Phi(1)'$. Given knowledge of $\Theta(1)$ and $A(1)$, we can recover

$$B_0^{-1} = A(1) \Theta(1),$$

and, once we know B_0^{-1} , we can proceed with the further analysis of the VAR model exactly as in the case of short-run identifying restrictions.

Most applications of long-run restrictions involve a close variation on the theme of Blanchard and Quah (1989), in which the aggregate supply shock is interpreted as an aggregate productivity shock or as a technology shock with permanent effects on real output. Even if more variables are included in VAR models based on long-run restrictions, the focus often is on identifying the responses to aggregate productivity shocks only as opposed to other structural shocks. Such extensions are straightforward. If we augment the original Blanchard and Quah model to include additional stationary variables x_t such that

$$z_t = \begin{pmatrix} \Delta gdp_t \\ ur_t \\ x_t \end{pmatrix} \sim I(0),$$

for example, the aggregate supply shock may be identified by imposing a lower-triangular structure on $\Theta(1)$, as long as we are only interested in

identifying the responses to the aggregate supply shock. The higher-dimensional models in Galí (1999) are a good example for this approach. There are other examples of the use of long-run restrictions, however, that involve models that are fully identified, possibly in conjunction with other identifying restrictions (see Section 10.4).

In Section 10.2 we present a formal framework due to King, Plosser, Stock, and Watson (1991) for imposing long-run restrictions on the effects of structural shocks, followed by several empirical examples in Section 10.3. Although our analysis focuses on models in which all variables are $I(0)$ or $I(1)$, as is common in applied work, it should be kept in mind that the idea of long-run restrictions in structural VAR models may be generalized to processes that are integrated of higher order or fractionally integrated (see, e.g., break name differently, e.g., Tschernig, Weber, and Weigand 2013). Our framework is also general enough to accommodate the imposition of additional restrictions on the impact effects of structural shocks. Thus, it can be used to combine long-run and short-run identifying restrictions. For examples, the reader is referred to Section 10.4. In Section 10.5 we review the limitations of long-run identifying restrictions. A more detailed discussion of the estimation of structural VAR models subject to long-run restrictions can be found in Chapter 11.

10.2 A General Framework for Imposing Long-Run Restrictions

Whereas Blanchard and Quah (1989) restricted the cumulative response of the growth rate of real GDP in a stationary VAR representation, a more general framework for studying the long-run effects of structural shocks utilizes the vector error correction representation of the VAR model.

10.2.1 The Long-Run Multiplier Matrix

Suppose that at least some components of the K -dimensional $\text{VAR}(p)$ process y_t are $I(1)$. In Chapter 3 we demonstrated that in this situation the VECM,

$$\Delta y_t = \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \quad (10.2.1)$$

is a convenient reparameterization of the VAR process, where it is assumed that the cointegrating rank is r and that α and β are $K \times r$ matrices of rank r , the former being the loading matrix and the latter the cointegration matrix. The other symbols have their usual meaning. In particular, $u_t \sim (0, \Sigma_u)$ is the reduced-form white noise error term with nonsingular covariance Σ_u . At this point we do not consider deterministic terms because they are not relevant for structural identification.

In Chapter 3 we also showed that this process has the Granger representation

$$y_t = \Xi \sum_{i=1}^t u_i + \Xi^*(L)u_t + y_0^*, \quad (10.2.2)$$

where

$$\Xi = \beta_\perp \left[\alpha'_\perp \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_\perp \right]^{-1} \alpha'_\perp, \quad (10.2.3)$$

β_\perp and α_\perp are orthogonal complements of β and α , respectively,¹ $\Xi^*(L)u_t = \sum_{j=0}^{\infty} \Xi_j^* u_{t-j}$ is an $I(0)$ process, and y_0^* contains the initial values. The matrix Ξ has rank $K - r$, i.e., its rank equals the dimension of the process minus the cointegration rank, or, in other words, the rank of Ξ equals the number of common trends.

Since the structural shocks are obtained from the reduced-form errors by a linear transformation, $w_t = B_0 u_t$, we can replace u_t by $B_0^{-1} w_t$ in the Granger representation to obtain

$$y_t = \Upsilon \sum_{i=1}^t w_i + \Xi^*(L)B_0^{-1} w_t + y_0^*, \quad (10.2.4)$$

where $\Upsilon = [\zeta_{kl}] = \Xi B_0^{-1}$. This representation is useful because it directly shows the long-run effects or permanent effects of the structural shocks on the level of the variables y_t . The matrix Υ is also known as the matrix of long-run multipliers. Since the coefficient matrices $\Xi_j^* B_0^{-1}$ in the stationary term $\Xi^*(L)B_0^{-1} w_t$ taper off to zero, as $j \rightarrow \infty$, it is clear that the long-run effects of the structural shocks can be obtained from Υ . This matrix also has rank $K - r$, just as Ξ , because it is obtained from Ξ by a nonsingular transformation.

Restrictions on the long-run effects of the shocks can be imposed directly on Υ . If a shock does not have any long-run effects at all, the corresponding column in Υ is restricted to zero. Expressing the Blanchard-Quah example in this notation yields

$$y_t = \begin{pmatrix} gdp_t \\ ur_t \end{pmatrix} = \begin{bmatrix} \zeta_{11} & 0 \\ \zeta_{21} & \zeta_{22} \end{bmatrix} \sum_{i=1}^t \begin{pmatrix} w_i^{AS} \\ w_i^{AD} \end{pmatrix} + \dots,$$

where the stationary terms have been suppressed. The (1, 2) element of the Υ matrix is zero because the aggregate demand shock does not have a long-run effect on gdp_t . Note that we rearranged the original Blanchard-Quah model

¹ Recall that if M is an $m \times n$ matrix of full column rank, an orthogonal complement of M , denoted by M_\perp , is an $m \times (m - n)$ matrix with $\text{rk}(M_\perp) = m - n$ such that $M'M_\perp = 0_{m \times (m-n)}$. The orthogonal complement of a nonsingular square matrix is 0 and the orthogonal complement of a zero matrix is I_m .

such that even gdp_t is included in levels in y_t . Whereas y_t contains an $I(1)$ variable, the variable z_t in the introductory example of this chapter is $I(0)$.

Since ur_t is assumed to be $I(0)$, the second row of Υ must be zero as well, because no structural shock can have a nonzero long-run effect on a stationary variable. In other words, we obtain a long-run effects matrix

$$\Upsilon = \begin{bmatrix} \xi_{11} & 0 \\ 0 & 0 \end{bmatrix}.$$

Thus, the aggregate demand shock has no long-run effects at all. The element ξ_{11} , in contrast, must be nonzero because the rank of Υ is one if only one of the two variables is $I(0)$. In this case the cointegrating rank is one, as shown in Chapter 3, and, thus, $K - r = 1$ for this bivariate system.

More generally, whenever there are $I(0)$ components in y_t , the corresponding row of Υ is zero. In the special case of a stationary VAR process, the cointegrating rank is K and there are no common trends, i.e., $\Upsilon = 0_{K \times K}$. As mentioned earlier, shocks cannot have permanent effects on $I(0)$ variables, so this result makes sense.

In contrast, when all variables are $I(1)$ and there is no cointegration such that $r = 0$, then β and α are zero matrices and their orthogonal complements are simply $K \times K$ identity matrices. Thus,

$$\Xi = \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right)^{-1}$$

and

$$\Upsilon = \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right)^{-1} B_0^{-1}. \quad (10.2.5)$$

In that case, the first differences of y_t have a reduced-form $\text{VAR}(p-1)$ representation

$$\Delta y_t = \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t$$

and a structural-form representation

$$B(L) \Delta y_t = w_t,$$

where $B(L) = B_0 \Gamma(L) = B_0(I_K - \Gamma_1 L - \cdots - \Gamma_{p-1} L^{p-1})$. The structural impulse responses are obtained from the structural MA representation

$$\Delta y_t = B(L)^{-1} w_t = \Gamma(L)^{-1} B_0^{-1} w_t,$$

and the long-run effects on the levels y_t are the cumulated impulse responses $\Upsilon = \Gamma(1)^{-1} B_0^{-1}$ (see expression (10.2.5)).

10.2.2 Identification of Structural Shocks

Section 10.2.1 demonstrated that restrictions on the long-run effects of shocks can be placed directly on the $K \times K$ matrix Υ . Thus, we can identify some or potentially all shocks by restrictions on this matrix. For example, if a specific shock is known to have no long-run effect on a particular variable, a zero restriction can be placed on the corresponding element of Υ as in the Blanchard-Quah example.

In imposing restrictions on Υ , the properties of this matrix have to be taken into account. In particular, it is important to remember that the matrix has reduced rank $K - r$. An immediate implication of this property is that at most r shocks can have transitory effects only. In other words, at most r columns can be zero. This does not mean, however, that exactly r shocks must be purely transitory. In fact, all shocks can have permanent effects, because a $K \times K$ matrix of rank $K - r$ does not necessarily have zero elements. For example, in a bivariate system, the matrix

$$\Upsilon = \begin{bmatrix} 1 & \frac{1}{2} \\ 2 & 1 \end{bmatrix}$$

has rank one, although all elements are nonzero. We first consider the simpler case of $r = 0$, in which Υ has full rank K , before turning to the more difficult case, in which Υ has rank $0 < r < K$.

Unit Roots without Cointegration. One case that is particularly easy to handle arises when the cointegrating rank r is zero such that Δy_t is stationary and, thus, Υ is nonsingular. In that case, identification may be achieved, for example, by specifying a lower-triangular long-run effects matrix. The implied restrictions on B_0 are easy to impose by using the relationship (10.2.5). Defining $\Gamma(L) = I_K - \sum_{i=1}^{p-1} \Gamma_i L^i$, that relationship implies

$$\begin{aligned}\Upsilon\Upsilon' &= \Gamma(1)^{-1} B_0^{-1} B_0'^{-1} \Gamma(1)'^{-1} \\ &= \Gamma(1)^{-1} \Sigma_u \Gamma(1)'^{-1}.\end{aligned}$$

The latter expression is easy to compute from the reduced form. A lower-triangular Υ can then be obtained from the lower-triangular Cholesky decomposition of $\Upsilon\Upsilon'$ as

$$\Upsilon = \text{chol}(\Upsilon\Upsilon').$$

Hence,

$$B_0 = [\Gamma(1)\text{chol}(\Upsilon\Upsilon')]^{-1}.$$

This simple device for computing the restricted B_0 matrix and thereby the structural shocks has proved attractive for applied work. Applications of this

approach can be found in Binswanger (2004) and Lütkepohl and Velinov (2016).

Unit Roots with Cointegration. If $0 < r < K$, the analysis becomes more complicated because the rank of Υ must be taken into account when determining the number of restrictions that have to be imposed for full identification of the structural shocks. Recall that in the standard setup $K(K - 1)/2$ restrictions have to be imposed on B_0 or on the impact multiplier matrix B_0^{-1} to identify the structural shocks. In Chapter 8 we saw that in that context identification may be achieved by imposing a recursive structure such that B_0 and B_0^{-1} are lower triangular. In contrast, in the present context, simply counting the zero restrictions on the long-run multiplier matrix is not enough to ensure identification. Recursive restrictions on the matrix Υ alone will not achieve identification because of its reduced rank. Consider, for example, the 3×3 matrix

$$\Upsilon = \begin{bmatrix} \zeta_{11} & 0 & 0 \\ \zeta_{21} & \zeta_{22} & 0 \\ \zeta_{31} & \zeta_{32} & \zeta_{33} \end{bmatrix}.$$

Assuming that ζ_{33} is nonzero, this matrix can have rank 1 only if also $\zeta_{11} = \zeta_{21} = \zeta_{22} = 0$. Thus, Υ cannot be lower triangular with all elements below the diagonal nonzero. In fact, imposing the two zero restrictions in the last column implies that the first two elements in columns one and two also must be zero. Hence, they do not count as separate identifying restrictions. Besides, there may be $I(0)$ components in y_t that imply zero rows for Υ and do not serve as identifying restrictions. In other words, simply counting zero restrictions on the long-run multiplier matrix is not enough to ensure identification of the structural shocks. Only if the variables are not cointegrated, and hence Υ has full rank K , is it possible to identify all K shocks by a recursive structure on the long-run effects or, equivalently, by specifying Υ to be triangular.

It is useful to illustrate in more detail the nature of this problem. Recall that, if there is only one shock with purely transitory effects, this shock is identified as the complement to the shocks with permanent effects. If there are two or more shocks with only transitory effects, in contrast, these shocks must be identified by additional restrictions on B_0 or B_0^{-1} . Suppose a $K \times K$ matrix Υ has rank $K - r$ and there are exactly r purely transitory shocks with no long-run effects at all, where $r > 1$. Let the last r shocks be the transitory shocks. Then Υ has the form

$$\Upsilon = [\Upsilon_1, 0_{K \times r}].$$

In other words, the last r columns are zero. Clearly, the last r shocks must be distinguished by features other than their long-run effects. This can be accomplished, for example, by imposing restrictions on the last r columns of the

transformation matrix B_0 or its inverse. The first $K - r$ shocks that have permanent effects may be identified by restrictions on Υ_1 . This discussion highlights that occasionally it may be necessary to complement long-run identifying restrictions with short-run restrictions on B_0 or B_0^{-1} .

It is important to keep in mind, however, that, given the reduced form of the VAR model, all such restrictions jointly constrain the transformation matrix B_0 . Suppose that we have a set of linear restrictions

$$R_l \text{vec}(\Upsilon) = r_l \quad \text{or} \quad R_l \text{vec}(\Xi B_0^{-1}) = r_l$$

on the long-run multiplier matrix, where R_l is a suitable given restriction matrix and r_l a given fixed vector. These restrictions can be written as

$$R_l(I_K \otimes \Xi) \text{vec}(B_0^{-1}) = r_l \quad (10.2.6)$$

by using the rules for the vec operator and the Kronecker product. Because Ξ is fully determined by the reduced form (see expression (10.2.3)), it is not constrained by the structural identifying restrictions. In fact, expression (10.2.6) shows that the structural restrictions can be represented as linear restrictions on $\text{vec}(B_0^{-1})$,

$$R_L \text{vec}(B_0^{-1}) = r_l$$

with a restriction matrix $R_L = R_l(I_K \otimes \Xi)$.

If there are additional linear restrictions on the impact effects,

$$R_S \text{vec}(B_0^{-1}) = r_s,$$

then the long-run restrictions and the short-run restrictions can be combined as

$$\begin{bmatrix} R_L \\ R_S \end{bmatrix} \text{vec}(B_0^{-1}) = \begin{pmatrix} r_l \\ r_s \end{pmatrix}. \quad (10.2.7)$$

For a fully identified set of structural shocks there must be at least $K(K - 1)/2$ linearly independent restrictions, i.e., the restriction matrix

$$\begin{bmatrix} R_L \\ R_S \end{bmatrix}$$

must have a rank of at least $K(K - 1)/2$.

If overidentifying restrictions are considered, there are some further restrictions that have to be taken into account due to the reduced rank of the long-run multiplier matrix Υ . As pointed out earlier, in a model with cointegrating rank r , at most r of the structural innovations can have purely transitory effects and at least $K - r$ of them must have permanent effects. Lütkepohl (2008) shows that this fact limits the number of exclusion restrictions we can impose on B_0^{-1} . He proves that, under weak conditions, the number of admissible zero restrictions placed on columns of B_0^{-1} associated with transitory shocks cannot exceed $r - 1$. For example, if $r = 1$ and there is one transitory shock, as in

Blanchard and Quah (1989), there cannot be any zero restriction on the column of B_0^{-1} corresponding to the transitory shock. This result is intuitive, because in the bivariate model, the transitory shock is identified as the residual after identifying the permanent shock, so no further restrictions on B_0^{-1} are required. If $r = 2$ and $K = 3$, there can be at most one zero restriction on each of the columns of B_0^{-1} associated with the transitory shocks. If more zero restrictions are imposed, then Σ_u becomes singular. Such a singularity would contradict the premise that there must be as many shocks as variables in the structural VAR model. The same argument can also be invoked in reverse, by starting with the premise of r transitory shocks, which limits the number of zero restrictions that can be placed on the columns of the long-run structural multiplier matrix that are associated with the permanent shocks.

Although this discussion suggests that it can be difficult to combine short-run and long-run restrictions because the restriction accounting becomes more complicated, it must be kept in mind that overidentifying restrictions are the exception rather than the rule in structural VAR analysis. There are in fact situations in which the additional flexibility offered by long-run restrictions is helpful in achieving full identification. Also, in some cases a specific shock of interest may be particularly easy to identify in the present framework. For example, one may only be interested in a shock with permanent effects in a model where only one such shock is present.

10.3 Examples of Long-Run Restrictions

The approach in Section 10.2.2 requires expressing the VAR model as a VECM or, in the absence of cointegration, as a VAR model in first differences. This section presents a number of examples from the literature.

10.3.1 A Real Business Cycle Model with and without Nominal Variables

King, Plosser, Stock, and Watson (1991) apply the general framework considered in the previous section to the analysis of the Real Business Cycle (RBC) model. Their baseline model includes real output, real consumption, and real investment. An extended model includes in addition the aggregate price level, nominal interest rates, and money holdings.

The Baseline Three-Variable Model. It is useful to start with the baseline VAR model including only logged data for U.S. real output (gnp_t), real consumption (c_t), and real investment (inv_t). Unlike in Blanchard and Quah (1989), in this model all real variables are affected by the same productivity shock in the long run. Given a productivity shock with a stochastic trend, balanced growth under uncertainty implies that real consumption, real investment, and real output are

cointegrated such that $c_t - gnp_t \sim I(0)$ and $inv_t - gnp_t \sim I(0)$. This means that the VAR model for $y_t = (gnp_t, c_t, inv_t)'$ may equivalently be written as a reduced-form VECM as in the previous section with cointegration rank $r = 2$ and known cointegrating matrix β .

King et al. are interested in using this model to identify the responses to the common productivity shock. The remaining two transitory shocks remain economically unidentified. In this sense, the model is only partially identified. Assuming that there are indeed two transitory shocks and placing them last in the vector of shocks such that the first shock is the permanent shock, we obtain

$$\Upsilon = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix},$$

where $*$ denotes an unrestricted element. No further restrictions are necessary to identify the permanent shock.

It is possible to identify the remaining two transitory shocks in this model by combining the restrictions on the long-run multiplier matrix with short-run identifying restrictions. For local just-identification of all structural elements of B_0^{-1} , we need $K(K - 1)/2 = 3$ restrictions in this model. Since the long-run effects matrix Υ has rank $K - r = 1$, the two zero columns stand for two independent restrictions only. Clearly, the transitory shocks are not identified without further restrictions. One restriction on the last two columns of B_0^{-1} is sufficient to disentangle the two transitory shocks. For example, we may impose

$$\Upsilon = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix} \quad \text{and} \quad B_0^{-1} = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & * & * \end{bmatrix},$$

where $*$ indicates again that no restriction is imposed, which allows the first transitory shock (w_{2t}) to have instantaneous effects on all variables and prevents the second transitory shock (w_{3t}) from having an impact effect on real consumption. Whether such a restriction makes economic sense is a different matter. Indeed, in the baseline model of King et al. one would be hard-pressed to justify an additional exclusion restriction on B_0^{-1} , and the authors are content to focus on the effects of the balanced growth shock, w_{1t} . Which restriction is used to statistically identify the transitory shocks leaves the permanent shock and its effects unaffected. This means that we may simply impose an arbitrary exclusion restriction on the last two columns of B_0^{-1} if all we are interested in is the responses to the common productivity shock.

King, Plosser, Stock, and Watson (1991) investigate the ability of w_{1t} to explain the variability of the three model variables. They find impulse response patterns that are consistent with simple theoretical models in that all three variables increase in response to a positive balanced growth shock, but real output

and real investment respond more strongly than real consumption. Most of the adjustment is complete within four years. Structural forecast error variance decompositions indicate that 45–58% of the variability of real GNP growth at the short horizons is explained by the balanced-growth shock. This increases to 68% at the two-year horizon and 81% at the six-year horizon.

An equivalent way of writing the King et al. model is as a stationary VAR model for $z_t \equiv (\Delta gnp_t, c_t - gnp_t, inv_t - gnp_t)'$, where the balanced growth shock affects gnp_t in the long run, but not the stationary ratios $c_t - gnp_t$ and $inv_t - gnp_t$. This is, in essence, the representation chosen by Blanchard and Quah (1989). If we dropped the last variable in z_t , the VAR model for $(\Delta gnp_t, c_t - gnp_t)'$ could be analyzed using exactly the same approach used by Blanchard and Quah for $(\Delta gdp_t, ur_t)'$. In this sense, the analysis in King et al. may be viewed as a generalization of the approach in Blanchard and Quah (1989).

As mentioned earlier, the approach of relying on the stationary VAR representation rather than VECMs is also common when working with larger models. For example, Gali (1999) fits a stationary VAR model to $z_t \equiv (\Delta prod_t, \Delta h_t, \Delta m_t - \Delta p_t, i_t - \Delta p_t, \Delta^2 p_t)'$, where $prod_t$ denotes labor productivity, h_t stands for hours worked, m_t denotes money holdings, p_t the aggregate price level, and i_t the nominal interest rate. His identifying assumption is that only technology shocks have long-run effects on labor productivity. The four non-technology shocks in the model are not individually identified from an economic point of view. Thus, they can be identified arbitrarily from a statistical point of view. For example, one may restrict the accumulated effects matrix $\Theta(1)$ to be lower triangular. In this setup, as long as we are only interested in identifying the first structural shock, the ordering of the other shocks is inconsequential because the structural responses to the first shock will be invariant to the ordering of the identifying assumptions for the remaining variables.

In the baseline three-variable model of King, Plosser, Stock, and Watson (1991) this approach would involve specifying a VAR model for $z_t = (\Delta gnp_t, c_t - gnp_t, inv_t - gnp_t)' \sim I(0)$ and imposing

$$\Theta(1) = \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}.$$

The key difference is that in this case no further restrictions on B_0^{-1} are required because the restrictions on $\Theta(1)$ suffice to pin down the impact responses to the common productivity shock. However, the response of c_t to a productivity shock can only be constructed by cumulating the response of Δgnp_t and adding the implied response of gnp_t to that of $c_t - gnp_t$ (and similarly for inv_t). This discussion illustrates that the way we impose the identifying assumptions implied by a given economic model in estimation depends on how the structural VAR model is specified.

The Extended Six-Variable Model with Nominal Variables. The analysis becomes substantially more complicated once we allow for the possibility that there is more than one permanent shock. This situation is illustrated by the second example considered in King, Plosser, Stock, and Watson (1991). King et al. extend the baseline model such that $y_t = (gnp_t, c_t, inv_t, m_t - p_t, i_t, \Delta p_{t+1})'$, where i_t is the nominal interest rate, p_t is the log of the price level, and m_t the log of nominal money holdings. In this six-variable VAR system an additional cointegrating relationship arises that represents the money market equilibrium:

$$m_t - p_t - \beta_1 gnp_t + \beta_2 i_t \sim I(0), \quad (10.3.1)$$

where real balances ($m_t - p_t$), real output, and the nominal interest rate are assumed to be $I(1)$. At the same time, the balanced-growth paths must be allowed to depend on the real interest rate in recognition of the fact that growth theory predicts that higher real interest rates lower the share of output entering investment, while raising the share of output in consumption:

$$c_t - gnp_t - \phi_1(i_t - \Delta p_{t+1}) \sim I(0),$$

$$inv_t - gnp_t - \phi_2(i_t - \Delta p_{t+1}) \sim I(0).$$

Unlike in the baseline three-variable model, the consumption and investment shares are treated as $I(1)$ variables, as is the real interest rate.

Combining these results, we have three cointegrating relationships (and thus three common trends and three permanent shocks) in the model. The first permanent shock is the balanced-growth shock with long-run effects on real balances as well as real output, consumption, and investment; the second permanent shock is an inflation shock that affects the inflation rate and the nominal interest rate in the long run, but has no long-run effects on real output, consumption, or investment; and the third permanent shock is a real interest rate shock with long-run effects on the two ratios, the nominal interest rate, and real balances. These permanent shocks are assumed to be mutually uncorrelated as well as uncorrelated with the transitory shocks. As before, no attempt is made to identify the transitory shocks from an economic point of view.

Assuming that the three transitory shocks are placed last in the vector of shocks, the 6×6 long-run structural multiplier matrix takes the form $\Upsilon = [\Upsilon_1, 0_{6 \times 3}]$, where Υ_1 is restricted as

$$\Upsilon_1 = {}_{(6 \times 3)}^{\left[\begin{array}{ccc} * & 0 & 0 \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & * \\ * & * & 0 \end{array} \right]}.$$

Taking into account the results of a detailed cointegration analysis, King et al. decompose this matrix and impose further restrictions compatible with the rank of this matrix. They interpret the three permanent shocks as a balanced-growth shock (w_t^{growth}), a neutral inflation shock ($w_t^{\text{inflation}}$), and a real interest rate shock ($w_t^{\text{real interest}}$). The interpretation of the latter two shocks is not directly motivated based on economic models, and indeed these shocks are not structural in the sense discussed in Chapter 8. If only the balanced-growth shock is of interest, the identification of the other permanent shocks is not important, of course. In other words, the impulse responses to the first shock are not affected by the identifying restrictions for the second and third shocks.

Like in the three-variable model, constructing an estimate of the first column of B_0^{-1} requires additional ad hoc restrictions on the elements of the last three columns of B_0^{-1} . Given that we have already imposed three restrictions on Υ to identify $K - r$ shocks with permanent effects, we need to impose as many additional restrictions on the structural impact multiplier matrix as are required to identify the remaining $r = 3$ structural shocks. This may be accomplished, for example, by setting

$$B_0^{-1} = \begin{bmatrix} * & * & * & * & 0 & 0 \\ * & * & * & * & * & 0 \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \\ * & * & * & * & * & * \end{bmatrix}.$$

King et al. show that in the extended model the explanatory power of the balanced-growth shock for real output is substantially reduced. Much of the short-run variability in output and investment is associated with the permanent real interest rate shock. The permanent inflation shock explains little of the variation in the real variables.

10.3.2 A Model of Neutral and Investment-Specific Technology Shocks

Most models based on long-run restrictions identify only one permanent shock. Fisher (2006) considers a growth model with two permanent shocks. The motivation is that, from a theoretical point of view, conventional technology shocks that are neutral in that they affect the production of all goods homogeneously are not the only possible source of permanent effects on labor productivity. There also are investment-specific technology shocks that are embodied in capital. Omitting the latter shocks from the empirical analysis is likely to bias the VAR estimates of the response to neutral technology shocks. Hence, Fisher designs a model that incorporates both types of technology shocks.

Let $y_t = (p_t, prod_t, h_t)'$, where p_t is the log real price of investment goods, $prod_t$ is the log of labor productivity, and h_t is the log of per capita hours worked. The first two variables are treated as $I(1)$, whereas $h_t \sim I(0)$. Moreover, it is assumed that p_t and $prod_t$ are not cointegrated. Thus, $(\Delta p_t, \Delta prod_t, h_t)'$ is a stationary vector. The cointegrating rank of the model for y_t is $r = 1$ due to the inclusion of one $I(0)$ variable, so there must be at least two shocks with permanent effects.

The three structural shocks are a capital-embodied technology shock, w_t^{cet} , a labor productivity shock, w_t^{lp} , and a transitory shock, w_t^{trans} . They are ordered as $w_t = (w_t^{cet}, w_t^{lp}, w_t^{trans})'$. Clearly the last shock is identified by a zero column in the long-run effects matrix Υ . There are three assumptions for identifying the permanent shocks, which are explicitly derived from a real business cycle model. First, only capital-embodied technology shocks have a long-run effect on the log-level of the price of investment goods. Second, both neutral technology shocks and capital-embodied technology shocks have a long-run effect on the log-level of labor productivity. Third, shocks to investment-specific technology raise labor productivity in the long run by an amount that is proportionate to the amount by which they lower the log-level price of investment goods in the long run. The constant of proportionality is presumed known. The third assumption is not necessary for just-identification of the model, but serves as an overidentifying assumption.

These identifying restrictions can be imposed on the matrix of accumulated long-run effects $\Theta(1)$ for the $I(0)$ VAR model for $(\Delta p_t, \Delta prod_t, h_t)'$ as in Fisher (2006). This approach results in

$$\Theta(1) = \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}$$

with $\theta_{22}(1) = \alpha\theta_{21}(1)$, where $\theta_{ij}(1)$ denotes the i,j^{th} element of $\Theta(1)$ and α is known. Fisher (2006) imposes $\alpha = 1/3$. Fisher shows that the results are largely unaffected by the imposition of the overidentifying restriction. Alternatively, and equivalently, these restrictions can be expressed in terms of the structural long-run effects matrix:

$$\Upsilon = \begin{bmatrix} \zeta_{11} & 0 & 0 \\ \zeta_{21} & \alpha\zeta_{21} & 0 \\ 0 & 0 & 0 \end{bmatrix},$$

where the last row of zeros is due to h_t being $I(0)$ and, hence, none of the shocks moves hours permanently. Fisher's model may be augmented to include additional variables (see Fisher 2006).

10.3.3 A Model of Real and Nominal Exchange Rate Shocks

Although most applications of long-run restrictions focus on productivity shocks, there are other applications. For example, Enders and Lee (1997) propose a bivariate model of real and nominal exchange rates. They distinguish between real shocks that affect both real and nominal exchange rates equally in the long run and nominal shocks that affect only nominal exchange rates in the long run. Based on U.S. dollar exchange rates for Canada, Japan, and Germany since 1973, Enders and Lee find that the real shock explains much of the observed variability of real and nominal exchange rate movements. Their approach is formally identical to that in Blanchard and Quah (1989).

Let r_t denote the real exchange rate between the United States and a foreign country and e_t the corresponding nominal rate. Assuming that both variables are $I(1)$ but not cointegrated, we can set up a VAR model in first differences of the original variables and specify $z_t = (\Delta r_t, \Delta e_t)'$. The structural shocks are $w_t = (w_t^{\text{real}}, w_t^{\text{nominal}})'$. As before, we write

$$B_0^{-1} = A(1)\Theta(1),$$

where $\Theta(1)$ is the matrix of accumulated long-run effects of the model in first differences and identification is achieved by restricting the upper right element of $\Theta(1)$ to zero, consistent with the notion of the long-run neutrality of nominal shocks. This restriction suffices to identify both shocks. The precise type of the nominal and real shocks is left unspecified in the model. Enders and Lee (1997), however, make the case that the distinction between nominal and real shocks is consistent with a wide class of theoretical models including the well-known Dornbusch (1976) overshooting model.

An alternative way of imposing this identifying restriction on the matrix of accumulated long-run effects, $\Theta(1)$, would be to work with the framework of Section 10.2 for the levels variables $y_t = (r_t, e_t)'$. The matrix $\Theta(1)$ is identical to the structural long-run effects matrix Υ for the levels variables. Thus, using a VAR model for the levels variables $y_t = (r_t, e_t)'$, the structural shocks w_t are identified by one zero restriction on Υ ,

$$\Upsilon = \begin{bmatrix} * & 0 \\ * & * \end{bmatrix}.$$

10.3.4 A Model of Expectations about Future Productivity

An influential study by Beaudry and Portier (2006) focuses on the problem of capturing shifts in expectations about future productivity. They start with a bivariate model. Let $y_t = (tfp_t, sp_t)'$, where tfp_t denotes the log of quarterly total factor productivity and sp_t the log of the Standard & Poors 500 composite stock price index deflated by the quarterly GDP price deflator. Given that these

two variables appear cointegrated, Beaudry and Portier focus on the VECM representation of the VAR model for y_t .

They consider two identification schemes. In the first specification, they impose a recursive ordering on B_0^{-1} such that the second structural shock does not contemporaneously affect tfp_t , while both structural shocks are allowed to affect sp_t instantaneously. Put differently,

$$\begin{pmatrix} u_t^{tfp} \\ u_t^{sp} \end{pmatrix} = \begin{bmatrix} * & 0 \\ * & * \end{bmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix}.$$

This identifying assumption is consistent with the view that stock prices incorporate new information about productivity instantaneously and that stock prices anticipate increases in productivity that are yet to come. Beaudry and Portier interpret w_{2t} as a “news shock” by which they mean an anticipated change in future productivity.² In contrast, they interpret w_{1t} as an unanticipated productivity shock.

In the second specification, they impose instead a long-run identifying assumption. If we let $y_t = (tfp_t, sp_t)'$, this involves restricting the long-run multiplier matrix as

$$\Upsilon = \begin{bmatrix} * & 0 \\ * & * \end{bmatrix}.$$

In this alternative specification w_{1t} refers to a shock with long-run effects on tfp_t , whereas w_{2t} has no long-run effects on tfp_t . Equivalently, one could work with $z_t = (\Delta tfp_t, sp_t - tfp_t)'$ by analogy to Blanchard and Quah (1989).

Beaudry and Portier observe that the responses to w_{2t} in the first specification and to w_{1t} in the second specification appear very similar and that the respective shock series are almost perfectly correlated. They proceed to show that this finding is robust to augmenting the VAR model to include real consumption or real consumption and hours worked. Beaudry and Portier (2006) conclude from this evidence that these shocks are effectively the same shock, which implies that permanent changes in productivity growth are preceded by stock market booms. They also show that the observed impulse response patterns are qualitatively consistent with theoretical models in which technological innovation affects productivity with a delay.

This result has attracted considerable attention in the business cycle literature because it differs sharply from the conventional view of the business cycle being driven by unanticipated changes in total factor productivity. For example, it explains how booms and busts can happen absent large changes in fundamentals and why no technological regress is required to generate recessions.

² In Chapter 7 we noted that this “news shock” terminology is misleading and unrelated to the earlier literature on news shocks properly defined.

It also provides an alternative explanation for the observed comovement of macroeconomic aggregates. Indeed, estimates of VECMs including additional macroeconomic variables suggest that the “news shocks” identified by Beaudry and Portier are associated with increases in consumption, investment, output, and hours on impact and appear to constitute an important source of business cycle fluctuations. These findings have spurred the development of theoretical models capable of generating news-driven comovement among macroeconomic aggregates.

At the same time, there has been growing skepticism about the empirical approach used by Beaudry and Portier (2006). In particular, Kurmann and Mertens (2014) show that in the VECMs with more than two variables estimated by Beaudry and Portier, their identification scheme fails to determine news about total factor productivity. This point is important because the higher-dimensional VECMs are what allows Beaudry and Portier to quantify the business cycle effects of the “news shock”. Without that evidence the importance of these shocks remains unclear.

In the words of Kurmann and Mertens (2014), the identification problem arises from the interplay of two assumptions. First, Beaudry and Portier’s identification scheme imposes the restriction that one of the non-news shocks has no permanent impact on either TFP or consumption. Second, the VECMs estimated by Beaudry and Portier (2006) postulate that total factor productivity and consumption are cointegrated. As a result, total factor productivity and consumption have the same permanent component, which makes one of the two long-run restrictions redundant and leaves an infinite number of possible solutions with very different implications for the business cycle. The results reported in Beaudry and Portier (2006) represent just one arbitrary choice among these solutions, making it impossible to draw any conclusions about the role of news shocks.

More formally, consider the example of a VECM for

$$y_t = \begin{pmatrix} tfp_t \\ sp_t \\ c_t \end{pmatrix},$$

where all variables are in logs and c_t denotes consumption. Similar to their baseline model, Beaudry and Portier impose the short-run exclusion restriction that news shocks have no contemporaneous effect on tfp_t such that the (1,2) element of B_0^{-1} is zero:

$$B_0^{-1} = \begin{bmatrix} * & 0 & * \\ * & * & * \\ * & * & * \end{bmatrix}.$$

In addition, Beaudry and Portier impose two long-run exclusion restrictions. Note that all three variables share a common trend such that $K = 3$, $r = 2$, and $K - r = 1$. This means that the $K \times K$ long-run multiplier matrix Υ has rank $K - r = 1$. Beaudry and Portier (2006) impose the exclusion restriction that the third structural shock does not effect the level of tfp_t or the level of c_t in the long run such that the (1, 3) and (3, 3) element of Υ are zero. Given that Υ is of rank one, which means that all its rows are proportionate or all its columns are proportionate, we can conclude that either

$$\Upsilon = \begin{bmatrix} * & * & 0 \\ * & * & 0 \\ * & * & 0 \end{bmatrix} \quad \text{or} \quad \Upsilon = \begin{bmatrix} 0 & 0 & 0 \\ * & * & * \\ 0 & 0 & 0 \end{bmatrix},$$

where * denotes an unrestricted element. Both specifications are consistent with the identifying assumptions of Beaudry and Portier (2006). Although for the first of these alternative specifications of Υ all three shocks would be identified, this is not the case for the second specification. Given that the long-run restrictions on Υ in the latter case do not contribute to the identification of the structural shocks and given that there is only one restriction imposed on B_0^{-1} , the matrix B_0^{-1} is unidentified. Hence, without further identifying restrictions, the empirical analysis in Beaudry and Portier (2006) is uninformative.

As discussed in Kurmann and Mertens (2014), this problem cannot be addressed by simply not imposing the cointegrating restrictions in estimation, because this change in the model specification does not make the cointegration between tfp_t and c_t disappear, if such cointegration indeed exists in the data. If we imposed the absence of cointegration between tfp_t and c_t , however, disregarding the possible presence of cointegration, the shock implied by the remaining identifying restrictions would be largely unrelated to total factor productivity. In short, the approach employed by Beaudry and Portier (2006) is not informative about the question of interest.

10.4 Examples of Models Combining Long-Run and Short-Run Zero Restrictions

As the previous empirical example illustrated, in models with more than two variables it is common to combine long-run restrictions with short-run zero restrictions on B_0^{-1} or B_0 , allowing for additional shocks with transitory effects to be identified. This section provides some additional examples.

10.4.1 The IS-LM Model Revisited

A case in point is the IS-LM model of Galí (1992). Galí's objective is to further disentangle the effects of transitory money demand shocks, money supply

shocks, and shocks to the IS curve. His approach is to treat the textbook IS-LM model as a description of the interactions of the VAR model variables conditional on past data. Galí considers a quarterly model for $z_t = (\Delta gnp_t, \Delta i_t, i_t - \Delta p_t, \Delta m_t - \Delta p_t)'$. Movements in these macroeconomic variables are determined by four types of exogenous disturbances: aggregate supply shocks (w_t^{AS}), money supply shocks (w_t^{MS}), money demand shocks (w_t^{MD}), and shocks to the IS curve (w_t^{IS}). Thus, $w_t = (w_t^{AS}, w_t^{MS}, w_t^{MD}, w_t^{IS})'$. Ignoring the lagged dependent variables for expository purposes, the unrestricted structural VAR model can be written as

$$\begin{aligned}\Delta gnp_t &= -b_{12,0}\Delta i_t - b_{13,0}(i_t - \Delta p_t) - b_{14,0}(\Delta m_t - \Delta p_t) + w_t^{AS}, \\ \Delta i_t &= -b_{21,0}\Delta gnp_t - b_{23,0}(i_t - \Delta p_t) - b_{24,0}(\Delta m_t - \Delta p_t) + w_t^{MS}, \\ i_t - \Delta p_t &= -b_{31,0}\Delta gnp_t - b_{32,0}\Delta i_t - b_{34,0}(\Delta m_t - \Delta p_t) + w_t^{MD}, \\ \Delta m_t - \Delta p_t &= -b_{41,0}\Delta gnp_t - b_{42,0}\Delta i_t - b_{43,0}(i_t - \Delta p_t) + w_t^{IS},\end{aligned}$$

where $b_{ij,0}$ denotes the ij^{th} element of B_0 . Galí interprets the first equation as an aggregate supply function, the second equation as a money supply function, the third equation as the money demand function, and the last equation as an IS function. He imposes six identifying restrictions on the accumulated long-run effects of selected shocks and on B_0 . First, money supply shocks (w_t^{MS}), money demand shocks (w_t^{MD}), and IS shocks (w_t^{IS}) have no long-run effects on real GNP. Only aggregate supply shocks (w_t^{AS}) affect real GNP in the long run. This implies the restrictions $\theta_{12}(1) = \theta_{13}(1) = \theta_{14}(1) = 0$. Second, money demand shocks and money supply shocks do not have contemporaneous effects on output, which distinguishes them from IS shocks. This implies $b_{12,0} = 0$ and $b_{13,0} = 0$. Third, the monetary authority is assumed not to react contemporaneously to changes in the price level. This implies that $b_{23,0} + b_{24,0} = 0$, which imposes a linear restriction on B_0 . The restricted cumulated long-run effects $\Theta(1)$ and B_0 are of the form

$$\Theta(1) = \begin{bmatrix} * & 0 & 0 & 0 \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \quad \text{and} \quad B_0 = \begin{bmatrix} * & 0 & 0 & * \\ * & * & b_{23,0} & -b_{23,0} \\ * & * & * & * \\ * & * & * & * \end{bmatrix}.$$

Equivalently let $y_t = (gnp_t, i_t, i_t - \Delta p_t, m_t - p_t)'$. Then, using the notation of Section 10.2, Galí's restrictions can be expressed in terms of the following constraints on Υ and B_0 :

$$\Upsilon = \begin{bmatrix} * & 0 & 0 & 0 \\ * & * & * & * \\ 0 & 0 & 0 & 0 \\ * & * & * & * \end{bmatrix} \quad \text{and} \quad B_0 = \begin{bmatrix} * & 0 & 0 & * \\ * & * & b_{23,0} & -b_{23,0} \\ * & * & * & * \\ * & * & * & * \end{bmatrix},$$

where * indicates elements that are not explicitly restricted. The third row of zeros in Υ is due to the stationarity of the real interest rate.³

Having estimated the structural model, Gálí examines how well the model matches traditional Keynesian views. While the timing and magnitude of the structural impulse responses is largely consistent with the predictions of more elaborate New Keynesian models, a structural forecast error variance decomposition suggests that aggregate supply shocks have played a larger role in explaining economic fluctuations than traditional Keynesian views suggest. Historical decompositions indicate that recessions historically were caused by the coincidence of several adverse structural shocks of different types, with the mix of the adverse shocks varying considerably across recessions.

10.4.2 A Model of the Neoclassical Synthesis

A second example is Shapiro and Watson (1988). This study proposes a model of the U.S. economy that exploits insights from neoclassical economics about long-run behavior, while allowing for Keynesian explanations of short-run behavior. Unlike the preceding example, Shapiro and Watson do not take a stand on the economic model underlying the short-run behavior. Let h_t denote the log of hours worked, o_t the price of oil, gdp_t the log of real GDP, π_t inflation, and i_t the nominal interest rate. Shapiro and Watson decompose fluctuations in the $I(0)$ vector $z_t = (\Delta h_t, \Delta o_t, \Delta gdp_t, \Delta \pi_t, i_t - \pi_t)'$ in terms of labor supply shocks (w_t^{LS}), oil price shocks ($w_t^{\text{oil price}}$), technology shocks ($w_t^{\text{technology}}$), and two aggregate demand shocks ($w_t^{\text{AD-IS}}$ and $w_t^{\text{AD-LM}}$). The first identifying assumption is that aggregate demand shocks have no long-run effects on real GDP or hours worked. The second identifying assumption is that the long-run labor supply is exogenous, which allows Shapiro and Watson to separate the effects of shocks to technology from those to labor supply. The third identifying assumption is that exogenous oil price shocks have a permanent effect on the level of all $I(1)$ variables except hours worked. The two aggregate demand shocks may be interpreted as goods market (IS) and money market (LM) shocks. No effort is made to identify the two aggregate demand shocks separately.

Using our framework to state the restrictions formally, we use a vector of levels variables $y_t = (h_t, o_t, gdp_t, \pi_t, i_t - \pi_t)'$. All variables but the real interest rate are assumed to be $I(1)$ and not cointegrated, whereas $i_t - \pi_t \sim I(0)$. With this set of variables and the five shocks $w_t = (w_t^{\text{LS}}, w_t^{\text{oil price}}, w_t^{\text{technology}}, w_t^{\text{AD-IS}}, w_t^{\text{AD-LM}})'$ characterized earlier, we obtain the following matrix of

³ A critical discussion of the identifying assumptions imposed in this model is provided in Pagan and Pesaran (2008), who show that two out of the three short-run restrictions imposed by Gálí (1992) are not required when restrictions consistent with the cointegration properties of the variables are imposed.

long-run effects and B_0 :

$$\Upsilon = \begin{bmatrix} * & 0 & 0 & 0 & 0 \\ 0 & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & * \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B_0 = \begin{bmatrix} * & * & * & * & * \\ 0 & * & 0 & 0 & 0 \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix},$$

where * indicates that no explicit restriction is imposed. The last row of zeros in Υ is due to $i_t - \pi_t$ being $I(0)$. Note that the structure of Υ is not recursive. The additional short-run restrictions arise because the change in the price of oil is treated as exogenous white noise.

10.4.3 A U.S. Macroeconomic Model

Fisher, Huh, and Pagan (2016) stress the need for extending the traditional classification of shocks to include shocks with permanent effects on the level of at least one $I(1)$ variable and shocks with transitory effects on all variables. When there are additional $I(0)$ variables included in the VAR model, the corresponding additional shocks may have purely transitory effects on all variables or may have transitory effects on some variables and permanent effects on others. In the latter case, there will be more structural shocks with permanent effects on at least one of the $I(1)$ variables than suggested by the rank of Υ . The concern is that in this situation extra care is required to avoid some shocks in the model having unintended permanent effects. For example, nominal shocks may have unintended long-run effects on real variables and relative prices, unless the researcher is careful in specifying the structural model.

This point may be illustrated using the example of Peersman (2005). Peersman postulates a quarterly model of the U.S. economy based on a covariance stationary structural VAR model for the percent change in the nominal price of oil (Δo_t), real output growth (Δq_t), consumer price inflation (Δp_t), and the short-term nominal interest rate (i_t). The nominal interest is considered $I(0)$. The three $I(1)$ variables o_t , q_t , and p_t are expressed in first differences. Cointegration among the variables in levels is ruled out. In particular, the real price of oil is implicitly assumed to be $I(1)$. Because we want to study the model within the framework of Section 10.2, we consider the vector of levels variables $y_t = (o_t, q_t, p_t, i_t)$.

The vector of structural shocks includes a nominal oil price shock ($w_t^{\text{oil price}}$), a domestic aggregate supply shock (w_t^{AS}), a domestic aggregate demand shock (w_t^{AD}), and a domestic monetary policy shock ($w_t^{\text{monetary policy}}$). Identification involves two long-run exclusion restrictions and four contemporaneous exclusion restrictions. Neither aggregate demand shocks nor monetary policy shocks are allowed to affect the level of real output in the long run. Accounting also for the stationarity of the nominal interest rate (i_t), the

structural long-run effects matrix has the form

$$\Upsilon = \begin{bmatrix} * & * & * & * \\ * & * & 0 & 0 \\ * & * & * & * \\ 0 & 0 & 0 & 0 \end{bmatrix}. \quad (10.4.1)$$

The short-run restrictions arise from treating the oil price as predetermined with respect to all other variables, providing three exclusion restrictions, and from assuming monetary policy shocks not to have a contemporaneous effect on real output:

$$u_t = \begin{pmatrix} u_t^o \\ u_t^q \\ u_t^p \\ u_t^d \\ u_t^i \end{pmatrix} = B_0^{-1} w_t = \begin{bmatrix} * & 0 & 0 & 0 \\ * & * & * & 0 \\ * & * & * & * \\ * & * & * & * \end{bmatrix} \begin{pmatrix} w_t^{\text{oil price}} \\ w_t^{\text{AS}} \\ w_t^{\text{AD}} \\ w_t^{\text{monetary policy}} \end{pmatrix}.$$

Writing the long-run restrictions as in equation (10.4.1) reveals an obvious problem with this VAR model specification, which was first highlighted by Fisher, Huh, and Pagan (2016). In particular, the monetary policy shock in this model may have a permanent effect on the real price of oil because the nominal oil price may fall by more than the price level in the long-run, which is inconsistent with the maintained notion of long-run monetary neutrality. Likewise, aggregate demand shocks may have a long-run effect on the level of the real price of oil, invalidating the analysis in Peersman (2005).

To arrive at a VAR model with more economically defensible long-run properties, we may replace the nominal price of oil (o_t) by the real price of oil ($o_t - p_t$), allowing us to impose the required additional long-run restrictions. Specifically, we need to impose that neither aggregate demand nor monetary policy shocks affect the level of the real price of oil and the level of real output in the long run. In addition to these four long-run exclusion restrictions, we require two contemporaneous restrictions for exact identification, one to separate the aggregate demand shock from the monetary policy shock and the other to separate the real oil price shock from the aggregate supply shock. Fisher et al. impose the restrictions that U.S. aggregate demand and U.S. aggregate supply shocks have no contemporaneous effects on the real price of oil. In other words, in a model for $y_t = (o_t - p_t, q_t, p_t, i_t)$ their restrictions for the long-run effects and the impact effects are

$$\Upsilon = \begin{bmatrix} * & * & 0 & 0 \\ * & * & 0 & 0 \\ * & * & * & * \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad \text{and} \quad B_0^{-1} = \begin{bmatrix} * & 0 & 0 & * \\ * & * & * & * \\ * & * & * & * \\ * & * & * & * \end{bmatrix}.$$

Fisher, Huh, and Pagan (2016) show that price and output puzzles that were absent in the original specification of Peersman (2005) reemerge when the model is restricted to enforce the absence of unintended long-run effects.

10.5 Limitations of Long-Run Restrictions

There are a number of concerns related to using long-run restrictions for identifying structural VAR models. Our discussion in this section focuses on concerns with long-run restrictions that arise within the framework described in Section 10.2. Some of these concerns are of more general nature and arise in one form or another also in the context of short-run restrictions. Others are specific to the use of long-run identifying restrictions. Concerns specifically related to estimating structural VAR models with long-run restrictions that arise from the imprecision of estimates of the long-run impulse responses are addressed in Chapter 11, which focuses on the estimation of structural VAR models based on long-run restrictions.

10.5.1 Long-Run Restrictions Require Exact Unit Roots

One important limitation of the long-run identification schemes presented so far is that they require us to take a stand on the presence of exact unit roots in the autoregressive lag polynomial $A(L)$. This means that this alternative approach is more limited in scope than VAR models based on short-run restrictions, which remain valid regardless of the order of integration of the variables.

One alternative that allows for departures from the exact $I(1)$ hypothesis is to impose long-run identifying restrictions in structural VAR models of fractionally integrated variables (see Chapter 2). Tschernig, Weber, and Weigand (2013) propose an extension of the Granger representation for fractionally integrated variables of the form

$$y_t = \begin{bmatrix} \Delta_+^{-\delta_1} & 0 \\ 0 & \ddots & \Delta_+^{-\delta_K} \end{bmatrix} \boldsymbol{\Xi} B_0^{-1} \sum_{i=1}^t w_i + \begin{bmatrix} \Delta_+^{b-\delta_1} & 0 \\ 0 & \ddots & \Delta_+^{b-\delta_K} \end{bmatrix} \boldsymbol{\Xi}_+^*(L_b) B_0^{-1} w_t + y_{0t}^*,$$

where y_{0t}^* denotes the initial values, $\delta_1, \dots, \delta_K$ and b are real numbers, Δ^d stands for the fractional differencing operator defined as $\Delta^d = (1 - L)^d \equiv \sum_{i=0}^{\infty} (-1)^i \binom{d}{i} L^i$, and Δ_+^d denotes a truncated version of this expansion. Similarly, $\boldsymbol{\Xi}_+^*(L_b)$ is a truncated operator. The operator L_b denotes the fractional lag operator defined as $L_b \equiv 1 - \Delta^b$. The important point here is that, for certain

values of δ_j , restrictions may be placed on the long-run effects of the structural shocks on the model variables by restricting the matrix ΞB_0^{-1} .

Tschernig et al. investigate a bivariate system of U.S. log real GDP (gdp_t) and the log of its implicit deflator (p_t). They specify two shocks that can be interpreted as aggregate demand and aggregate supply shocks. The aggregate demand shock is identified as a shock having no persistent effect on gdp_t . This identifying assumption is imposed by restricting the upper right-hand element of ΞB_0^{-1} to zero. There are no restrictions on the long-run effects of the aggregate supply shock. The latter shock has persistent but not necessarily permanent effects on gdp_t .

While this approach provides an alternative to the exact unit root framework, it also involves additional complications. For example, it requires the user to assess the fractional and cofractional properties of the model variables. In addition, justifying long-run restrictions on economic grounds is likely to be more difficult in this framework than in the VECM framework.

10.5.2 Sensitivity to Omitted Variables

In low-dimensional models such as the bivariate model of Blanchard and Quah (1989) the aggregate demand and aggregate supply shocks must be viewed as aggregates of a larger number of demand and supply shocks (see Faust and Leeper 1997). For example, in reality there may be labor supply shocks and productivity shocks rather than just one aggregate supply shock. As Blanchard and Quah (1989) point out, this fact may invalidate the economic interpretation of their shock estimates. For example, even if none of the underlying demand shocks affect real output in the long run, the estimated aggregate demand shock in their model will represent a mixture of *both* the underlying demand and supply disturbances. Blanchard and Quah (1989) provide a theorem clarifying when this problem does not occur, but the conditions underlying that theorem are quite restrictive.

Faust and Leeper (1997) demonstrate that, in general, one cannot extract aggregate demand shocks in Blanchard and Quah's 1989 bivariate model that only involve the underlying demand shocks, nor can one extract aggregate supply shocks that only involve the underlying supply shocks, because each of these shocks involves different dynamic responses. The source of the problem is that the DGP has more shocks than the estimated model, and that each shock triggers different dynamic responses. One potential solution is to estimate larger VAR models that allow for more shocks (see Faust and Leeper 1997; Erceg, Guerrieri, and Gust 2005). Indeed, many applied researchers have augmented the Blanchard and Quah (1989) model to include additional variables (see, e.g., Galí 1999). There is evidence, however, that the results may be sensitive to the choice of the additional variables. This means that it is difficult to draw general lessons from estimates of models based on long-run restrictions. Note, however, that omitted variables also distort impulse

responses when short-run restrictions are used. In fact, much the same problem arises also in DSGE models. For example, an aggregate technology shock in an RBC model is merely a convenient fiction that obscures the fact that there are many potential sources of variation in aggregate technology with potentially different effects. The point of raising this issue in the context of structural VAR models with long-run restrictions is that it is important to be aware of the fact that this problem cannot be circumvented by using long-run restrictions.

10.5.3 Lack of Robustness at Lower Data Frequencies

It may seem that models based on long-run restrictions would apply equally regardless of the frequency of the data, making this approach particularly attractive when dealing with, say, quarterly or annual data, for which conventional short-run identifying assumptions are more difficult to justify. Faust and Leeper (1997) caution against this interpretation on the grounds that time aggregation will tend to invalidate the assumption of orthogonal structural shocks, even if that assumption applies at higher frequencies.

10.5.4 Nonuniqueness Problems without Additional Sign Restrictions

Estimates of the impulse responses in VAR models identified by short-run or long-run restrictions are identified only up to their sign. This fact matters both for the construction of impulse response point estimates and for the construction of simulated confidence intervals (see Chapter 12). In solving for the unknown elements of B_0^{-1} , it is typically implicitly assumed that the k^{th} shock has a positive effect on the k^{th} variable. There are situations in which such an assumption is natural. For example, in typical semistructural VAR models of monetary policy we would expect a contractionary monetary policy shock to be associated with a higher interest rate. Likewise, fully structural macroeconomic VAR models based on short-run restrictions can usually be written such that the k^{th} shock has a positive effect on the k^{th} variable (see Taylor 2004).

In other situations, the normalization can be less clear. This is especially true for models identified based on long-run restrictions. Consider, for example, the model of Blanchard and Quah (1989). Recall that $B_0^{-1} = \boldsymbol{\Gamma}(1)\boldsymbol{\Upsilon}$, where

$$\boldsymbol{\Upsilon} = \begin{bmatrix} * & 0 \\ * & * \end{bmatrix} = \text{chol}(\boldsymbol{\Gamma}(1)^{-1}\Sigma_u\boldsymbol{\Gamma}(1)^{-1})$$

and * denotes an unrestricted element. The solution $\boldsymbol{\Upsilon}$ is unique due to the uniqueness of the Cholesky decomposition. However, for any positive definite matrix Ω , one may reverse the signs of all elements in any column of $\text{chol}(\Omega)$ and still preserve $\Omega = \text{chol}(\Omega)\text{chol}(\Omega)'$. Thus, as Taylor (2004) notes, the solution for B_0^{-1} is identified only up to a transformation. Equivalent solutions may be obtained by multiplying any of the columns of B_0^{-1} by -1 , resulting in 2^K possible solutions, all of which satisfy $B_0^{-1}B_0^{-1'} = \Sigma_u$ (see also Lütkepohl

2013a). The practical effect of flipping the sign is to flip the structural impulse response functions in question about the horizontal axis. An obvious implication is that users of VAR models based on long-run restrictions need to make explicit which additional sign restrictions they are imposing for identification.

To obtain a unique solution, in practice, we need additional information from economic theory about the sign of the short-run or the long-run response. In some cases, this information is obvious. For example, Taylor (2004) discusses a bivariate VAR model of U.S. real GDP and its implicit price deflator, in which aggregate supply shocks have permanent effects on real GDP, but aggregate demand shocks do not. Of the $2^2 = 4$ possible solutions for B_0^{-1} , three can be immediately ruled out because we know that aggregate demand shocks move prices and quantities in the same direction on impact, while aggregate supply shocks move them in opposite directions. In other cases, identification may be less straightforward. For example, economic theory may not be informative about the sign of the impact response in question, in which case it is unclear how to proceed.

A case in point is the debate about the sign of the impact response of real output to a productivity shock, with some economists suggesting an initial decline and others an initial increase. Long-run restrictions were considered appealing in this context, precisely because of the perception that they leave short-run responses unrestricted. Without an additional normalization, however, models based on long-run restrictions cannot answer this question, and choosing the normalization based on the sign of the short-run response of real GDP simply amounts to assuming the answer. In this context, a more appealing strategy therefore is to normalize the sign of the response function based on the sign of the long-run response of real output. A similar concern also arises in the debate on the liquidity effect (see Section 8.5.3).

This problem is also important when conducting inference on models with long-run restrictions. In this case, an obvious concern is that, without an explicit sign assignment at each iteration, bootstrap replications of the model solution may correspond to different sign normalizations, invalidating inference (see Chapter 12). Moreover, Lütkepohl (2013a) observes that if an explicit sign assignment is carried out on an impact response coefficient that is close to zero, additional problems arise because the estimated sign need not coincide with the actual sign. For example, if one of the diagonal elements of B_0^{-1} is zero in population, and we normalize its estimate to be positive, then we will make the other elements flip their sign with positive probability regardless of their true value. This fact will inflate the width of bootstrap confidence intervals. Lütkepohl (2013a) illustrates this point in the context of the Blanchard and Quah (1989) model. In this model, the impact response of real output to an aggregate supply shock is close to zero. Normalizing the response on this coefficient dramatically inflates the width of bootstrap confidence intervals and changes the statistical significance of the impulse response estimates compared with normalizing on the unemployment response.

10.5.5 Sensitivity to Data Transformations

It has been observed that the conclusion from Blanchard-Quah type VAR models are sensitive to whether the second variable (e.g., hours worked) is entered in levels or differences. For example, specifying a VAR model with both hours worked and labor productivity in differences, Galí (1999) finds that hours worked initially drop after a positive technology shock, a finding that lends support to models with embedded frictions. On the other hand, Christiano, Eichenbaum, and Vigfusson (2004) provide support for the predictions of standard real business cycle (RBC) models, with hours worked rising immediately after a positive productivity shock, using the same long-run identification scheme, but allowing hours worked to enter the model in levels (see Section 11.5).

Gospodinov, Maynard, and Pesavento (2011) clarify the source of the extensive debate on the effect of technology shocks on unemployment/hours worked that ensued from these conflicting empirical results. They find that the contrasting conclusions from specifying the second VAR variable in levels as opposed to differences can be explained by small, but important, low-frequency co-movement between hours worked and labor productivity, which is allowed for in the levels specification but is implicitly set to zero in the differenced specification. Their theoretical analysis shows that, even when the root of hours is very close to one and the low-frequency co-movement is quite small, assuming away or explicitly removing the low-frequency component can have important implications for the long-run identifying restrictions, giving rise to biases large enough to account for the empirical difference between the two specifications. We defer a more formal discussion of this problem to Chapter 11. For a closely related analysis, see also Canova, López-Salido, and Michelacci (2010).

Which specification is right is ultimately an economic question and continues to be debated. For example, Fernald (2007) makes the case that the observed low-frequency correlation in the data is spurious and arises from breaks in both productivity and hours worked in the early 1970s and mid-1990s. Francis and Ramey (2009) instead attribute the observed correlation to common low-frequency trends in demographics and in public employment that are beyond the scope of the economic model. This view implies that the low-frequency component ought to be removed prior to the analysis, which leads to results that support the earlier findings of Galí (1999). On the other hand, if there is a true low-frequency correlation in the population model, as maintained by Christiano, Eichenbaum, and Vigfusson (2004), then any procedure that removes the low-frequency correlation between hours and productivity, whether by differencing, HP-filtering, or removing a deterministic time trend with breaks, will result in substantial bias in the estimates.

11 Estimation Subject to Long-Run Restrictions

This chapter illustrates the estimation of structural VAR models subject to long-run identifying restrictions. A variety of estimation methods have been proposed to estimate these models, including the method of moments, instrumental variable (IV) estimators, and full information maximum likelihood (FIML) estimators. We first focus on structural models based on long-run identifying restrictions only, followed by structural models that impose both long-run and short-run identifying restrictions.

11.1 Model Setup

Let y_t be a K -dimensional vector of variables that may be integrated of order 1 and possibly cointegrated. Consider the structural form

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t = B Y_{t-1} + w_t, \quad (11.1.1)$$

where $Y'_{t-1} \equiv (y'_{t-1}, \dots, y'_{t-p})$ and $B \equiv [B_1, \dots, B_p]$. Deterministic terms are neglected and the structural errors w_t are assumed to have a unit covariance matrix, $w_t \sim (0, I_K)$, unless stated otherwise. The corresponding reduced form is

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t = A Y_{t-1} + u_t, \quad (11.1.2)$$

where $A \equiv [A_1, \dots, A_p] = B_0^{-1} B$ and $u_t = B_0^{-1} w_t \sim (0, \Sigma_u = B_0^{-1} B_0^{-1})$.

Suppose that the structural model is identified by long-run restrictions. As discussed in Chapter 10, there are two alternative representations for estimating structural VAR models subject to long-run restrictions. If the $I(1)$ variables are not cointegrated or the cointegrating relations are known, then all $I(1)$ variables may be transformed to $I(0)$ variables by expressing them in first differences or as cointegration relations. As in the previous chapter, we denote these transformed variables as z_t . For the reduced-form and structural-form VAR representations of z_t we use the same notation as for y_t in equations (11.1.1) and (11.1.2). In other words, $z_t = A_1 z_{t-1} + \cdots + A_p z_{t-p} + u_t$ is the reduced-form

representation and $B_0 z_t = B_1 z_{t-1} + \cdots + B_p z_{t-p} + w_t$ denotes the structural form.

In that case, identification involves imposing long-run exclusion restrictions on the elements of the $K \times K$ cumulative structural impulse response matrix,

$$\Theta(1) = \sum_{i=0}^{\infty} \Theta_i = B(1)^{-1} = A(1)^{-1} B_0^{-1},$$

of the stationary process z_t , where

$$\begin{aligned}\Theta(L) &= \sum_{i=0}^{\infty} \Theta_i L^i \equiv B(L)^{-1} = (B_0 - B_1 L - \cdots - B_p L^p)^{-1} \\ &= A(L)^{-1} B_0^{-1} = (I_K - A_1 L - \cdots - A_p L^p)^{-1} B_0^{-1}\end{aligned}$$

and L is the lag operator, $A(L) \equiv I_K - A_1 L - \cdots - A_p L^p$, $B(L) \equiv B_0 - B_1 L - \cdots - B_p L^p$, $A(1)$ is $A(L)$ evaluated at $L = 1$ and $B(1)$ is $B(L)$ evaluated at $L = 1$.

An alternative and more general reduced-form representation of the VAR model (11.1.2) that also allows for unknown cointegrating relations is the VECM

$$\Delta y_t = \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t, \quad (11.1.3)$$

where α and β are $K \times r$ matrices of rank r . In the latter case, the long-run effects of the reduced-form shocks on the level of the model variables are given by the $K \times K$ matrix

$$\Xi = \beta_{\perp} \left[\alpha'_{\perp} \left(I_K - \sum_{i=1}^{p-1} \Gamma_i \right) \beta_{\perp} \right]^{-1} \alpha'_{\perp},$$

where β_{\perp} and α_{\perp} are orthogonal complements of β and α , respectively. As was shown in Chapter 10, the long-run effects of the structural shocks on the level variables in the VEC representation are given by the $K \times K$ matrix

$$\Upsilon = \Xi B_0^{-1}.$$

Thus, long-run restrictions may also be imposed on the elements of the long-run multiplier matrix Υ of the VECM. Since Ξ is determined by the reduced-form parameters of the VAR model, all constraints on Υ ultimately restrict B_0^{-1} and, hence B_0 .

For estimation purposes it is sometimes useful to consider a structural representation of the VECM, which is obtained by pre-multiplying the reduced form (11.1.3) by B_0 ,

$$B_0 \Delta y_t = \alpha' \beta' y_{t-1} + \Gamma_1' \Delta y_{t-1} + \cdots + \Gamma_{p-1}' \Delta y_{t-p+1} + w_t,$$

where $\alpha^\dagger = B_0\alpha$ and $\Gamma_i^\dagger = B_0\Gamma_i$, $i = 1, \dots, p - 1$. Note that in the error correction term this transformation only affects the loading coefficients, but not the cointegration relationships $\beta'y_{t-1}$. We are interested in the estimation of B_0 and/or B_0^{-1} . Obviously, any estimate of B_0 can be inverted to obtain an estimate of B_0^{-1} and vice versa.

11.2 Models Subject to Long-Run Restrictions Only

In this section, we focus on an illustrative example based on Galí (1999) who studies the relationship between technology shocks, employment, and aggregate economic fluctuations using quarterly U.S. data. The observables are the log of productivity (denoted by $prod_t$) and the log of total employee hours in nonagricultural establishments, averaged to quarterly frequency (denoted by h_t). The productivity variable is constructed by subtracting the log of hours from the log of real GDP. Let $y_t = (prod_t, h_t)'$ and suppose that both variables are $I(1)$ but not cointegrated. Galí postulates that y_t evolves according to a VAR(5) process with intercept.

Stationary Representation of the Model. The stationary VAR representation of Galí's model is a VAR(4) model with intercept for $z_t = \Delta y_t = (\Delta prod_t, \Delta h_t)'$,

$$z_t = v + A_1 z_{t-1} + \dots + A_4 z_{t-4} + u_t.$$

The sample period for the transformed data is 1947q2-1998q3. The unrestricted reduced-form LS estimates are

$$\begin{aligned}\widehat{A}_1 &= \begin{bmatrix} -0.1288 & -0.1283 \\ 0.2955 & 0.5809 \end{bmatrix}, \\ \widehat{A}_2 &= \begin{bmatrix} 0.0881 & -0.1258 \\ 0.1833 & -0.1060 \end{bmatrix}, \\ \widehat{A}_3 &= \begin{bmatrix} -0.0240 & -0.0464 \\ 0.1190 & 0.1545 \end{bmatrix}, \\ \widehat{A}_4 &= \begin{bmatrix} 0.0251 & -0.0697 \\ -0.0052 & -0.1112 \end{bmatrix},\end{aligned}$$

and

$$\widehat{\Sigma}_u = \begin{bmatrix} 0.4596 & -0.0469 \\ -0.0469 & 0.5343 \end{bmatrix}.$$

There is only one identifying restriction in the model of Galí (1999). The technology shock is identified by the assumption that only technology shocks have a long-run effect on the level of productivity. Thus, if the technology shock is the first structural shock, the matrix of structural long-run multipliers is lower

triangular such that

$$\Theta(1) = \left(I_2 - \sum_{i=1}^4 A_i \right)^{-1} B_0^{-1} = \begin{bmatrix} \theta_{11}(1) & 0 \\ \theta_{21}(1) & \theta_{22}(1) \end{bmatrix}.$$

The second structural shock is interpreted as a non-technology shock. The structure of this model mirrors that in Blanchard and Quah (1989).

Vector Error Correction Representation of the Model. Equivalently, this model can be written in the general reduced-form representation of the VAR model (11.1.2) as a VECM. In that case, we have

$$\Delta y_t = \nu + \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_4 \Delta y_{t-4} + u_t.$$

Given that both variables in y_t are $I(1)$ but not cointegrated, the cointegration rank is $r = 0$, so that $\alpha = \beta = 0$ and $\alpha_\perp = \beta_\perp = I_2$. Hence, the VECM reduces to a VAR model in first differences, $\Delta y_t = (\Delta \text{prod}_t, \Delta h_t)'$, and the VEC representation coincides with the stationary VAR representation for z_t .

The unrestricted reduced-form estimates are

$$\begin{aligned}\widehat{\Gamma}_1 &= \begin{bmatrix} -0.1288 & -0.1283 \\ 0.2955 & 0.5809 \end{bmatrix}, \\ \widehat{\Gamma}_2 &= \begin{bmatrix} 0.0881 & -0.1258 \\ 0.1833 & -0.1060 \end{bmatrix}, \\ \widehat{\Gamma}_3 &= \begin{bmatrix} -0.0240 & -0.0464 \\ 0.1190 & 0.1545 \end{bmatrix}, \\ \widehat{\Gamma}_4 &= \begin{bmatrix} 0.0251 & -0.0697 \\ -0.0052 & -0.1112 \end{bmatrix},\end{aligned}$$

and

$$\widehat{\Sigma}_u = \begin{bmatrix} 0.4596 & -0.0469 \\ -0.0469 & 0.5343 \end{bmatrix}.$$

In Galí's model $\alpha_\perp = \beta_\perp = I_2$, so that

$$\boldsymbol{\Xi} = \boldsymbol{\Gamma}(1)^{-1} = \left(I_2 - \sum_{i=1}^4 \boldsymbol{\Gamma}_i \right)^{-1}.$$

Thus, the identifying restriction that the non-technology shock has no long-run effects on the level of real GDP implies that

$$\boldsymbol{\Upsilon} = \boldsymbol{\Xi} B_0^{-1} = \left(I_2 - \sum_{i=1}^4 \boldsymbol{\Gamma}_i \right)^{-1} B_0^{-1} = \begin{bmatrix} \xi_{11} & 0 \\ \xi_{21} & \xi_{22} \end{bmatrix}$$

is lower triangular. In this particular case, $\Theta(1)$ coincides with $\boldsymbol{\Upsilon}$.

11.2.1 Method-of-Moments Estimation

Estimating the Model Using the Cholesky Decomposition

Stationary Representation. Recall from Chapter 10 that $\Theta(1)\Theta(1)' = A(1)^{-1}\Sigma_u A(1)^{-1}'$, where $A(1) = (I_2 - A_1 - A_2 - A_3 - A_4)$ and where we have imposed $\Sigma_w = I_2$. Given that $\Theta(1)$ has a lower-triangular structure, it can be estimated as

$$\begin{aligned}\widehat{\Theta}(1) &= \begin{bmatrix} \widehat{\theta}_{11}(1) & 0 \\ \widehat{\theta}_{21}(1) & \widehat{\theta}_{22}(1) \end{bmatrix} = \text{chol}(\widehat{A}(1)^{-1}\widehat{\Sigma}_u\widehat{A}(1)^{-1}) \\ &= \begin{bmatrix} 0.6157 & 0 \\ -0.2745 & 1.1125 \end{bmatrix},\end{aligned}$$

where

$$\widehat{A}(1)^{-1} = \begin{bmatrix} 0.6689 & -0.5141 \\ 0.8227 & 1.4435 \end{bmatrix}$$

and where chol denotes a function that returns the lower-triangular Cholesky decomposition. Suitable functions that implement this Cholesky decomposition are available in commonly used software.

Given estimates $\widehat{A}(1)$ and $\widehat{\Theta}(1)$, the structural impact multiplier matrix can be computed as

$$\widehat{B}_0^{-1} = \widehat{A}(1)\widehat{\Theta}(1) = \begin{bmatrix} 0.5384 & 0.4119 \\ -0.4971 & 0.5359 \end{bmatrix}.$$

Note that the use of long-run identifying restrictions allows \widehat{B}_0^{-1} to be nonrecursive. Given the estimate of B_0^{-1} , it is straightforward to compute the mutually uncorrelated structural residuals as

$$\widehat{w}_t = \widehat{B}_0 \widehat{u}_t$$

and to compute the implied structural impulse responses from the model in differences. All structural responses in Figure 11.1 have been cumulated and correspond to the response in the log-level. A positive technology shock is associated with a persistent upward shift in productivity and a decline in hours worked. A positive non-technology shock, in contrast, implies a short-lived increase in productivity and a positive response of hours worked.

Vector Error Correction Representation. Equivalently, this model may be estimated as a VECM. In this particular example, the VECM reduces to a VAR model in differences because $r = 0$. Note that

$$\begin{aligned}\Upsilon\Upsilon' &= \Gamma(1)^{-1}B_0^{-1}B_0'^{-1}\Gamma(1)'^{-1} \\ &= \Gamma(1)^{-1}\Sigma_u\Gamma(1)'^{-1},\end{aligned}$$

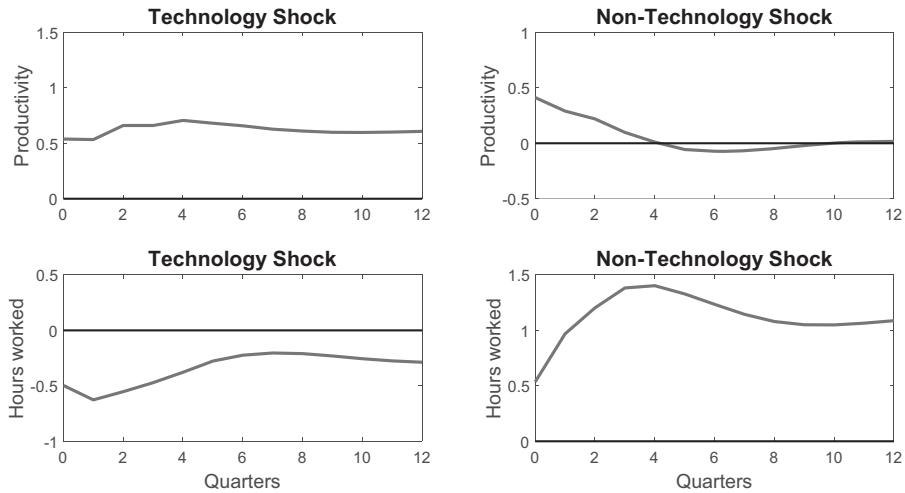


Figure 11.1. Responses to technology and non-technology shocks.

where Υ is lower triangular. Replacing the parameters in the latter expression by the reduced-form estimates and imposing $\Sigma_w = I_K$, an estimate of Υ can then be obtained from the lower-triangular Cholesky decomposition of

$$\widehat{\Upsilon}\widehat{\Upsilon}' \equiv \widehat{\Gamma}(1)^{-1} \widehat{\Sigma}_w \widehat{\Gamma}(1)'^{-1},$$

such that

$$\widehat{\Upsilon} = \text{chol}(\widehat{\Upsilon}\widehat{\Upsilon}').$$

The implied estimate of the structural impact multiplier matrix B_0^{-1} is given by

$$\widehat{B}_0^{-1} = \widehat{\Gamma}(1) \text{chol}(\widehat{\Upsilon}\widehat{\Upsilon}').$$

For our empirical example we obtain

$$\begin{aligned} \widehat{\Gamma}(1)^{-1} &= (I_2 - \widehat{\Gamma}_1 - \widehat{\Gamma}_2 - \widehat{\Gamma}_3 - \widehat{\Gamma}_4)^{-1} = \begin{bmatrix} 0.6689 & -0.5141 \\ 0.8227 & 1.4435 \end{bmatrix}, \\ \widehat{\Upsilon} &= \begin{bmatrix} 0.6157 & 0 \\ -0.2745 & 1.1125 \end{bmatrix}, \end{aligned}$$

and

$$\widehat{B}_0^{-1} = \begin{bmatrix} 0.5384 & 0.4119 \\ -0.4971 & 0.5359 \end{bmatrix},$$

which matches the estimate obtained using the stationary representation.

The approach of taking the Cholesky decomposition is computationally convenient, but cannot be employed unless the long-run multiplier matrix is

lower triangular. A case in point is the baseline three-variable model of King, Plosser, Stock, and Watson (1991), in which the structural long-run multiplier matrix takes the form

$$\Upsilon = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix},$$

where $*$ denotes an unrestricted element (see Section 10.3.1). Generalizations of the VECM estimation approach that can handle this situation are discussed in Section 11.3.

Estimating the Model Using a Nonlinear Equation Solver. An alternative approach is to solve the system of nonlinear equations implicitly defining the elements of B_0^{-1} using a nonlinear equation solver that finds the vector x such that $F(x) = 0$, where $F(x)$ is a system of nonlinear equations in x .¹ As before, we normalize $\Sigma_w = I_2$, while leaving the diagonal elements of B_0 unrestricted. We vectorize

$$B_0^{-1}B_0^{-1\prime} - \widehat{\Sigma}_u = 0$$

and impose the additional identifying restriction that $\theta_{12}(1) = 0$. The objective is to find the unknown elements of B_0^{-1} such that

$$\begin{bmatrix} \text{vech}(B_0^{-1}B_0^{-1\prime} - \widehat{\Sigma}_u) \\ \theta_{12}(1) \end{bmatrix} = 0, \quad (11.2.1)$$

given estimates of the reduced-form parameters $\widehat{\Sigma}_u$ and $\widehat{\Gamma}(1)^{-1}$ and an initial guess for B_0^{-1} such that $\Theta(1) = \widehat{\Gamma}(1)^{-1}B_0^{-1}$. The nonlinear equation solver iterates on expression (11.2.1) until convergence, given an initial guess for B_0^{-1} such that $\Theta(1) = \widehat{\Gamma}(1)^{-1}B_0^{-1}$. As noted in Chapter 9, the sign of the columns in \widehat{B}_0^{-1} is not unique and may flip, depending on how this numerical procedure is initialized. In practice, an additional normalization may be required. The estimate

$$\widehat{B}_0^{-1} = \begin{bmatrix} 0.5384 & 0.4119 \\ -0.4971 & 0.5359 \end{bmatrix}$$

obtained using this procedure exactly matches the earlier estimate based on $\widehat{\Theta}(1)$. Indeed, in this simple example, there is no advantage to using a nonlinear equation solver, given the recursive structure of $\Theta(1)$.

Under the alternative normalization that Σ_w is diagonal with positive elements on the diagonal, while the diagonal of B_0 is restricted to a vector of ones, we have $\Theta(1) = \Gamma(1)^{-1}B_0^{-1}\Sigma_w^{1/2}$, where $\Sigma_w^{1/2}$ is obtained by taking the square

¹ An example of such a nonlinear equation solver is *fso* in the MATLAB optimization toolbox.

root of the diagonal elements of Σ_w . Having directly imposed the zero restrictions on the off-diagonal elements of Σ_w and ones on the diagonal elements of B_0 , the objective is to find the unknown elements of B_0 and Σ_w such that

$$\begin{bmatrix} \text{vech}(B_0^{-1}\Sigma_w B_0^{-1\prime} - \widehat{\Sigma}_u) \\ \theta_{12}(1) \end{bmatrix} = 0, \quad (11.2.2)$$

given estimates of the reduced-form parameters and an initial guess for the off-diagonal elements of B_0 and for the diagonal elements of Σ_w .

Expression (11.2.2) may be iterated until convergence using a nonlinear equation solver, yielding $\widehat{\Theta}(1) = \widehat{\Gamma}(1)^{-1}\widehat{B}_0^{-1}\widehat{\Sigma}_w^{1/2}$. The implied estimate of the structural impact multiplier matrix, $\widehat{B}_0^{-1}\widehat{\Sigma}_w^{1/2}$, is the same as when using the Cholesky decomposition.

Estimating the Model Using the Algorithm of Rubio-Ramírez et al. (2010). Rubio-Ramírez, Waggoner, and Zha (2010) propose an algorithm for estimating exactly identified structural VAR models with possibly nonrecursive structure that tends to be computationally more efficient than using a nonlinear equation solver. This algorithm may be adapted easily to allow for the imposition of long-run restrictions. Let $\Sigma_w = I_K$. When imposing restrictions on the response of variable i at the infinite horizon, it is assumed that the i^{th} variable is expressed in first differences. The long-run structural impulse response of variable k to structural shock l corresponds to the element in row k and column l of the matrix

$$\Theta(1) = \left(I_K - \sum_{i=1}^p \boldsymbol{\Gamma}_i \right)^{-1} B_0^{-1},$$

which has K columns corresponding to the K structural shocks and K rows because we need to allow for restrictions on the long-run responses of any of the K variables. A candidate draw for $\Theta(1)$, denoted L_∞ , is constructed from the reduced-form estimates $\widehat{\boldsymbol{\Gamma}}_i$, given an initial guess of the structural impact multiplier matrix of $B_0^{-1} = \text{chol}(\widehat{\Sigma}_u)$, denoted by L_0 .

Zero restrictions on the long-run structural responses are represented by matrices Z_j for $1 \leq j \leq K$. The number of columns in Z_j is equal to the number of rows in L_∞ . If the rank of Z_j is z_j , then z_j is the number of zero restrictions associated with the j^{th} shock. The total number of zero restrictions is $z = \sum_{j=1}^K z_j$. The structural parameters satisfy the zero restrictions if and only if $Z_j L_\infty e_j = 0$ for $1 \leq j \leq K$, where e_j is the j^{th} column of I_K . In what follows, let Z_j represent the zero restrictions with the equations of the model ordered such that $z_j \leq K - j$, where $1 \leq j \leq K$. Observe that $Z_j L_\infty e_j = 0$ and $\bar{Z}_j L_\infty e_j = 0$ are equivalent statements, provided \bar{Z}_j exists, where \bar{Z}_j is defined by deleting all rows of zeros in Z_j .

An Algorithm for Solving Exactly Identified Models Based on Long-Run Restrictions.

1. Let $j = 1$.
2. If $j = 1$, then $Q'_j = \tilde{Z}_1 L_\infty$. For $j > 1$, form the matrix

$$Q'_j = \begin{bmatrix} \tilde{Z}_j L_\infty \\ q'_1 \\ \vdots \\ q'_{j-1} \end{bmatrix}.$$

3. There exists a vector q_j of unit length such that $Q'_j q_j = 0$. To find q_j such that $Q'_j q_j = 0$, use the QR decomposition $Q_j = \tilde{Q} R$, where \tilde{Q} is orthogonal and R is upper triangular. Choose q_j to be the last column of \tilde{Q} .
4. If $j = K$, stop. Otherwise set $j = j + 1$ and go to step 2.

If the model is exactly identified, this algorithm produces a $K \times K$ orthogonal rotation matrix

$$Q' = [q_1 \quad \cdots \quad q_K]$$

such that $L_0 Q'$ represents a solution for B_0^{-1} , given the estimates of the reduced-form VAR model and the identifying restrictions. As in the case of estimates based on nonlinear equation solvers, it may be necessary to normalize the sign of the columns of this solution.

Applying this algorithm to the example of Galí (1999), consider the candidate solution $L_0 = \text{chol}(\widehat{\Sigma}_u)$ and hence

$$L_\infty = \begin{bmatrix} 0.4890 & -0.3741 \\ 0.4579 & 1.0504 \end{bmatrix},$$

obtained as $L_\infty = (I_2 - \widehat{\Gamma}_1 - \widehat{\Gamma}_2 - \widehat{\Gamma}_3 - \widehat{\Gamma}_4)^{-1} \text{chol}(\widehat{\Sigma}_u)$. Let the first column refer to the long-run responses of the two model variables to the non-technology shock and the second column to the long-run responses to the technology shock.

Given that there is only one identifying restriction on the long-run response of productivity to the non-technology shock, the identifying restrictions can be expressed as

$$Z_1 = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad Z_2 = \begin{bmatrix} 0 & 0 \\ 0 & 0 \end{bmatrix}.$$

By deleting the rows of zeros from the Z_j s, we obtain

$$\bar{Z}_1 = [1 \quad 0].$$

Since all rows of Z_2 are zeros, there is no \bar{Z}_2 .

For $j = 1$, we obtain

$$Q'_1 = \bar{Z}_1 L_\infty = [0.4890 \quad -0.3741].$$

The QR decomposition for Q_1 yields

$$\tilde{Q} = \begin{bmatrix} -0.7942 & 0.6076 \\ 0.6076 & 0.7942 \end{bmatrix},$$

so $q'_1 = (0.6076, 0.7942)$, where q_1 by construction relates to the non-technology shock. For $j = 2$, we obtain

$$Q'_2 = q'_1 = [0.6076 \quad 0.7942]$$

because \bar{Z}_2 is the empty matrix. The QR decomposition for Q_2 yields

$$\tilde{Q} = \begin{bmatrix} -0.6076 & -0.7942 \\ -0.7942 & 0.6076 \end{bmatrix},$$

so $q'_2 = (-0.7942, 0.6076)$, where q_2 by construction relates to the technology shock.

Upon completing this subroutine, we obtain the following solution for the restricted rotation matrix:

$$Q' = [q_2 \quad q_1] = \begin{bmatrix} -0.7942 & 0.6076 \\ 0.6076 & 0.7942 \end{bmatrix},$$

where the ordering of q_1 and q_2 has been reversed to match the ordering of the structural shocks in the original model specification of Galí (1999). The implied solution for the structural impact multiplier matrix is

$$L_0 Q' = \begin{bmatrix} -0.5384 & 0.4119 \\ 0.4971 & 0.5359 \end{bmatrix}.$$

After flipping the sign of the first column such that positive technology shocks are associated with increases in productivity, we obtain the estimate

$$\widehat{B}_0^{-1} = \begin{bmatrix} 0.5384 & 0.4119 \\ -0.4971 & 0.5359 \end{bmatrix},$$

which matches the earlier solution. Of course, there is no advantage over the use of the Cholesky decomposition of the long-run variance-covariance matrix in the simple example considered here, but a similar algorithm could be used if the model is identified by both short-run and long-run restrictions or if the long-run variance covariance matrix does not have a lower-triangular representation, as illustrated in Section 11.3.

11.2.2 Full Information Maximum Likelihood Estimation

In this section we postulate that $\Sigma_w = I_K$, while the diagonal elements of B_0 remain unrestricted. If the errors are normally distributed, the reduced-form

VECM can be estimated using Johansen's ML method presented in Chapter 3. If the errors are not Gaussian, this method under suitable conditions can be interpreted as a quasi-ML method. The log-likelihood function is

$$\log l = \text{constant} - \frac{T}{2} \log(\det(\Sigma_u)) - \frac{1}{2} \sum_{t=1}^T u_t' \Sigma_u^{-1} u_t.$$

Substituting $\Sigma_u = B_0^{-1} B_0^{-1'}$ and concentrating out the intercept and slope parameters of the reduced form by replacing them by their ML estimates yields the concentrated log likelihood as a function of the structural parameters

$$\log l_c(B_0) = \text{constant} + \frac{T}{2} \log(\det(B_0)^2) - \frac{T}{2} \text{tr}(B_0' B_0 \tilde{\Sigma}_u), \quad (11.2.3)$$

where $\tilde{\Sigma}_u = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t'$ is the usual ML estimator of the reduced-form residual covariance matrix. This function has to be maximized with respect to B_0 subject to identifying restrictions on specific elements of $\Upsilon = \widehat{\Xi} B_0^{-1}$, where

$$\widehat{\Xi} = \widehat{\beta}_{\perp} \left[\widehat{\alpha}'_{\perp} \left(I_K - \sum_{i=1}^{p-1} \widehat{\Gamma}_i \right) \widehat{\beta}_{\perp} \right]^{-1} \widehat{\alpha}'_{\perp},$$

if $0 < r < K$ and $\widehat{\Xi} = \widehat{\Gamma}(1)^{-1}$ if $r = 0$. Maximizing the log likelihood may require iterative methods.

For example, for the Galí model $r = 0$ and the long-run restriction can be written as

$$\widehat{\gamma}(1)_{12} b_{11,0} + \widehat{\gamma}(1)_{22} b_{12,0} = 0,$$

where $\widehat{\gamma}(1)_{ij}$ denotes the i,j^{th} element of $\widehat{\Gamma}(1)$, where $b_{ij,0}$ is the i,j^{th} element of B_0 , and where we have exploited the fact that the triangularity of Υ is equivalent to $\Upsilon^{-1} = B_0 \Gamma(1)$ being triangular. In practice, the ML estimate of B_0 in the Galí model can be constructed as

$$\widetilde{B}_0 = \text{chol}(\widehat{\Gamma}(1) \tilde{\Sigma}_u^{-1} \widehat{\Gamma}(1)' \widehat{\Gamma}(1)^{-1}.$$

We obtain

$$\widetilde{B}_0 = \begin{bmatrix} 1.1115 & -0.8542 \\ 1.0308 & 1.1166 \end{bmatrix}.$$

Except for the degrees-of-freedom adjustment for $\tilde{\Sigma}_u$, this estimate equals the inverse of the estimate \widehat{B}_0^{-1} discussed earlier in this section.

11.2.3 Instrumental Variable Estimation

Pagan and Pesaran (2008) point out that in some cases models with cointegration relations and long-run restrictions may also be estimated by IV estimation

methods with the cointegration relations used as instruments. They consider a system with cointegrating rank r , $0 < r < K$, with exactly r structural shocks with purely transitory effects only and $K - r$ shocks with permanent effects. If the cointegration relations are known, the cointegration relations can be used as instruments for estimating some of the structural parameters. Suppose we place the permanent shocks first in the w_t vector and the transitory shocks last by partitioning w_t into $w_t = (w_t^p, w_t^{tr})'$, where w_t^p is $(K - r)$ -dimensional and w_t^{tr} is r -dimensional. Then the $K \times r$ matrix of loadings, α^\dagger , in the structural VECM has a $(K - r) \times r$ block of zeros,

$$\alpha^\dagger = \begin{bmatrix} 0_{(K-r) \times r} \\ \alpha_{(r \times r)}^\dagger \end{bmatrix}.$$

In other words, the cointegration relations do not appear in the first $K - r$ structural equations of the system. Partitioning the structural VECM,

$$B_0 \Delta y_t = \alpha^\dagger \beta' y_{t-1} + \Gamma_1^\dagger \Delta y_{t-1} + \cdots + \Gamma_{p-1}^\dagger \Delta y_{t-p+1} + w_t,$$

where the diagonal elements of B_0 have been normalized to unity, into the subset of equations that include an error correction term and the subset of equations that do not, yields

$$B_0^p \Delta y_t = \Gamma_1^{\dagger p} \Delta y_{t-1} + \cdots + \Gamma_{p-1}^{\dagger p} \Delta y_{t-p+1} + w_t^p, \quad (11.2.4)$$

$$B_0^{tr} \Delta y_t = \alpha_{(r \times r)}^\dagger \beta' y_{t-1} + \Gamma_1^{\dagger tr} \Delta y_{t-1} + \cdots + \Gamma_{p-1}^{\dagger tr} \Delta y_{t-p+1} + w_t^{tr}, \quad (11.2.5)$$

where w_t^p refers to shocks with permanent effects on the level of y_t and w_t^{tr} to shocks with purely transitory effects.

The objective is to estimate the unknown elements of B_0^p and B_0^{tr} . This involves instrumenting for the elements of Δy_t with unknown coefficients. We first estimate the set of equations (11.2.4). If the r cointegration relations $\beta' y_{t-1}$ are known, they can be used as instruments for the elements of Δy_t with unknown coefficients in equation (11.2.4). Instrumental variable estimation requires the instrument to be correlated with the variable to be instrumented, but uncorrelated with the error term of the regression to be estimated. The instruments $\beta' y_{t-1}$ satisfy the first condition because they are correlated with the lagged Δy_t and hence with the relevant elements of Δy_t . They satisfy the second condition because they are lagged and, hence, uncorrelated with w_t^p in equation (11.2.4). Having estimated equation (11.2.4), we next instrument for the elements of Δy_t with unknown coefficients in equation (11.2.5). For this purpose, the residuals \hat{w}_t^p from equation (11.2.4) are natural instruments. These instruments are valid because they are uncorrelated with w_t^{tr} .

In this setup, the diagonal elements of B_0 are usually normalized to one and the variances of the structural errors are left unrestricted. Hence, the

$(K - r) \times K$ matrix B_0^P has $(K - 1)(K - r)$ unknown elements that must be estimated. The r cointegration relations may not provide enough instruments to achieve full identification, and other restrictions may be required for estimating the structural parameters. One may, for example, use exclusion restrictions for the impact effects of the structural shocks. Estimation under joint restrictions on long-run and short-run effects is discussed in Section 11.3. A case in which enough instruments are available for estimating B_0^P arises when the cointegrating rank $r = K - 1$ and, hence, there is just one common trend and one permanent shock.

Pagan and Pesaran (2008) use the Blanchard-Quah model as an example. This example differs from the model in Galí (1999) in that the second variable in the level representation is $I(0)$ rather than $I(1)$. Recall from Chapter 10 that the Blanchard-Quah model consists of the log of U.S. real GDP (gdp_t) and the unemployment rate (ur_t). Thus, $y_t = (\text{gdp}_t, ur_t)'$. The first variable, gdp_t , is $I(1)$, whereas the second variable, ur_t , is $I(0)$. Thus, there is one trivial cointegration relation consisting of the second variable only. We write

$$\beta'y_t = [0, 1]y_t = ur_t.$$

Blanchard and Quah (1989) consider an aggregate supply shock (w_t^s) with permanent effects on gdp_t and a transitory aggregate demand shock (w_t^d). Using the insights from Pagan and Pesaran (2008), the two structural equations can be written as

$$\begin{aligned} \Delta\text{gdp}_t &= -b_{12,0}\Delta ur_t + \sum_{i=1}^{p-1} \gamma_{11,i}^\dagger \Delta\text{gdp}_{t-i} + \sum_{i=1}^{p-1} \gamma_{12,i}^\dagger \Delta ur_{t-i} + w_t^s, \\ \Delta ur_t &= -b_{21,0}\Delta\text{gdp}_t + \alpha^\dagger ur_{t-1} + \sum_{i=1}^{p-1} \gamma_{21,i}^\dagger \Delta\text{gdp}_{t-i} \\ &\quad + \sum_{i=1}^{p-1} \gamma_{22,i}^\dagger \Delta ur_{t-i} + w_t^d. \end{aligned}$$

In the first equation, ur_{t-1} can be used as an instrument for Δur_t in estimating $b_{12,0}$. The structural residual \widehat{w}_t^s can be used as an instrument for Δgdp_t in estimating $b_{21,0}$ in the second equation. Obviously, w_t^s is correlated with Δgdp_t and uncorrelated with the error term of the second equation by construction.

More generally, if the structural parameters in the first set of equations (11.2.4) can all be estimated, the corresponding residuals \widehat{w}_t^p can be used as instruments for estimating the structural parameters in B_0^{tr} , because the permanent shocks w_t^p are uncorrelated with the transitory shocks w_t^{tr} .

It is worth remembering that this IV method for estimating the structural parameters in the presence of long-run restrictions hinges on a set of important assumptions:

- The number of permanent shocks is equal to the number of common trends and the number of transitory shocks is equal to the cointegrating rank of the system. This condition is violated, for example, if the cointegration rank is greater than zero, but all structural shocks in the model have permanent effects (see Section 10.2.2).
- The cointegration relations are known (or super-consistently estimated).
- The r instruments provided by the cointegration relations suffice to estimate the structural parameters in the first $K - r$ equations.
- The permanent shocks suffice as instruments for estimating the structural parameters in the last r equations.

If these conditions are not satisfied, further identifying assumptions or instruments from other sources are needed.

11.3 Models Subject to Long-Run and Short-Run Restrictions

In this section, we consider two representative examples of structural VAR models that combine long-run and short-run identifying restrictions. The first example illustrates how these restrictions may be combined when estimating the structural model in its VAR representation. The second example illustrates how long-run and short-run restrictions may be imposed within the VECM framework.

11.3.1 Estimating the Model in VAR Representation

Consider a stylized VAR(4) model of U.S. monetary policy with only three quarterly variables. This example is based on Rubio-Ramírez, Waggoner, and Zha (2010). Let $z_t = (\Delta \text{gnp}_t, i_t, \Delta p_t)' \sim I(0)$, where gnp_t denotes the log of U.S. real GNP, p_t the corresponding GNP deflator in logs, and i_t the federal funds rate, averaged by quarter. The estimation period is restricted to 1954q4-2007q4 in order to exclude the recent period of unconventional monetary policy measures.

The Federal Reserve Board is assumed to control the interest rate by setting the policy innovation after observing the forecast errors for deflator inflation and real GNP growth. The model is fully identified and includes an aggregate demand shock and an aggregate supply shock in addition to the monetary policy shock. The monetary policy shock does not affect real GNP either within the current quarter or in the long run. The only shock to affect the log-level of real GNP in the long run is the aggregate supply shock. Defining $w_t = (w_t^{\text{policy}}, w_t^{\text{AD}}, w_t^{\text{AS}})'$, the identifying restrictions can be summarized

as

$$B_0^{-1} = \begin{bmatrix} 0 & * & * \\ * & * & * \\ * & * & * \end{bmatrix}$$

and

$$\Theta(1) = A(1)^{-1}B_0^{-1} = \begin{bmatrix} 0 & 0 & * \\ * & * & * \\ * & * & * \end{bmatrix}.$$

Since z_t is $I(0)$, the long-run restrictions are imposed on the cumulated impulse responses.

The unrestricted reduced-form estimates of a VAR(4) model with constant term are

$$\begin{aligned}\hat{A}_1 &= \begin{bmatrix} 0.2230 & 0.0097 & 0.3969 \\ 0.3147 & 1.0969 & 0.5979 \\ 0.0012 & 0.0636 & 0.4096 \end{bmatrix}, \\ \hat{A}_2 &= \begin{bmatrix} 0.2143 & -0.3862 & 0.1360 \\ 0.1867 & -0.4860 & 0.5037 \\ -0.0174 & -0.0510 & 0.2350 \end{bmatrix}, \\ \hat{A}_3 &= \begin{bmatrix} -0.0053 & 0.3407 & -0.5354 \\ 0.0275 & 0.4832 & -0.3212 \\ 0.0115 & -0.0052 & 0.0815 \end{bmatrix}, \\ \hat{A}_4 &= \begin{bmatrix} -0.0411 & 0.0013 & -0.0268 \\ -0.0226 & -0.1642 & -0.3320 \\ 0.0667 & -0.0137 & 0.2463 \end{bmatrix},\end{aligned}$$

and

$$\hat{\Sigma}_u = \begin{bmatrix} 0.6031 & 0.0795 & -0.0214 \\ 0.0795 & 0.6565 & 0.0375 \\ -0.0214 & 0.0375 & 0.0684 \end{bmatrix}.$$

Method-of-Moments Estimation

Estimating the Model Using a Nonlinear Equation Solver. In situations in which there are additional restrictions beyond the long-run identifying restriction, the structural VAR can no longer be estimated based on a lower-triangular Cholesky decomposition of the long-run variance-covariance matrix. However, estimates may still be constructed with the help of a nonlinear equation solver.

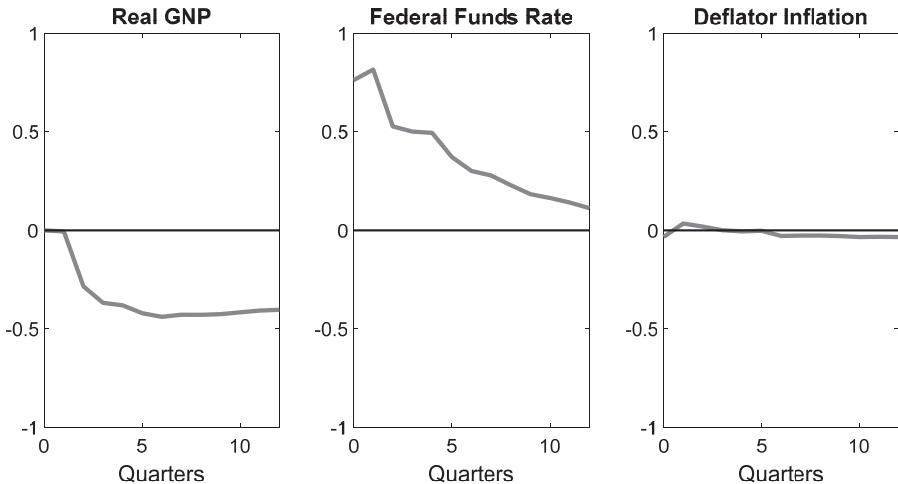


Figure 11.2. Responses to an unexpected U.S. monetary policy tightening.

The objective is to find the unknown elements of B_0^{-1} such that

$$\begin{bmatrix} \text{vech}(B_0^{-1} B_0^{-1\prime} - \widehat{\Sigma}_u) \\ b_0^{11} \\ \theta_{11}(1) \\ \theta_{12}(1) \end{bmatrix} = 0, \quad (11.3.1)$$

where the normalization $\Sigma_w = I_3$ has been imposed. The nonlinear equation solver iterates on expression (11.3.1) until convergence, given an initial guess for B_0^{-1} such that $\Theta(1) = \widehat{A}(1)^{-1} B_0^{-1}$, where $\widehat{A}(1)^{-1} = (I_3 - \sum_{i=1}^4 \widehat{A}_i)^{-1}$. The implied estimate of the structural impact multiplier matrix is

$$\widehat{B}_0^{-1} = \begin{bmatrix} 0 & 0.5845 & 0.5113 \\ 0.7625 & 0.2445 & -0.1239 \\ -0.0332 & 0.1491 & -0.2123 \end{bmatrix},$$

where the signs of the second column have been normalized such that a positive aggregate demand shock does not lower real GNP and inflation. Figure 11.2 shows the implied structural impulse response functions. An unexpected monetary policy tightening is associated with a persistent decline in real GNP but little response in inflation.

Estimating the Model as in Rubio-Ramírez et al. (2010). The same model may also be estimated using a variant of the algorithm proposed in Rubio-Ramírez, Waggoner, and Zha (2010). As before, $\Sigma_w = I_K$. When imposing restrictions on the infinite horizon, it is assumed that the i^{th} variable is expressed in first differences. The long-run structural impulse responses

correspond to the element in row i and column j of the matrix

$$\Theta(1) = \left(I_K - \sum_{i=1}^p A_i \right)^{-1} B_0^{-1}.$$

A candidate draw for $\Theta(1)$, denoted L_∞ , is constructed from the reduced-form estimates \widehat{A}_i , given an initial guess of the structural impact multiplier matrix of $B_0^{-1} = \text{chol}(\widehat{\Sigma}_u)$, denoted by L_0 .

It is convenient to stack the structural impulse response functions into a single matrix denoted by \mathbf{L} . If restrictions are imposed at horizons zero and infinity, then

$$\mathbf{L} = \begin{bmatrix} L_0 \\ L_\infty \end{bmatrix}$$

has K columns corresponding to the K structural shocks and $2K$ rows because we consider restrictions at two horizons for K variables.

Zero restrictions on the structural responses can be represented by matrices Z_j for $1 \leq j \leq K$. The number of columns in Z_j is equal to the number of rows in \mathbf{L} . If the rank of Z_j is z_j , then z_j is the number of zero restrictions associated with the j^{th} shock. The total number of zero restrictions is $z = \sum_{j=1}^K z_j$. The structural parameters satisfy the zero restrictions if and only if $Z_j \mathbf{L} e_j = 0$ for $1 \leq j \leq K$, where e_j is the j^{th} column of I_K . In what follows, let Z_j represent the zero restrictions with the equations of the VAR model ordered such that $z_j \leq K - j$, where $1 \leq j \leq K$. Observe that $Z_j \mathbf{L} e_j = 0$ and $\bar{Z}_j \mathbf{L} e_j = 0$ are equivalent statements, provided \bar{Z}_j exists, where \bar{Z}_j is defined by deleting all rows of zeros in Z_j .

An Algorithm for Solving Models That Are Exactly Identified by a Combination of Short-Run and Long-Run Restrictions.

1. Let $j = 1$.
2. If $j = 1$, then $Q'_j = \bar{Z}_1 \mathbf{L}$. For $j > 1$, form the matrix

$$Q'_j = \begin{bmatrix} \bar{Z}_j \mathbf{L} \\ q'_1 \\ \vdots \\ q'_{j-1} \end{bmatrix}.$$

3. There exists a vector q_j of unit length such that $Q'_j q_j = 0$. To find q_j such that $Q'_j q_j = 0$, use the QR decomposition $Q_j = \tilde{Q} R$, where \tilde{Q} is orthogonal and R is upper triangular. Choose q_j to be the last column of \tilde{Q} .
4. If $j = K$, stop. Otherwise, set $j = j + 1$ and go to step 2.

If the model is exactly identified, this algorithm produces a $K \times K$ orthogonal rotation matrix

$$Q' = [q_1 \quad \cdots \quad q_K]$$

such that $L_0 Q'$ represents a solution for B_0^{-1} , given the estimates of the reduced-form VAR model and the identifying restrictions. As in the case of estimates based on nonlinear equation solvers, it may be necessary to normalize the sign of the columns of this solution.

Returning to the example of the monetary policy VAR model, we need to impose restrictions on the structural impulse response functions at horizons 0 and ∞ . We first define a candidate solution for the structural impulse response functions of interest as

$$\mathbf{L} = \begin{bmatrix} L_0 \\ L_\infty \end{bmatrix} = \begin{bmatrix} 0.7766 & 0 & 0 \\ 0.1024 & 0.8037 & 0 \\ -0.0276 & 0.0501 & 0.2552 \\ 0.9396 & -0.3398 & -1.0189 \\ 6.2126 & 6.3718 & 15.0626 \\ -0.3180 & -0.4124 & 3.5063 \end{bmatrix}.$$

The exclusion restrictions for the monetary policy shock and the aggregate demand shock can be written as

$$Z_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$Z_2 = \begin{bmatrix} 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad Z_3 = 0_{3 \times 6},$$

where Z_1 embodies the restriction that w_t^{policy} does not affect real GDP in the short run or in the long run, and Z_2 embodies the restriction that w_t^{AD} does not affect real GDP in the long run. By deleting the rows of zeros from the Z_j s, we obtain

$$\bar{Z}_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 \end{bmatrix},$$

$$\bar{Z}_2 = [0 \quad 0 \quad 0 \quad 1 \quad 0 \quad 0].$$

Since all rows of Z_3 are zeros, there is no \bar{Z}_3 .

For $j = 1$, we obtain

$$Q' = \bar{Z}_1 \mathbf{L} = \begin{bmatrix} 0.7766 & 0 & 0 \\ 0.9396 & -0.3398 & -1.0189 \end{bmatrix}.$$

The QR decomposition for Q_1 yields

$$\tilde{Q} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & -0.3163 & -0.9487 \\ 0 & -0.9487 & 0.3163 \end{bmatrix},$$

so $q'_1 = (0, -0.9487, 0.3163)$. For $j = 2$, we obtain

$$Q'_2 = \begin{bmatrix} \tilde{Z}_2 \mathbf{L} \\ q'_1 \end{bmatrix} = \begin{bmatrix} 0.9396 & -0.3398 & -1.0189 \\ 0 & -0.9487 & 0.3163 \end{bmatrix}.$$

Applying the QR decomposition to Q_2 as before yields an orthogonal matrix, the last column of which is $q_2 = (0.7527, 0.2083, 0.6246)'$. Finally, we form

$$Q'_3 = \begin{bmatrix} q'_1 \\ q'_2 \end{bmatrix} = \begin{bmatrix} 0 & -0.9487 & 0.3163 \\ 0.7527 & 0.2083 & 0.6246 \end{bmatrix}$$

because \tilde{Z}_3 is the empty matrix. The last column of the orthogonal matrix obtained by applying the QR decomposition to Q_3 yields $q_3 = (-0.6584, 0.2381, 0.7140)'$.

Upon completing this subroutine, we obtain the following solution for the restricted rotation matrix:

$$Q' = [q_1 \ q_2 \ q_3] = \begin{bmatrix} 0 & 0.7527 & -0.6584 \\ -0.9487 & 0.2083 & 0.2381 \\ 0.3163 & 0.6246 & 0.7140 \end{bmatrix}.$$

The implied solution for the structural impact multiplier matrix is

$$L_0 Q' = \begin{bmatrix} 0 & 0.5845 & -0.5113 \\ -0.7625 & 0.2445 & 0.1239 \\ 0.0332 & 0.1491 & 0.2123 \end{bmatrix}.$$

After flipping the signs of the first column such that a positive interest rate innovation raises the interest rate and flipping the sign of the last column such that a positive aggregate supply shock is not associated with lower real GNP on impact, we obtain the estimate

$$\widehat{B}_0^{-1} = \begin{bmatrix} 0 & 0.5845 & 0.5113 \\ 0.7625 & 0.2445 & -0.1239 \\ -0.0332 & 0.1491 & -0.2123 \end{bmatrix},$$

which matches the estimate obtained using the nonlinear equation solver.

Other Estimation Methods. It is not possible to estimate our example model by IV methods because the short-run restriction is imposed on B_0^{-1} rather than B_0 , as would be required for IV estimation. In contrast, FIML estimation is straightforward. Because the VECM representation of the structural VAR

model accommodates more general models, the reader is referred to our discussion of the FIML estimator in the following section.

11.3.2 Estimating the Model in VECM Representation

Recall from Section 10.3.1 that King, Plosser, Stock, and Watson (1991) considered a baseline quarterly VAR(2) model in levels for the log of U.S. real GNP (gnp_t), the log of real consumption (c_t), and the log of real investment (inv_t). The estimation period is 1947q1–1988q4. In this model all variables are affected by the same productivity shock in the long run. This means that the VAR model for $y_t = (gnp_t, c_t, inv_t)'$ may equivalently be written as a reduced-form VECM as in the previous section with known cointegration rank $r = 2$ and unknown cointegrating matrix β :

$$\Delta y_t = \nu + \alpha \beta' y_{t-1} + \Gamma_1 \Delta y_{t-1} + u_t.$$

Let $\Sigma_w = I_3$. King et al. are interested in using this model to identify the responses to the common productivity shock. Assuming that there are two transitory shocks that are placed last in the vector of structural shocks, we obtain

$$\Upsilon = \begin{bmatrix} * & 0 & 0 \\ * & 0 & 0 \\ * & 0 & 0 \end{bmatrix},$$

where $*$ denotes an unrestricted element. No further restrictions are necessary to identify the permanent shock.

If we are only interested in quantifying the effects of the common productivity shock, we may, without loss of generality, impose an arbitrary exclusion restriction on the last two columns of B_0^{-1} . Without loss of generality, we choose to restrict b_0^{23} to zero for expository purposes such that

$$B_0^{-1} = \begin{bmatrix} * & * & * \\ * & * & 0 \\ * & * & * \end{bmatrix},$$

as in Lütkepohl (2005, chapter 9).

The reduced-form ML estimates of the VECM obtained by the Johansen method are

$$\tilde{\nu} = \begin{bmatrix} -0.009 \\ -0.028 \\ -0.301 \end{bmatrix},$$

$$\tilde{\alpha} = \begin{bmatrix} -0.225 & 0.204 \\ -0.062 & 0.072 \\ -0.112 & 0.255 \end{bmatrix},$$

$$\tilde{\beta}' = \begin{bmatrix} 1 & 0 & -1.020 \\ 0 & 1 & -1.099 \end{bmatrix},$$

$$\tilde{\Gamma}_1 = \begin{bmatrix} 0.123 & 0.090 & 0.159 \\ 0.208 & -0.207 & 0.025 \\ 0.703 & -0.169 & 0.331 \end{bmatrix},$$

and

$$\tilde{\Sigma}_u \times 1000 = \begin{bmatrix} 0.1259 & 0.0395 & 0.1580 \\ 0.0395 & 0.0547 & 0.0684 \\ 0.1580 & 0.0684 & 0.4897 \end{bmatrix},$$

where the cointegrating vectors in $\tilde{\beta}'$ have been normalized for expository purposes.

Method-of-Moments Estimation. As in the earlier example, we proceed in two steps. Having estimated the reduced form of the VECM, the objective is to find the unknown elements of B_0^{-1} such that

$$\begin{bmatrix} \text{vech}(B_0^{-1}B_0^{-1\prime} - \tilde{\Sigma}_u) \\ b_0^{23} \\ \zeta_{12} \\ \zeta_{13} \\ \zeta_{22} \\ \zeta_{23} \\ \zeta_{32} \\ \zeta_{33} \end{bmatrix} = 0, \quad (11.3.2)$$

where ζ_{ij} is the ij^{th} element of Υ . The nonlinear equation solver iterates on expression (11.3.2), given an initial guess for B_0^{-1} such that $\Upsilon = \tilde{\Xi}B_0^{-1}$, where

$$\tilde{\Xi} = \tilde{\beta}_{\perp} \left[\tilde{\alpha}'_{\perp} \left(I_K - \sum_{i=1}^{p-1} \tilde{\Gamma}_i \right) \tilde{\beta}_{\perp} \right]^{-1} \tilde{\alpha}'_{\perp},$$

which in this example reduces to

$$\tilde{\Xi} = \tilde{\beta}_{\perp} \left[\tilde{\alpha}'_{\perp} (I_K - \tilde{\Gamma}_1) \tilde{\beta}_{\perp} \right]^{-1} \tilde{\alpha}'_{\perp}.$$

Here $\tilde{\beta}_{\perp}$ and $\tilde{\alpha}_{\perp}$ are orthogonal complements of $\tilde{\beta}$ and $\tilde{\alpha}$, respectively. For a $K \times r$ matrix β of rank r the orthogonal complement β_{\perp} is computed by evaluating the singular value decomposition

$$\beta = Q \Lambda P,$$

where Q and P are orthogonal matrices of dimension $K \times K$ and $r \times r$, respectively, and Λ is a $K \times r$ diagonal matrix. We may choose β_{\perp} as the last $K - r$

columns of Q .² Thus,

$$\tilde{\beta}_\perp = \begin{pmatrix} 0.566 \\ 0.610 \\ 0.555 \end{pmatrix}$$

and

$$\tilde{\alpha}_\perp = \begin{pmatrix} 0.219 \\ -0.971 \\ 0.098 \end{pmatrix}.$$

Hence,

$$\tilde{\Xi} = \begin{bmatrix} -0.247 & 1.095 & -0.110 \\ -0.266 & 1.180 & -0.119 \\ -0.242 & 1.074 & -0.108 \end{bmatrix}.$$

The estimate of the structural impact multiplier matrix obtained using the nonlinear equation solver is

$$\tilde{B}_0^{-1} = \begin{bmatrix} 0.0008 & 0.0103 & -0.0045 \\ -0.0060 & 0.0043 & 0.0000 \\ 0.0026 & 0.0196 & 0.0100 \end{bmatrix}$$

and the long-run effects of the structural shocks on the level variables in this VECM are given by

$$\tilde{\Upsilon} = \tilde{\Xi} \tilde{B}_0^{-1} = \begin{bmatrix} -0.0071 & 0 & 0 \\ -0.0076 & 0 & 0 \\ -0.0069 & 0 & 0 \end{bmatrix}.$$

Figure 11.3 shows the implied responses of logged GNP, consumption, and investment to a common productivity shock.

Alternatively, if β' is known to be

$$\begin{bmatrix} 1 & 0 & -1 \\ 0 & 1 & -1 \end{bmatrix},$$

after imposing these restrictions on the VECM, the reduced-form model may be estimated by unrestricted ML, which allows us to replace $\tilde{\Sigma}_u$ in expression (11.3.2) by the corresponding estimate from the restricted reduced-form VECM and to proceed as before.

An even easier approach, when β is known, is to express the VECM as a VAR model for

$$z_t = \begin{bmatrix} \Delta gnp \\ c_t - gnp_t \\ inv_t - gnp_t \end{bmatrix} \sim I(0)$$

and to impose a lower-triangular structure on $\Theta(1)$.

² In MATLAB the singular value decomposition can be computed by the function *svd*.

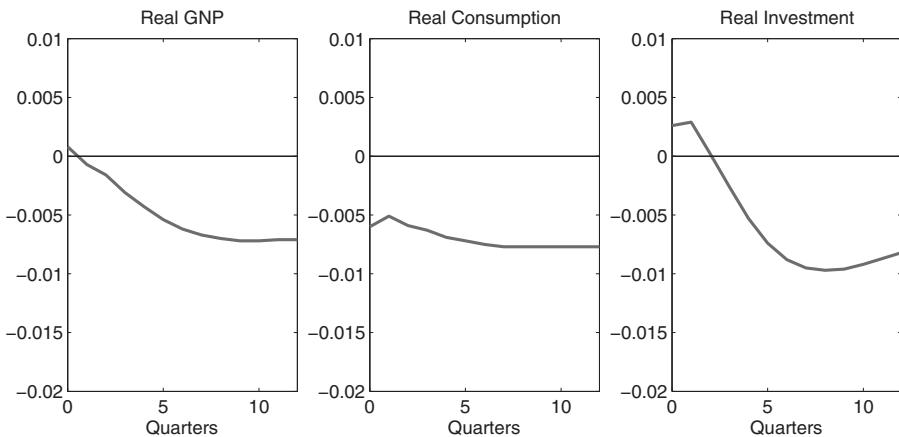


Figure 11.3. Responses to a productivity shock in the baseline model of King et al. (1991).

Full Information Maximum Likelihood Estimation. The Gaussian log-likelihood function can be set up as in Section 11.2.2. The concentrated log-likelihood is

$$\log l_c(B_0) = \text{constant} + \frac{T}{2} \log(\det(B_0)^2) - \frac{T}{2} \text{tr}(B_0' B_0 \tilde{\Sigma}_u),$$

where $\tilde{\Sigma}_u = T^{-1} \sum_{t=1}^T \hat{u}_t \hat{u}_t'$. Maximizing this function with respect to B_0 subject to the structural restrictions on B_0^{-1} and Υ is a numerical optimization problem. In the just-identified case, the solution for B_0 is obtained with a nonlinear equation solver. More generally, in the overidentified case, the likelihood has to be maximized numerically subject to all identifying restrictions.

In the empirical example based on King, Plosser, Stock, and Watson (1991) FIML estimation yields

$$\tilde{B}_0 = \begin{bmatrix} 34.905 & -155.206 & 15.707 \\ 48.705 & 15.991 & 21.917 \\ -104.537 & 9.010 & 52.958 \end{bmatrix}$$

which is the inverse of the \tilde{B}_0^{-1} that we obtained using the method of moments.

Instrumental Variable Estimation. As in the earlier monetary policy example, the model of King, Plosser, Stock, and Watson (1991) imposes identifying restrictions on B_0^{-1} , making it impossible to estimate this model by IV methods. There are other models, however, that combine long-run restrictions with identifying restrictions on B_0 (see, e.g., Pagan and Robertson 1998). Although

it is possible in principle to generalize the IV approach of Pagan and Pesaran (2008) in Section 11.2.3 to accommodate additional exclusion restrictions on B_0 , such applications appear to be rare.

11.4 Practical Limitations of Long-Run Restrictions

In Chapter 10 we already reviewed a number of conceptual challenges in imposing long-run identifying restrictions. In addition, there are challenges that arise when implementing these methods in practice. To build some intuition, consider a bivariate model with $\Theta(1)$ restricted to be lower triangular. Recall that

$$\Theta(1) = \Gamma(1)^{-1} B_0^{-1} = \begin{bmatrix} \theta_{11}(1) & 0 \\ \theta_{21}(1) & \theta_{22}(1) \end{bmatrix},$$

where $\Gamma(1) = (I_2 - \sum_{i=1}^{p-1} \Gamma_i)$, which may be estimated as

$$\widehat{\Theta}(1) = \begin{bmatrix} \widehat{\theta}_{11}(1) & 0 \\ \widehat{\theta}_{21}(1) & \widehat{\theta}_{22}(1) \end{bmatrix} = \text{chol}(\widehat{\Gamma}(1)^{-1} \widehat{\Sigma}_u \widehat{\Gamma}(1)^{-1})$$

such that

$$\widehat{B}_0^{-1} = \widehat{\Gamma}(1) \widehat{\Theta}(1).$$

Note that in constructing $\widehat{\Theta}(1)$ we have to invert $\widehat{\Gamma}(1)$. There are two potential problems that affect the reliability of the estimator of the long-run impulse responses. The first concern is that $\widehat{\Gamma}(1)$ and hence $\widehat{\Gamma}(1)^{-1}$ may be estimated imprecisely. For example, the matrix $\Gamma(1)$ may be estimated imprecisely if the finite-order VAR model is a poor approximation to an infinite-order DGP.

The second concern is that the estimate of $\Gamma(1)^{-1}$ (and hence the estimate of $\Theta(1)$) may be imprecise, even when $\Gamma(1)$ is estimated accurately. This problem arises when $\widehat{\Gamma}(1)$ is close to singular. If the matrix $\Gamma(1)$ is not well-conditioned, the estimator of its inverse may have high variance, even if the estimator of $\Gamma(1)$ has low variance. This problem is analogous to estimating the inverse $1/\gamma$ of a scalar parameter γ . Suppose γ is close to zero. Even if a very precise estimator $\hat{\gamma}$ is used, which often provides estimates very close to zero, the inverse $1/\hat{\gamma}$ has very large values and hence may have very large variance. Any small deviations from the true value of γ will be magnified by the estimator $1/\hat{\gamma}$. Thus, $1/\gamma$ may be estimated very imprecisely.³

This near-singularity problem is known to arise, for example, if one of the variables in the stationary representation of the VAR model is persistent. In fact, in the limiting case of an exact unit root, we know already that $\Gamma(1)$ is

³ It is useful to keep in mind, however, that the inversion of a matrix may also improve the reliability of an estimator. For example, reconsider the earlier scalar example. If the true γ is large, then large deviations of $\hat{\gamma}$ from γ can result in small deviations of $1/\hat{\gamma}$ from $1/\gamma$.

singular (see Chapter 3). How empirically relevant this near-singularity problem is, of course, depends on the choice of the model variables.

These potential problems with the use of long-run restrictions have been studied in the literature from different angles. This section summarizes the main arguments.

11.4.1 Estimators of the Long-Run Multiplier Matrix May Be Unreliable

As stressed by Faust and Leeper (1997), estimating the long-run multiplier of a VAR model is akin to estimating the spectral density of the data at frequency zero based on the estimated coefficients of the finite-order VAR approximation. When this finite-order approximation is poor, the model approximation error may contaminate the implied estimate of the model's long-run behavior. One approach to this problem is to increase the autoregressive sieve lag order, possibly beyond typical choices employed in empirical work. Another response has been to develop alternative nonparametric estimators of the long-run multiplier matrix (see, e.g., Christiano, Eichenbaum, and Vigfusson 2006a, 2006b). No comprehensive comparison of these approaches in finite samples exists at this point.

Some simulation evidence based on DSGE model DGPs is provided in Mertens (2012) who concludes that nonparametric estimators of the long-run variance-covariance matrix need not improve on conventional estimators. Like much of this literature, however, Mertens' work does not make any adjustments for small-sample bias in the LS estimates, nor do his lag-order choices recognize that the approximating VAR models are autoregressive sieves.

11.4.2 Lack of Power

Obtaining an accurate estimate of the impulse response at the infinite horizon amounts to pinning down the dominant autoregressive root of the process. We know that it is not possible to estimate accurately the long-run behavior of an economic time series from a short time span of data. For that reason one would expect estimates from such structural VAR models to be imprecise. This problem may persist even in large samples, as stressed by Faust and Leeper (1997). Indeed, it is well established that properly constructed confidence intervals for VAR models based on long-run restrictions tend to be wide compared with models based on short-run identifying restrictions (see, e.g., Erceg, Guerrieri, and Gust 2005; Christiano, Eichenbaum, and Vigfusson 2006b).⁴

⁴ One way of mitigating this problem is to replace the long-run identifying restrictions by finite-horizon restrictions. For example, we may identify a technology shock as the structural shock that has the most explanatory power for the level of real GDP at a large but finite horizon of 10 years (see, e.g., Francis, Owyang, Roush, and DiCecio 2014).

The mirror image of wide confidence intervals is a lack of power against alternative models. Faust and Leeper argue that, under long-run identifying assumptions, confidence intervals for structural impulse responses will asymptotically rule out false models only at a rate equal to the rejection rate of the true model (Faust and Leeper 1997, p. 347). In other words, models based on long-run identifying restrictions lack power. This statement allows for general models under the alternative.

This point is investigated further in Gust and Vigfusson (2009). They investigate the ability of impulse-response based inference in VAR models based on long-run restrictions to discriminate between alternative structural models. They make the case that, when the set of possible DGPs is restricted to a plausible class of DSGE models that satisfy the long-run identifying restrictions, the rejection probabilities for false models are greater than the rejection rates based on data coming from the true model, indicating nontrivial power. Gust and Vigfusson also observe that tests of the sign and shape of impulse response functions may have higher power than tests based on impulse response confidence intervals.

11.4.3 Near-Observational Equivalence of Shocks with Permanent Effects and Shocks with Persistent Effects

Erceg, Guerrieri, and Gust (2005) show that models based on long-run identifying restrictions have difficulty disentangling shocks with permanent effects from shocks with persistent but not permanent effects. This result is not surprising given the observational equivalence of VAR models with unit roots and with roots close to unity in small samples.

11.4.4 Weak Instrument Problems

Another concern is that the $I(0)$ variables used to achieve identification often themselves are quite persistent. The unemployment rate used in the Blanchard and Quah (1989) model is a good example. Blanchard himself makes it clear that he is conflicted about whether this variable should be considered $I(0)$ or $I(1)$ (see Blanchard 1989). The problem is that conventional asymptotics tend to become unreliable when the data are highly persistent. Gospodinov (2010) proves that the impulse responses of interest are not consistently estimable under the long-run identification scheme when the process for this variable is parameterized as local to unity, and that standard confidence intervals are invalid in this case. Gospodinov represents the Blanchard and Quah (1989) model as

$$(1 - \mathbf{C}L)y_t = \Psi(L)u_t,$$

where $y_t = (gdp_t, ur_t)$, $\Psi(L)$ is an MA operator, and

$$\mathbf{C} = \begin{bmatrix} 1 & 0 \\ 0 & 1 + \lambda/T \end{bmatrix},$$

with $\lambda \leq 0$ a fixed constant. For $\lambda = 0$, this model would reduce to a specification with both variables in differences. For $\lambda < 0$, the ur_t process is restricted to be in the neighborhood of a difference stationary process. The larger T , the closer the ur_t process becomes to a unit root process. The purpose of this alternative representation is to serve as an asymptotic device that allows us to mimic a situation in which we are unable to tell whether the unemployment rate is $I(0)$ or $I(1)$ in finite samples. For a more detailed exposition of local-to-unity asymptotics in VAR models, the reader is referred to Chapter 12.

Within this framework, Gospodinov studies the statistical properties of the impulse response estimator. He expresses the impulse response estimator as a function of parameters to be estimated by instrumental variables. He demonstrates that these parameters cannot be consistently estimated because roots local to unity in the second equation of the Blanchard-Quah model necessarily cause a weak-instrument problem. It follows immediately that conventional structural impulse response estimators are inconsistent and conventional impulse response confidence intervals are invalid. In response to this problem, Gospodinov (2010) proposes an alternative estimator of the structural impulse responses in the Blanchard and Quah (1989) model that does not suffer from the weak instrument problem. This impulse response estimator has much lower bias than conventional estimators in finite samples and lower variance.⁵ Related work also includes Chevillon, Mavroeidis, and Zhan (2015).

11.5 Can Structural VAR Models Recover Responses in DSGE Models?

11.5.1 The Origin of This Controversy

The reason why this seemingly arcane question has risen to great prominence in recent years is that Galí (1999) in particular used evidence from structural VAR models based on long-run restrictions to make the case that the class of real business cycle (RBC) models is inconsistent with U.S. macroeconomic data. The RBC model, as proposed by Kydland and Prescott (1982), is an example of a DSGE model. Galí (1999) observed that evaluating the fit of this model based on unconditional second moments alone, as proposed by Kydland and Prescott (1982), can be misleading in that a DSGE model may do well

⁵ The problem of weak instruments in long-run identified VAR models was first discussed in Pagan and Robertson (1998) and Cooley and Dwyer (1998). As noted by Watson (2006), conventional tests for weak instruments are not designed for highly persistent data and may not be able to detect weak instrument problems of the type discussed by Gospodinov (2010).

according to this criterion and yet provide a highly distorted impression of the economy's response to a technology shock.

The situation Galí had in mind was the following. The basic RBC model, as originally proposed, implies a high positive correlation between hours worked and productivity, whereas the data show a correlation near zero. This observation led researchers subsequently to augment the RBC model with non-technology shocks, which allow the model to replicate the near-zero correlation in the data. While the same unconditional correlation near zero can alternatively be explained by a suitably specified New Keynesian model, these two theoretical models differ in their implications for the responses of hours worked to a technology shock. Whereas the New Keynesian model implies that hours worked decline in response to a positive technology shock, the augmented RBC model implies a positive response of hours. Galí (1999) provided empirical evidence that hours decline persistently and significantly in response to a positive technology shock, which contradicts the augmented RBC model.

His evidence was based on structural VAR models in the tradition of Blanchard and Quah (1989), with one important difference. Given that RBC models do not generate data on the unemployment rate, Galí followed Cogley and Nason (1995) in replacing the unemployment rate in Blanchard and Quah's model with per capita hours worked, and expressed this variable in differences. Galí interpreted the impulse response estimates and related evidence from the same VAR model as implying that technology shocks do not play an important role in driving the business cycle. In view of the important role of technology shocks in explaining the business cycle in the recent literature, this evidence has been referred to as a potential paradigm shifter (see Francis and Ramey 2005).

Not surprisingly, proponents of the augmented RBC model have risen to the challenge of refuting Galí's conclusions. Some studies, including Francis and Ramey (2005), showed that the fall in hours could also be explained by RBC models that allow for even richer preference and technology structures. Others, including Christiano, Eichenbaum, and Vigfusson (2004), challenged the conclusion of Galí (1999) on the grounds that hours worked should be modeled as a stationary process in levels rather than in differences, in which case a positive shock to technology drives up per capita hours worked and real output, reversing Galí's conclusion. Finally, Chari, Kehoe, and McGrattan (2008) challenged the findings of Galí (1999) on methodological grounds. They made the case that the response estimator in Galí (1999) is prone to severe bias. Chari et al. suggested that VAR approximations to the DSGE model may generate significant declines in hours worked, even when applied to data from a DSGE model in which the response of hours is known to be positive. Their conclusion was that structural VAR models based on long-run restrictions simply should not be used to assess the empirical content of business cycle theory.

The latter conclusion has been challenged in turn by Christiano, Eichenbaum, and Vigfusson (2006b). First, Christiano et al. take issue with the realism of the RBC model specification employed by Chari et al. Second, they take issue with the specification of hours in log differences in the VAR model. They show that spurious rejections do not arise when hours are specified in levels. Third, they make the case that we ought to focus on the coverage accuracy of confidence intervals for the structural impulse responses generated from DSGE models rather than the accuracy of the impulse response point estimates. They provide evidence that users of VAR models are unlikely to spuriously reject the underlying DSGE model after allowing for sampling uncertainty in the impulse response estimates. They also provide examples in which users of structural VAR models may successfully discriminate between alternative DSGE models. Their study in turn drew a response from Kehoe (2006) who took issue with the econometric methodology for evaluating structural impulse response estimators employed in Christiano, Eichenbaum, and Vigfusson (2006b).

11.5.2 *The Position of Chari et al. (2008)*

Because this ongoing debate about the VAR methodology has received much attention, it is useful to outline each position in turn, starting with Chari, Kehoe, and McGrattan (2008). Notably, Kehoe (2006), building on his earlier work with Chari and McGrattan, makes the case that there is only one reasonable approach to evaluating DSGE models.

Suppose that we are interested in the question of how well a calibrated DSGE model fits the U.S. data, as measured by the impulse responses to structural shocks. Kehoe suggests that we first fit the structural VAR model of our choice to the actual U.S. data of sample length T . Then we generate repeated draws of length T from the calibrated DSGE model. For each of these draws, we fit the structural VAR model of interest, making sure to have transformed the data prior to the analysis exactly like the actual data. Finally, we compare the average of the implied structural impulse response estimates obtained conditional on the simulated data to the impulse response estimate of the structural VAR model fit to actual data. The distance between the average of the model-based structural impulse responses and the structural impulse responses based on the U.S. data becomes a measure of model fit.

Kehoe (2006) and Chari, Kehoe, and McGrattan (2008) refer to this approach as the Sims-Cogley-Nason approach with reference to Sims (1989) and, more importantly, Cogley and Nason (1995). The approach they describe actually is exactly how Kydland and Prescott would have evaluated a DSGE model, except that the latter would have focused on the cross-autocorrelation statistics rather than on structural impulse response statistics. We refer to this approach as the CKM approach.

Kehoe lists three reasons for preferring this approach to the conventional approach of comparing VAR estimates of the structural impulse responses to the population responses in the DSGE model. He observes, first, that his preferred approach does not suffer from small-sample bias. This claim is correct in that by conditioning on the actual data when evaluating the structural VAR model, one effectively takes a Bayesian approach. Conditional on the data, there is no small-sample bias.

Second, Kehoe suggests that this approach avoids lag-truncation bias by which he means that both the model fit to the U.S. data and the model fit to the simulated data are truncated in the same way. This does not mean that the truncation bias has been avoided so much as that this bias is the same whether working with simulated data or U.S. data.

Third, Kehoe emphasizes that this approach does not require the identifying assumptions used in the structural VAR to be correct, because we can compare the model-based statistics to the statistics based on U.S. data, even if those statistics have no economic meaning. This statement is technically correct, but it effectively amounts to reinterpreting the structural impulse responses as descriptive statistics only rather than as structural objects. Indeed, Kehoe hastens to add that in this case we might as well dispense with the identification of structural VAR models and revert back to using cross-autocorrelation statistics. Of course, this argument just amounts to ignoring the concern in Galí (1999) (as well as Cogley and Nason 1995) that evaluating the fit of a DSGE model based on unconditional second moments alone can be misleading.

According to Chari et al., the fit of a given DSGE model is measured by the extent to which the structural impulse responses obtained by fitting a VAR model to the U.S. data deviate from the average of the corresponding structural impulse response estimates obtained by fitting the same VAR model to each of many random draws of data generated from the calibrated DSGE model. Chari, Kehoe, and McGrattan (2008) are not interested in exploring the substantive implications of this approach, however. Rather, the objective of their analysis is to demonstrate that the fit of a DSGE model cannot be evaluated by comparing VAR model estimates of structural impulse responses based on U.S. data with the population impulse responses in DSGE models.

For this purpose, they design a thought experiment in which the data are generated from a DSGE model with known structural parameters. They then draw 1,000 data sequences of length T from this DSGE model and fit a structural VAR(p) model of the type employed by Galí (1999) to each sequence of simulated data. They compute the mean of the structural impulse responses implied by these structural VAR model estimates. Finally, they compare this average of the structural impulse responses to the corresponding population impulse response in the DSGE model. If the latter differs substantially from the average structural response estimates based on the simulated data, as measured by the sign or the magnitude of the response, the structural VAR model

is considered inappropriate for assessing the validity of DSGE models. Chari et al. demonstrate that the VAR estimates of the structural impulse response may look quite different on average than the corresponding population impulse responses in their DGP.

Based on this exercise, Chari, Kehoe, and McGrattan (2008) conclude that, for common choices of p and T , a researcher comparing structural VAR estimates of the impulse responses obtained from U.S. data to the population responses in the DSGE model is likely to incorrectly conclude that the data are not generated from the DSGE model. Chari et al. attribute the apparent failure of the structural VAR approach in this thought experiment mainly to lag-order truncation bias.

11.5.3 *The Position of Christiano et al. (2006)*

How does the CKM approach differ from Christiano et al.'s approach? Christiano, Eichenbaum, and Vigfusson (2006b) take the position, consistent with the frequentist approach to econometrics, that the data are random, but that the DSGE model parameters are non-random, because, after all, we know their exact values when we specify the DSGE model. This fact allows one to compute the exact population value of the structural impulse responses of interest based on the $\text{VAR}(\infty)$ representation of the DSGE model.

The objective is to judge whether the population value of the statistic of interest in the DSGE model is consistent with the data. One addresses this question by fitting a structural VAR model to the U.S. data and approximating the sampling distribution of the statistic of interest by bootstrap or by asymptotic methods. After forming a confidence interval for the statistic of interest, one rejects the DSGE model at the chosen significance level γ , if the population value of the statistic of interest in the DSGE model lies outside of the $(1 - \gamma) 100\%$ confidence interval.

In the context of Christiano et al.'s study, the focus is on the response of hours worked to a technology innovation. In practice, they consider two summary statistics. The first summary statistic is the bias of the structural impulse response estimator. The bias is computed as the difference between the population impulse responses in the DSGE model and the average of the structural impulse response estimates obtained by fitting an approximating VAR model to each of 1,000 simulated data sets of sample length T generated by the DSGE model. The second summary statistic refers to the probability that pointwise bootstrap confidence intervals for the structural impulse responses, constructed from the finite-order VAR approximation, include the population value of the structural impulse response in repeated sampling (see Inoue and Kilian 2002b). The closer this probability is to the nominal coverage probability $1 - \gamma$, the more accurate the interval. Christiano et al. evaluate the effective coverage probability by keeping track of how many of the bootstrap interval

estimates – each obtained after fitting a VAR model to one of the 1,000 data sets of length T coming from the DSGE model – include the population value of the response.

Christiano, Eichenbaum, and Vigfusson (2006b) make the case that even if the VAR point estimates of the structural impulse responses are inaccurate in small samples, after accounting for sampling uncertainty, researchers would rarely reject a DSGE model incorrectly. Although the confidence bands may be wide, they are not so wide as to be consistent with any possible DSGE model. Christiano et al. show by example that, at least in some cases, the confidence bands are tight enough to allow the researcher to discriminate between competing DSGE models.

The approach employed by Christiano et al. is standard in frequentist econometrics. Thus, Kehoe's assertion that Christiano et al.'s approach lacks statistical foundations is without basis. If anything, the statistical foundations of his own preferred approach seem debatable, especially when compared to a proper Bayesian analysis. The “confidence intervals” reported in figure 4 and table 1 of Chari, Kehoe, and McGrattan (2008) are neither regions of highest posterior density nor confidence intervals in the frequentist sense, calling into question the authors' conclusion that the structural VAR model in question spuriously rejects the DSGE model that generated the data.

11.5.4 Understanding the Simulation Evidence

Kehoe (2006) raises several concerns about the reliability of Christiano et al.'s approach. Many of these concerns relate to the bias of the impulse response estimates obtained after imposing long-run restrictions. There are four distinct potential sources of bias. One concern is that all structural VAR models, but especially those based on long-run identifying restrictions, are subject to truncation bias in the lag order. This argument reflects a view exemplified by Cooley and Dwyer (1998, p. 76) that, as a theoretical matter, finite-order VAR models cannot accommodate the VARMA representations of DSGE models. This view is mistaken (see Chapter 2). Indeed, Chari, Kehoe, and McGrattan (2008) recognize that, under suitable conditions, a VARMA model can be represented as a $\text{VAR}(\infty)$ model. It is well known that under these conditions the finite-order $\text{VAR}(p)$ model can be viewed as a semiparametric approximation, which becomes arbitrarily accurate asymptotically if the lag order increases with T at a suitable rate (see, e.g., Lewis and Reinsel 1985; Lütkepohl and Poskitt 1991; Inoue and Kilian 2002a). This means that the truncation bias refers to finite-sample approximation error rather than omitted variable bias. This distinction is not just a matter of semantics. It not only invalidates standard procedures for selecting the lag order and affects how one conducts inference about structural impulse responses, but it has been shown that the lag order required for reasonable sieve approximations tends to be

larger than for conventional models with finite lag order (see, e.g., Inoue and Kilian 2002b). Thus, evidence that a VAR(4) approximation may yield impulse response estimates subject to large approximation error, for example, is hardly unexpected.

The question is whether this approximation error can be mitigated by the choice of a much larger lag order. Chari et al.'s premise is that for common choices of T , VAR models necessarily require short lag orders. This is not the case. There is considerable room for increasing the lag order in practice. Chari, Kehoe, and McGrattan (2008) provide an example where increasing the lag order from 4 to about 40 greatly improves the approximation to the population responses. While this example may be extreme, it remains to be seen what the benefit is more generally from increasing the lag order beyond the customary 4 or 8 quarterly lags.

Many researchers are reluctant to increase the VAR lag order because of the perception that adding more lags necessarily greatly inflates the variance of the impulse response estimator and renders the model uninformative. This perception is not correct. Not only are there counterexamples in empirical work showing that statistically significant impulse responses may be obtained even when using many autoregressive lags (see, e.g., Kilian 2009), but it has been shown that impulse responses from VAR models that are underparameterized tend to be highly misleading (see Kilian 2001).

There are several ways of approaching this problem. For example, Christiano, Eichenbaum, and Vigfusson (2006b) experiment with alternative non-parametric estimators of the long-run variance matrix in VAR models with long-run restrictions. An obvious question is whether this problem could be solved simply by increasing the approximating lag order substantially. The answer will be case specific. Intuitively, the lag order required for an accurate approximation depends on the magnitude of the MA roots in the VARMA representation of the DSGE model. One would not expect VAR models of given lag order to work equally well for all DSGE models for that reason. It therefore would be useful to know in particular what lag order it takes to achieve a good approximation for the DSGE models of interest to policymakers.

The second source of bias is small-sample bias of the type discussed in Pope (1990) and Kilian (1998c) (see Section 2.3.3). Chari, Kehoe, and McGrattan (2008) suggest that this small-sample bias is small. This finding is not unexpected if both real output and per capita hours worked are expressed in log-differences. Small-sample bias is expected to be higher when per capita hours are expressed in levels, as in Christiano et al.'s analysis. Indeed, Christiano, Eichenbaum, and Vigfusson (2006b, p. 26) present evidence of small-sample bias. In this case, it would make sense to consider using well-known methods of correcting for this bias, at least in constructing confidence bands. In contrast, Christiano et al. rely on Runkle's (1987) bootstrap method that ignores small-sample bias, and the number of bootstrap replications used in their study

is unreasonably small, raising questions about the reliability of the results (see Chapter 12).

The third source of bias is asymptotic in nature and arises in models with per capita hours worked expressed in levels. This is the type of model used in Christiano, Eichenbaum, and Vigfusson (2006b). Hours worked in levels are quite persistent. As shown by Gospodinov (2010), such estimates are biased even asymptotically when hours follow a local-to-unity process, which is expected to be a good approximation when we are unsure of whether hours are $I(0)$ or $I(1)$ in the data. Gospodinov proposes an alternative impulse response estimator that is consistent and that is more precise than conventional estimators. The latter method does not appear to have been used in the debate at hand. Although this method was designed for finite-order VAR processes, extensions to VAR(∞) models should be feasible. It is also worth pointing out that there is some tension between the RBC model used by Christiano et al., which implies that per capita hours are stationary, and the actual data which may or may not be stationary.

Fourth, we already highlighted several reasons why VAR inference based on long-run restrictions may be unreliable in practice. One concern was the difficulty of interpreting the model estimates, when the model includes only one non-technology shock as in Blanchard and Quah (1989). This result is consistent with Christiano et al.'s finding that VAR models with four rather than two variables are better at recovering the DSGE model responses. Extending the model may not be enough, however, if there are non-technology shock processes in the model that themselves are highly persistent (see also Erceg, Guerrieri, and Gust 2005).

Fifth, none of the studies in question explains how its authors dealt with the problem that the impulse response functions implied by long-run restrictions are not unique without normalizing the sign of the response function of interest (see Section 10.5.4). This question is of obvious importance, given that this debate was triggered by questions about the sign of the response of hours worked to a technology shock. It also matters because the choice of normalization affects the coverage accuracy of impulse response confidence intervals and our ability to discriminate between alternative DSGE models.

To summarize, there are many potential sources of bias in the VAR impulse response estimates, some of which are specific to VAR models based on long-run identifying assumptions. The magnitude of the bias in the response estimates can also be sensitive to the specification of the underlying DSGE model, making it difficult to generalize from specific examples. For example, Christiano, Eichenbaum, and Vigfusson (2006b) demonstrate by simulation that recursively identified VAR models are quite accurate at recovering the population responses in the DSGE model if these short-run identifying restrictions are also satisfied in the underlying DGP. Such restrictions exist in a subset of DSGE models, which does not include the standard RBC model. As Kehoe (2006) correctly points out, this result does not mean that recursively

identified VAR models can be used more broadly to discriminate between alternative theoretical models. In particular, one cannot reject theoretical models merely because their implications are not consistent with the estimates of recursively identified structural VAR models, unless these models also satisfy the identifying assumptions used in the VAR models. However, the results in Christiano, Eichenbaum, and Vigfusson (2006b) and Kascha and Mertens (2009) provide further evidence that truncation bias may not be the main problem, nor for that matter small-sample bias, but the identification method.

Chari, Kehoe, and McGrattan (2008) and Kehoe (2006) also question the ability of structural VAR models to provide accurate inference about the validity of DSGE models, but that assessment is not based on the methods of inference actually being used by practitioners, but rather on their own diagnostics. The evidence in Chari, Kehoe, and McGrattan (2008) is based on a VAR model in which hours worked are log-differenced. Christiano, Eichenbaum, and Vigfusson (2006b) make the case that – from a theoretical point of view – a VAR specification with hours worked in levels is the only specification that is consistent with the underlying DSGE model. Even Chari, Kehoe, and McGrattan (2008) concede that in the latter case, no spurious rejection of the DSGE model arises, even using their nonstandard approach to inference. That conclusion is consistent with simulation evidence on the coverage accuracy of structural impulse response confidence intervals in Christiano, Eichenbaum, and Vigfusson (2006b), which shows that the use of conventional pointwise confidence intervals tends to protect researchers from spurious inference, as long as enough of the variability in hours worked is driven by the structural shocks of interest. This evidence, in short, suggests that much of this debate actually is about the question of how to specify the log of hours worked in VAR models based on long-run restrictions.

Equally importantly, it can be shown by example that these confidence intervals at least in some cases help discriminate among competing DSGE models, undermining the claim that structural VAR models based on long-run restrictions are altogether useless for developing business cycle theory. Nevertheless, much more work remains to be done to answer the questions raised by this debate. Perhaps the most useful insight provided by this literature so far is that whether or not estimates from a structural VAR model will be informative depends on the structure of the economy, on the specification of the structural VAR model, on there being enough variability in the data to allow us to identify the model parameters of interest, and potentially even on unmodeled features of the U.S. data.

11.5.5 Summary

It is important to keep in mind that this discussion relates to the ability of VAR models with long-run restrictions to recover impulse responses from a make-believe world described by a stylized DSGE model. This focus is helpful in that

it allows us to rule out a number of potential misspecification issues that we would confront with actual data. The results have to be viewed with caution, however, precisely because actual data need not conform to the assumptions of the DSGE model. A case in point is the analysis in Gospodinov, Maynard, and Pesavento (2011) of how features beyond the scope of standard DSGE business cycle models (such as the low-frequency correlation between hours and productivity) may affect the reliability of alternative specifications of VAR models based on long-run restrictions.

Keeping in mind these caveats, what have we learned from this debate about the reliability of structural VAR models based on long-run restrictions? First, the answer depends on the DSGE model of interest. On the one hand, even after making sure that the VAR model has as many structural shocks as the DSGE model and that the identifying restrictions are consistent with the DGP, there are situations in which confidence intervals based on structural VAR models imposing long-run restrictions may have serious coverage deficiencies. On the other hand, there are also many DSGE model specifications for which structural VAR models based on long-run restrictions are accurate with reasonably tight intervals, allowing the rejection of economically interesting alternatives (see also Erceg, Guerreri, and Gust 2005).

Second, when working with long-run identifying assumptions, the VAR lag order is important for determining the accuracy of the approximation. The simulations in Chari, Kehoe, and McGrattan (2008, figure 3), which represent the least favorable DSGE model example discussed in this literature, show that with four autoregressive lags, the approximation to the true impulse response is poor, but with 40 lags the bias, as defined in that paper, appears reasonably small. Of course, these numbers are specific to their example. It remains to be seen to what extent the performance of structural models based on long-run restrictions may be improved by more careful attention to the lag order and which features of the DSGE model may facilitate the use of low-order VAR approximations (see, e.g., Pagan and Robinson 2016).

Third, there are additional concerns about the use of structural VAR models based on long-run restrictions in particular. Kascha and Mertens (2009), for example, attribute the biases of the impulse response estimators reported in the literature to the use of long-run identifying restrictions rather than to any truncation bias in the approximating VAR model. If there were no alternatives to the use of long-run restrictions in practice, one could argue that the precise reason for these biases would be of little consequence. Even granting that sometimes it can be difficult to justify conventional short-run exclusion restrictions in DSGE models, however, this is not the only alternative-identification scheme. For example, Canova and Paustian (2011) report considerable success in recovering the qualitative characteristics of the underlying structural impulse responses when fitting structural VAR models with sign restrictions to data generated from DSGE models. This alternative class of

models is discussed in Chapter 13. The current debate has largely ignored this and other alternative identification strategies.

Finally, as stressed by Kehoe (2006), one must not draw conclusions based on estimates of structural VAR models derived under a given set of identifying assumptions about theoretical models that do not satisfy these assumptions. This is only half of the message, however. The other half is that it is difficult to generalize from one or two DSGE models to the ability of structural VAR models to recover population responses in general. It may seem that if the structural VAR approach does not work in a simple case, it cannot be trusted more generally. This conclusion is unwarranted, because we know that the extent to which truncation bias in particular matters will differ from one DSGE model to the next, depending on the magnitude of the MA roots. Hence, if one wants to make a case against the use of structural VAR models, it is not enough to show that for some stylized and grossly unrealistic DSGE model the structural VAR approach can be inaccurate. Rather, one needs to show that structural VAR models are unable to capture the underlying population responses of DSGE models that deserve our attention because they fit the data reasonably well and because they are economically plausible.

12 Inference in Models Identified by Short-Run or Long-Run Restrictions

Chapters 9 and 11 reviewed the estimation of the structural impulse responses at horizon i , Θ_i , in structural VAR models subject to short-run and/or long-run restrictions. This chapter focuses on inference in exactly identified structural VAR models. We discuss the leading methods of constructing confidence intervals for structural impulse responses and related statistics such as structural forecast error variance decompositions. Similar methods of inference can be applied to all statistics discussed in Chapter 4 with the exception of historical decompositions. The latter statistics do not permit the construction of confidence intervals because they condition on a particular realization of the data.

Section 12.1 presents the delta method, followed by bootstrap methods in Sections 12.2 and 12.3. Section 12.4 discusses potential pitfalls in conducting inference about structural impulse responses in stationary models. In Section 12.5 we summarize the evidence on the finite-sample properties of alternative confidence intervals for structural impulse responses. Section 12.6 extends the analysis to integrated and cointegrated processes. Section 12.7 discusses local-to-unity asymptotic and bootstrap confidence intervals for structural impulse responses. In Section 12.8, we review an alternative and seemingly simpler approach to inference about impulse responses based on local projections and discuss its practical and conceptual limitations. In Section 12.9, we provide practical recommendations for conducting pointwise inference about structural impulse responses. Section 12.10 contrasts Bayesian and frequentist approaches to pointwise inference. Section 12.11 focuses on joint inference about sets of structural impulse responses. Section 12.12 briefly discusses other bootstrap applications including bootstrap prediction and the construction of bootstrap critical values for test statistics. Section 12.13 provides a number of empirical examples. It also discusses extensions to overidentified structural VAR models. Inference in other classes of structural VAR models including set-identified models is considered in subsequent chapters.

12.1 Delta Method Intervals for Structural Impulse Responses

We begin with a summary of asymptotic results based on the delta method for structural impulse response estimates from exactly identified VAR models, building on the analysis in Lütkepohl (1990). Consider the VAR(p) model

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where y_t is a K -dimensional vector and u_t is an iid error term with at least four finite moments. Under suitable conditions,

$$\sqrt{T} \begin{pmatrix} \widehat{\alpha} - \alpha \\ \widehat{\sigma} - \sigma \end{pmatrix} \xrightarrow{d} \mathcal{N} \left(0, \begin{bmatrix} \Sigma_{\widehat{\alpha}} & 0 \\ 0 & \Sigma_{\widehat{\sigma}} \end{bmatrix} \right), \quad (12.1.1)$$

where T is the sample size, $\widehat{\alpha}$ and $\widehat{\sigma}$ denote consistent estimators of $\alpha = \text{vec}([A_1, \dots, A_p])$ and $\sigma = \text{vech}(\Sigma_u)$, respectively. Block-diagonality of the asymptotic covariance matrix in expression (12.1.1) is assumed here for deriving the asymptotic properties of impulse response estimators because it simplifies the asymptotic distributions. This assumption holds, for example, when the errors are iid and symmetrically distributed with four finite moments. More generally, $\widehat{\alpha}$ and $\widehat{\sigma}$ are not asymptotically independent, and hence, the asymptotic covariance matrix is not block diagonal. Nevertheless, the same techniques may be used to derive the asymptotic distribution of the structural impulse response estimator (see, e.g., Lütkepohl 2005, section 3.7, remark 3, or Brüggemann, Jentsch, and Trenkler 2016, theorem 2.1). The expression for the asymptotic covariance matrix will be different, however.

In the special case of Gaussian iid innovations, the covariance matrix $\Sigma_{\widehat{\sigma}}$ simplifies to $\Sigma_{\widehat{\sigma}} = 2\mathbf{D}_K^+(\Sigma_u \otimes \Sigma_u)\mathbf{D}_K^{+'}$, where $\mathbf{D}_K^+ = (\mathbf{D}_K'\mathbf{D}_K)^{-1}\mathbf{D}_K'$ is the Moore-Penrose inverse of \mathbf{D}_K , and \mathbf{D}_K is the $K^2 \times \frac{1}{2}K(K+1)$ duplication matrix, as defined in Lütkepohl (2005, appendix A12.2), and $\Sigma_{\widehat{\alpha}} = \Gamma_Y(0)^{-1} \otimes \Sigma_u$, where $\Gamma_Y(0)$ may be estimated as ZZ'/T , given $Z \equiv (Y_0, \dots, Y_{T-1})$ and $Y_t \equiv (y'_t, \dots, y'_{t-p+1})'$. If the regression includes deterministic regressors, $\Sigma_{\widehat{\alpha}}$ corresponds to the last $K^2 p$ rows and columns of the variance-covariance matrix of the VAR regressors.

Note that each structural impulse response coefficient $\theta_{jk,i}$ may be expressed as a component of a continuously differentiable function $g(\alpha, \sigma) = (g_1(\alpha, \sigma), \dots, g_m(\alpha, \sigma))'$ with values in the m -dimensional Euclidean space, where m is the number of structural impulse response coefficients under consideration. If the partial derivatives $\partial g_i / \partial \alpha'$ and $\partial g_i / \partial \sigma'$ are nonzero at the true parameter values α and σ for $i = 1, \dots, m$, an application of the delta method yields

$$\sqrt{T} (g(\widehat{\alpha}, \widehat{\sigma}) - g(\alpha, \sigma)) \xrightarrow{d} \mathcal{N} \left(0, \frac{\partial g}{\partial \alpha'} \Sigma_{\widehat{\alpha}} \frac{\partial g'}{\partial \alpha} + \frac{\partial g}{\partial \sigma'} \Sigma_{\widehat{\sigma}} \frac{\partial g'}{\partial \sigma} \right). \quad (12.1.2)$$

Standard asymptotic inference requires the limiting variance to be nonzero. That condition may not hold if there are constraints on the model coefficients or if the model variables are integrated and/or cointegrated, in which case $\Sigma_{\hat{\alpha}}$ is singular (see Chapter 3). Even if $\Sigma_{\hat{\alpha}}$ itself is nonsingular, however, the partial derivatives $\partial g_i / \partial \alpha'$ may be zero, invalidating inference. This may occur in some points of the parameter space because these derivatives are sums of products of the reduced-form VAR slope parameters. An example would be a finite-order VAR process in which some variable does not respond to a given structural shock such that the impulse response is zero in population (see Benkowitz, Lütkepohl, and Neumann 2000). We will return to this point later in this chapter.

The precise form of the variance-covariance matrices differs, depending on whether we are conditioning on a finite-order VAR DGP or approximating an infinite-order VAR process. The following results draw on Lütkepohl (1990) and Lütkepohl and Poskitt (1991). While we focus on the asymptotic variance-covariance matrix for structural impulse responses, similar results for related statistics such as cumulative impulse responses and structural forecast error variance decompositions can be found in Lütkepohl (2005).

12.1.1 Finite-Order VAR Models

Let y_t be generated by a stationary VAR model of known finite lag order p with iid white noise errors. Extensions of the delta method to VAR models with conditionally heteroskedastic errors are discussed in Brüggemann, Jentsch, and Trenkler (2016). In the latter case, the simple additive structure of the variance in equation (12.1.2) is lost because the asymptotic covariance of $\hat{\alpha}$ and $\hat{\sigma}$ in equation (12.1.1) is not zero.

We are interested in conducting inference on the elements of the matrix $\hat{\Theta}_i$ of estimates of the structural impulse responses at horizon i , where i is fixed with respect to the sample size T . For expository purposes, let B_0^{-1} equal the lower-triangular Cholesky decomposition of Σ_u such that $B_0^{-1}B_0^{-1\prime} = \Sigma_u$. Then

$$\sqrt{T} \text{vec}(\hat{\Theta}_i - \Theta_i) \xrightarrow{d} \mathcal{N}\left(0, C_i \Sigma_{\hat{\alpha}} C_i' + \bar{C}_i \Sigma_{\hat{\sigma}} \bar{C}_i'\right), \quad i = 0, 1, 2, \dots,$$

where

$$\begin{aligned} C_0 &= 0 \\ C_i &= (B_0^{-1\prime} \otimes I_K) G_i, \quad i = 1, 2, \dots \\ \bar{C}_i &= (I_K \otimes \Phi_i) H, \quad i = 0, 1, 2, \dots \end{aligned}$$

and

$$G_i = \frac{\partial \text{vec}(\Phi_i)}{\partial \alpha'} = \sum_{m=0}^{i-1} J(\mathbf{A}')^{i-1-m} \otimes \Phi_m, \quad i = 1, 2, \dots$$

$$H = \frac{\partial \text{vec}(B_0^{-1})}{\partial \sigma'} = \mathbf{L}'_K \{ \mathbf{L}_K (I_{K^2} + \mathbf{K}_{KK}) (B_0^{-1} \otimes I_K) \mathbf{L}'_K \}^{-1},$$

where

$$\Phi_m = J \mathbf{A}^m J',$$

$$\begin{matrix} J \\ K \times K_p \end{matrix} = [I_K, 0_{K \times K(p-1)}],$$

$$\begin{matrix} \mathbf{A} \\ K_p \times K_p \end{matrix} = \begin{bmatrix} A_1 & A_2 & \dots & A_{p-1} & A_p \\ I_K & 0 & \dots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \dots & I_K & 0 \end{bmatrix},$$

\mathbf{L}_K is the elimination matrix defined such that for any $m \times m$ matrix F , $\text{vech}(F) = \mathbf{L}_m \text{vec}(F)$, and \mathbf{K}_{mn} is the commutation matrix defined such that for any $m \times n$ matrix G , $\mathbf{K}_{mn} \text{vec}(G) = \text{vec}(G')$ (see Lütkepohl 2005, appendix A12.2). If B_0^{-1} is not recursive, the expression for H must be modified accordingly.

Let diag denote the operator that retains the diagonal elements of a square matrix and stacks them into a column vector. Then the $K^2 \times 1$ vector $\sigma_{\text{vec}(\widehat{\Theta}_i)}$, consisting of the asymptotic standard errors of the elements of $\text{vec}(\widehat{\Theta}_i)$, contains the square roots of the elements of the vector

$$\text{diag}(C_i \Sigma_{\widehat{\alpha}} C'_i + \bar{C}_i \Sigma_{\widehat{\sigma}} \bar{C}'_i) / T. \quad (12.1.3)$$

The vector of nominal $(1 - \gamma)$ pointwise asymptotic confidence intervals may be formed as

$$\text{vec}(\widehat{\Theta}_i) \pm z_{\gamma/2} \widehat{\sigma}_{\text{vec}(\widehat{\Theta}_i)}, \quad i = 0, 1, 2, \dots,$$

where $\widehat{\sigma}_{\text{vec}(\widehat{\Theta}_i)}$ is obtained by replacing the unknowns in $\sigma_{\text{vec}(\widehat{\Theta}_i)}$ by consistent estimates, and $z_{\gamma/2}$ denotes the $\gamma/2$ quantile point of the standard normal distribution. In practice, it is most common to report one-standard error or two-standard error confidence intervals, corresponding, respectively, to (approximate) 68% or 95% two-sided confidence intervals.

The estimates $\widehat{\sigma}_{\text{vec}(\widehat{\Theta}_i)}$ may be constructed by bootstrap methods, as discussed later in this chapter, or using asymptotic closed-form solutions based

on expression (12.1.3). In applied work it is common to rely on closed-form solutions for $\Sigma_{\hat{\alpha}}$ and $\Sigma_{\hat{\sigma}}$ that have been derived under the assumption of Gaussian white noise errors. Specifically we replace the unknown parameters in $\Sigma_{\hat{\alpha}} = \Gamma_Y(0)^{-1} \otimes \Sigma_u$ and $\Sigma_{\hat{\sigma}} = 2\mathbf{D}_K^+(\Sigma_u \otimes \Sigma_u)\mathbf{D}_K^{+'}$ by consistent estimators. The advantage of bootstrap estimates of the standard error is that they allow us to relax the assumption of Gaussian white noise reduced-form errors, provided a suitable bootstrap method is employed.

12.1.2 Infinite-Order VAR Models

Let y_t be generated by a stationary $\text{VAR}(\infty)$ process with conditionally homoskedastic white noise errors to be approximated by a VAR model of finite lag order p_T , where p_T is assumed to increase with T at an appropriate rate (see Chapter 2). We are interested in conducting inference on the elements of $\widehat{\Theta}_i$ at horizon i , where i is fixed with respect to T . For expository purposes, let B_0^{-1} again equal the lower-triangular Cholesky decomposition of Σ_u such that $B_0^{-1}B_0^{-1'} = \Sigma_u$. Then

$$\sqrt{T}\text{vec}(\widehat{\Theta}_i(p_T) - \Theta_i) \xrightarrow{d} \mathcal{N}(0, \Omega_\theta(i)), \quad i = 0, 1, 2, \dots,$$

where

$$\Omega_\theta(i) = \left(I_K \otimes \sum_{j=0}^{i-1} \Phi_j \Sigma_u \Phi_j' \right) + \bar{C}_i \Sigma_{\hat{\sigma}} \bar{C}_i'$$

and $\Sigma_{\hat{\sigma}}$ is the asymptotic covariance matrix of $\sqrt{T}\text{vech}(\frac{1}{T} \sum_{t=1}^T u_t u_t' - \Sigma_u)$. The $K^2 \times 1$ vector $\sigma_{\text{vec}(\widehat{\Theta}_i)}$, consisting of the asymptotic standard errors of the elements of $\text{vec}(\widehat{\Theta}_i)$, contains the square roots of the elements of the vector

$$\text{diag}(\Omega_\theta(i))/T. \tag{12.1.4}$$

A nominal $(1 - \gamma)100\%$ pointwise asymptotic confidence interval may be formed as

$$\text{vec}(\widehat{\Theta}_i) \pm z_{\gamma/2} \widehat{\sigma}_{\text{vec}(\widehat{\Theta}_i)}, \quad i = 0, 1, 2, \dots,$$

where $\widehat{\sigma}_{\text{vec}(\widehat{\Theta}_i)}$ is obtained by replacing the unknowns in $\sigma_{\text{vec}(\widehat{\Theta}_i)}$ by consistent estimates, and $z_{\gamma/2}$ denotes the $\gamma/2$ quantile of the standard normal distribution. As in the finite-order case, expression (12.1.4) may be estimated by bootstrap methods or using asymptotic closed-form solutions derived under the assumption of Gaussian white noise innovations.

12.1.3 Discussion

Our analysis highlights that different asymptotic thought experiments imply different closed-form solutions for the asymptotic standard errors of the structural impulse responses. In the case of a known and finite lag order, the asymptotic standard errors of the structural impulse responses ultimately decline toward zero as the horizon increases. In sharp contrast, if the VAR model is viewed as an autoregressive sieve approximation, the asymptotic standard errors of the structural impulse responses are a nondecreasing function of the horizon and always nonzero. This fact forces us to choose between one or the other asymptotic thought experiment when constructing delta method confidence intervals. This choice does not arise when using the bootstrap methods discussed later in this chapter. Inoue and Kilian (2002b) and Gonçalves and Kilian (2007) show that confidence intervals based on the standard bootstrap methods remain asymptotically valid under either asymptotic thought experiment. Simulation evidence for structural VAR models based on short-run identifying restrictions of the type discussed in Chapter 8 shows that approximating VAR models can provide reliable estimates of structural impulse responses generated from stationary invertible VARMA models, as long as a suitably large lag order is chosen. For further discussion of the autoregressive sieve bootstrap method, see also Paparoditis (1996) and Kreiss, Paparoditis, and Politis (2011).

12.1.4 Extensions to Other Statistics

Although we focus on inference about structural responses as our leading example, the delta method may be applied to a wide range of statistics of interest in structural VAR analysis, including cumulative structural impulse responses and structural forecast error variance decompositions (see Chapter 4). Closed-form solutions for the asymptotic standard errors of these statistics under the assumption of Gaussian white noise errors are provided in Lütkepohl (2005). Alternatively, these standard errors may be estimated by appropriate bootstrap simulation methods, as discussed in the next section. Of course, confidence intervals based on multiples of bootstrap estimates of the standard error are subject to the same caveats regarding the limiting variance-covariance matrix as the delta method.

Bootstrap methods in practice tend to be the method of choice for two reasons. First, there is no alternative to the use of bootstrap methods when dealing with statistics, for which closed-form solutions for the asymptotic standard error have yet to be derived. Second, bootstrap methods are recommended because, in practice, VAR innovations rarely are Gaussian, calling into question the use of standard closed-form solutions for the asymptotic variance derived under the assumption of Gaussian white noise errors.

12.1.5 On the Choice of the Significance Level

As noted by Leamer (1978) and Sims (1987), there is no universally valid standard for statistical significance. The choice of the significance level in statistical testing and in constructing confidence intervals depends on how tightly parameterized the model in question is and how large the sample size is. The conventional choice of a 5% significance level (corresponding to a 95% confidence interval) is merely a useful convention that facilitates the communication of the reliability of estimates from highly restricted traditional DSEMs. There is no reason to expect this convention to be useful in discussing estimates from largely unrestricted VAR models estimated on comparatively short samples. Many applied users instead favor 68% confidence intervals (corresponding to +/- one standard error bands) over 95% confidence intervals (approximately corresponding to +/- two standard error bands) in reporting VAR estimates of structural impulse responses. Occasionally, researchers report intervals for more than one significance level, but this is not always an option because it tends to make the impulse response plots look too crowded.

When bootstrap methods are used in constructing confidence intervals, as discussed in the next section, an additional consideration is that the number of draws required to accurately estimate the 2.5th and 97.5th percentiles of the sampling distribution used in constructing a 95% confidence interval tends to be much larger than the number of draws required for the 16th and 84th percentiles used in the construction of a 68% confidence interval.

12.2 Bootstrap Intervals for Structural Impulse Responses

The central idea underlying the bootstrap approach is that we approximate the distribution of the statistic of interest based on its sample analogue, allowing us to mimic the underlying sampling experiment (see, e.g., Efron and Tibshirani 1993). Most bootstrap applications rely on simulation methods in practice. One advantage of bootstrap methods in our context is that they allow inference about smooth differentiable functions $g(\alpha, \sigma)$ even when closed-form solutions of the variance of the estimator are not available or inconvenient to use.

Another advantage is that the bootstrap approach is more general than standard asymptotic inference. For example, the closed-form solution of the variance of the structural impulse response estimator discussed in Section 12.1 relies on the VAR errors being iid Gaussian, whereas the standard residual-based bootstrap estimator remains asymptotically valid under weaker conditions.

A third advantage is that suitably constructed bootstrap confidence intervals in many applications tend to be more accurate in small samples than asymptotic approximations. In addition, under some conditions, the bootstrap methods may improve the accuracy of the asymptotic approximation.

However, bootstrap methods are no panacea. First, in some cases, they may not be applicable at all. Second, they must be used carefully. Inappropriate bootstrap applications may result in considerably less accurate inference than using first-order asymptotic approximations based on the delta method. Finally, bootstrap methods should only be used when they have asymptotic justification. Their rationale – like that of the delta method – is ultimately asymptotic and relies on the sample being large enough to be representative of the DGP.

In what follows, we again focus on the problem of inference about structural impulse responses. All bootstrap methods discussed can also be used to construct confidence intervals for related statistics such as cumulative structural impulse responses or structural forecast error variance decompositions.

12.2.1 The Standard Residual-Based Recursive-Design Bootstrap

The first study to propose the use of bootstrap methods for structural impulse responses in stationary VAR models was Runkle (1987). In fact, this work pre-dates the development of delta method intervals for structural impulse responses and was borne out of necessity, given the absence of asymptotic closed-form solutions at the time. It was not until Bose (1988) that the first-order asymptotic validity of this residual-based bootstrap method was established for autoregressive slope parameters.

Let u_t be iid white noise with distribution F , or $u_t \stackrel{iid}{\sim} F$ for short. We approximate the unknown stationary $\text{VAR}(p)$ DGP of known order p ,

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where $u_t \stackrel{iid}{\sim} F$, conditional on the observed data, by the bootstrap DGP

$$y_t^* = \hat{v} + \hat{A}_1 y_{t-1}^* + \cdots + \hat{A}_p y_{t-p}^* + u_t^*,$$

with $u_t^* \stackrel{iid}{\sim} \hat{F}_T$. Here $[\hat{v}, \hat{A}_1, \dots, \hat{A}_p]$ denotes the LS estimate of the model parameters conditional on the observed sample $\{y_t\}_{t=-p+1}^T$ and \hat{F}_T is the implied estimate of the error distribution F . To distinguish the realizations of the bootstrap DGP from the actual realizations they are marked with a superscript $*$. The algorithm, as described here, allows for nonzero means in the data, but not for other deterministic regressors such as a linear time trend or seasonal dummies. Such extensions may be handled in the obvious way. Moreover, the least-squares estimator may be replaced by any other consistent estimator without affecting the analysis.

The maintained assumption is that $\mathbb{E}(u_t) = 0$ and that u_t has finite moments of the appropriate order. The estimator \widehat{F}_T may be parametric or nonparametric. If we knew that $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_u)$, for example, then $u_t^* \stackrel{iid}{\sim} \mathcal{N}(0, \widehat{\Sigma}_u)$, where $\widehat{\Sigma}_u$ is the least-squares estimate of Σ_u . Draws for u_t^* would be generated by drawing from the estimated Gaussian distribution, which represents \widehat{F}_T .

In the more likely case that we do not know the parametric family of the distribution of the error term, there is no need to estimate F explicitly. Rather, we draw u_t^* with replacement from the set of regression residuals $\{\widehat{u}_t\}_{t=1}^T$, where $\widehat{u}_t = y_t - \widehat{v} - \widehat{A}_1 y_{t-1} - \cdots - \widehat{A}_p y_{t-p}$. This ensures that u_t^* in bootstrap population is iid, conditional on the data, and has the same distribution as the residuals, whatever that distribution might be. If the regression model does not include an intercept, we instead resample the demeaned residuals $\{\widehat{u}_t - \bar{u}\}_{t=1}^T$, where $\bar{u} = \sum_{t=1}^T \widehat{u}_t / T$, to ensure that the bootstrap population mean of u_t^* is zero, consistent with the mean of u_t in the DGP being zero.

Kilian (1998b) compares parametric and nonparametric resampling schemes for u_t^* in vector autoregressions and shows that in practice there is no noticeable efficiency gain from imposing parametric assumptions even when that assumption is true, whereas imposing the wrong parametric structure tends to undermine the accuracy of bootstrap inference. Hence, the nonparametric approach is strictly preferred in practice.

Given the practical importance of drawing residuals at random with replacement, it is useful to briefly review the mechanics of this procedure. For expository purposes, consider a univariate AR(p) process with innovation u_t . The set of residuals \widehat{u}_t represents one sample realization of the iid process for u_t . Our objective is to replicate this sampling process conditional on the observed set of residuals. In practice, we start with the row index of the column vector \widehat{u} of residuals \widehat{u}_t . For example, for $T = 5$ we may have

$$\text{index} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix}, \quad \widehat{u} = \begin{pmatrix} 0.10 \\ -0.22 \\ 0.35 \\ 0.01 \\ -0.40 \end{pmatrix}.$$

We specify a uniform distribution over the set of integers $\{1, \dots, T\}$ in the index, ensuring that each residual has the same probability $1/T$ of being selected. For each bootstrap replication $r = 1, \dots, R$, of length T , we create a new index, index^{*r} , by scaling T draws from a uniform random number generator defined on the interval $[0, 1]$ by T and rounding the results such that the scaled random draws form a column vector of numbers between 1 and T . The elements of u^{*r} are the elements in \widehat{u} with row numbers given by index^{*r} .

In our example, we may obtain

$$\text{index}^{*r} = \begin{pmatrix} 5 \\ 2 \\ 4 \\ 3 \\ 5 \end{pmatrix} \Rightarrow u^{*r} = \begin{pmatrix} -0.40 \\ -0.22 \\ 0.01 \\ 0.35 \\ -0.40 \end{pmatrix}.$$

Note that this procedure destroys any stochastic dependence among the residuals \hat{u}_t and ensures that u_t^* is iid, consistent with the premise that u_t is iid. Although the support of the population distribution of u_t is continuous and that of u_t^* is finite, for reasonably large T , the approximation error may be safely ignored.

The ultimate objective of creating draws for u_t^* is to recreate the underlying sampling experiment that generated the observed sample $\{y_t\}_{t=-p+1}^T$, with the bootstrap data generating process serving as a proxy for the unknown DGP. This allows us to generate as many sequences of bootstrap realizations $\{y_t^*\}_{t=-p+1}^T$ as required to approximate the distribution of the statistic of interest. In practice, we use a finite number R of bootstrap replications of the sampling experiment, where R is so large that approximation error may be ignored. How large R should be depends on the statistic to be bootstrapped. For example, a bootstrap estimate of the bias or the variance of an estimator may require only $R = 200$, but estimating the 95th percentile of the distribution of an estimator tends to require at least $R = 2,000$. The reason is that there are fewer realizations in the tail of the distribution than in the center, necessitating more draws to ensure high enough accuracy.

Given random draws for u_t^{*r} , $t = 1, \dots, T$, and initial conditions $[y_{-p+1}^{*r}, \dots, y_0^{*r}]$, we recursively generate for each bootstrap replication $r = 1, \dots, R$, a sequence of bootstrap realizations $\{y_t^{*r}\}_{t=-p+1}^T$ as

$$\begin{aligned} y_1^{*r} &= \hat{v} + \hat{A}_1 y_0^{*r} + \dots + \hat{A}_p y_{-p+1}^{*r} + u_1^{*r}, \\ y_2^{*r} &= \hat{v} + \hat{A}_1 y_1^{*r} + \dots + \hat{A}_p y_{-p+2}^{*r} + u_2^{*r}, \\ &\vdots \\ y_T^{*r} &= \hat{v} + \hat{A}_1 y_{T-1}^{*r} + \dots + \hat{A}_p y_{T-p}^{*r} + u_T^{*r}, \end{aligned}$$

where the additional superscript r reminds us which of the R bootstrap replications this sequence of draws refers to.

The standard approach is to draw the initial conditions $[y_{-p+1}^*, \dots, y_0^*]$ at random with replacement as a block of p consecutive vector valued observations from the observed data $\{y_t\}_{t=-p+1}^T$. For each $r = 1, \dots, R$, a new draw for $[y_{-p+1}^*, \dots, y_0^*]$ is selected. This approach ensures that the bootstrap

presample observations are drawn from the same distribution as the remaining observations without making a parametric assumption about that distribution. Alternatively, the initial conditions may be fixed at their sample values, $[y_{-p+1}, \dots, y_0]$ for all $r = 1, \dots, R$. This will cause the sampling uncertainty in the statistic of interest to be understated, but in stationary models the approximation error declines with T and is negligible in the limit. Finally, one could also set the bootstrap presample observations equal to an arbitrary value such as the sample average $\bar{y} = \sum_{t=1}^T y_t / T$ for all $r = 1, \dots, R$. Because subsequent realizations of y_t^* depend on the initial values, in the latter case we need to generate $\tau + T$ observations of y_t^* , where τ is chosen to allow the effects of the choice of the initial values on y_t^* to become negligible, and discard the first τ observations. The choice of τ depends both on how far the initial values are from the steady state and how persistent the process is. In practice, τ may be chosen based on visual inspection of the bootstrap data.

In this example we treated the lag order as known. If the lag order p is unknown and has to be estimated from the same data set on which the final model is estimated, we need to reestimate the lag order for each bootstrap replication $r = 1, \dots, R$ to account for the sampling uncertainty in the lag-order estimate (see Kilian 1998a). It should be noted, however, that this approach does not overcome the problems with post-model selection inference discussed in Leeb and Pötscher (2005) (see Chapter 2).

Given a sequence of data $\{y_t^{*r}\}_{t=-p+1}^T$, for $r = 1, \dots, R$, we proceed exactly as we did when confronted with the observed sample of data $\{y_t\}_{t=-p+1}^T$. We fit the VAR(p) model with intercept to obtain the LS estimates $[\hat{v}^{*r}, \hat{A}_1^{*r}, \dots, \hat{A}_p^{*r}]$ and $\hat{\Sigma}_u^{*r}$. This allows us to construct the implied bootstrap estimates $\hat{\theta}_{ik,h}^{*r}$, $r = 1, \dots, R$, of the structural impulse responses, as discussed earlier. From the bootstrap approximation to the unknown empirical distribution of the structural impulse response estimates $\hat{\theta}_{ik,h}^{*r}$ we construct confidence intervals. Runkle (1987) in particular reported the percentile interval

$$[\hat{\theta}_{ik,h,\gamma/2}^*, \hat{\theta}_{ik,h,1-\gamma/2}^*],$$

where $\hat{\theta}_{ik,h,\gamma/2}^*$ and $\hat{\theta}_{ik,h,1-\gamma/2}^*$ are the critical points defined by the $\gamma/2$ and $1 - \gamma/2$ quantiles of the distribution of $\hat{\theta}_{ik,h}^{*r}$. Note that the validity of this approach hinges on the distribution of $\hat{\theta}_{ik,h}^{*r}$ becoming close to the asymptotic normal distribution of $\hat{\theta}_{ik,h}$, as the sample size goes to infinity. This will typically be the case in the stationary VAR(p) model with some exceptions noted below. We will return to the question of how to evaluate the distribution of the bootstrap estimator of interest in more detail later in the chapter. For now our focus is on the question of how to generate bootstrap approximations to the sampling distribution.

12.2.2 The Standard Residual-Based Fixed-Design Bootstrap

An alternative and less common approach is to treat the design matrix as fixed in repeated sampling. Under stationarity, ignoring the randomness of the regressors does not affect the asymptotic validity of the bootstrap. This allows us to approximate the unknown stationary VAR(p) data generating process

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where $u_t \stackrel{iid}{\sim} F$, conditional on the observed data, by the bootstrap DGP

$$y_t^* = \hat{v} + \hat{A}_1 y_{t-1} + \cdots + \hat{A}_p y_{t-p} + u_t^*$$

with $u_t^* \stackrel{iid}{\sim} \hat{F}_T$. Bootstrap samples are created by adding the resampled residuals to the fitted model, holding the regressor matrix fixed in repeated sampling. Given a sequence of data $\{y_t^{*r}\}_{t=-p+1}^T$, we regress y_t^{*r} on the original regressors $(1, y_{t-1}, \dots, y_{t-p})$ to obtain the least-squares estimates $[\hat{v}^{*r}, \hat{A}_1^{*r}, \dots, \hat{A}_p^{*r}]$ and $\hat{\Sigma}_u^{*r}$. After repeating this procedure for $r = 1, \dots, R$, we construct the empirical distribution of $\hat{\theta}_{ik,h}^{*r}$, $r = 1, \dots, R$, and the implied bootstrap confidence intervals as before. Simulation evidence in Gonçalves and Kilian (2004) suggests that this algorithm is slightly less accurate in finite samples than the recursive-design bootstrap for autoregressive processes.

12.2.3 The Residual-Based Wild Bootstrap

Both bootstrap approaches discussed so far require the regression error to be iid. This assumption is quite strong, especially when working with monthly data that are often characterized by conditional heteroskedasticity of unknown form. An alternative approach is based on the wild bootstrap. The wild bootstrap was originally developed in the context of the classical linear regression model (see Wu 1986). Its main use has been to make inference in cross-sectional regressions robust to unconditional heteroskedasticity in the error term (see Davidson and Flachaire 2008). Its asymptotic validity for the parameters of dynamic regression models with conditional heteroskedasticity was established in Gonçalves and Kilian (2004, 2007).

The wild bootstrap may be used in the context of fixed regressor bootstrap methods as well as recursive bootstrap methods. The only difference compared with the methods discussed earlier is how we create draws for u_t^* . The central idea is to multiply each element of the residual vector by a scalar draw, η_t , from an auxiliary distribution (or external distribution) that has mean zero and variance 1 such as the $\mathcal{N}(0, 1)$ distribution. Thus

$$u_t^{*r} = (u_{1t}^{*r}, u_{2t}^{*r}, \dots, u_{Kt}^{*r})'$$

where $u_t^{*r} \equiv \widehat{u}_t \eta_t$ and $\eta_t \stackrel{iid}{\sim} (0, 1)$. Other examples of possible η_t distributions include the two-point distribution $\eta_t = -(\sqrt{5} - 1)/2$ with probability $\mathcal{P} = (\sqrt{5} + 1)/(2\sqrt{5})$ and $\eta_t = (\sqrt{5} + 1)/2$ with probability $1 - \mathcal{P}$, as proposed by Mammen (1993), and the two-point distribution $\eta_t = 1$ with probability $\mathcal{P} = 0.5$ and $\eta_t = -1$ with probability $\mathcal{P} = 0.5$, as proposed by Liu (1988). There is no indication that the choice of η_t makes a difference in practice (see Gonçalves and Kilian 2004).

When applying the wild bootstrap, there is no need to demean the residuals even when no intercept is included in the regression model. The vector \widehat{u}_t is treated as a constant in the bootstrap world. Hence, conditional on the data, $\mathbb{E}^*(u_t^*) = 0$ and $V(u_t^*) = \mathbb{E}^*(u_t^{*2}) = \widehat{u}_t^2 \cdot 1 = \widehat{u}_t^2$, where \mathbb{E}^* is the expectation under the bootstrap probability measure. This feature of the wild bootstrap is closely related to the robust variance estimator proposed by Nicholls and Pagan (1983). The resulting bootstrap distributions may be evaluated in the usual manner. Any t -statistics or Wald statistics based on the wild bootstrap, of course, will have to be computed based on heteroskedasticity-robust standard errors, not the usual LS standard errors.

The wild bootstrap approach is preferable to modeling the error term of the VAR model as a parametric multivariate GARCH model or stochastic volatility model. Multivariate GARCH models, for example, typically require strong and ad hoc restrictions on the conditional covariance matrix for estimation to be computationally feasible. Moreover, there are many different GARCH and stochastic volatility specifications that are mutually exclusive. The wild bootstrap is nonparametric, it accounts for shifts in the volatility of the data, it avoids ad hoc modeling choices, and it has high finite-sample accuracy even when the regression error is iid (see Gonçalves and Kilian 2004, 2007).

The residual-based wild bootstrap was originally designed to approximate the distribution of the autoregressive slope parameters. One important drawback of this method is that it will recover the asymptotic distribution of the reduced-form impulse responses in the limit, but not the asymptotic distribution of the error variance-covariance matrix, as shown in Brüggemann, Jentsch, and Trenkler (2016). Intuitively, the problem is that the wild bootstrap does not correctly replicate the relevant fourth-moment structure of the VAR innovations. As a result, the wild bootstrap fails to recover the second component of the limiting variance in

$$\sqrt{T} (g(\widehat{\alpha}, \widehat{\sigma}) - g(\alpha, \sigma)) \xrightarrow{d} \mathcal{N} \left(0, \frac{\partial g}{\partial \alpha'} \Sigma_{\widehat{\alpha}} \frac{\partial g'}{\partial \alpha} + \frac{\partial g}{\partial \sigma'} \Sigma_{\widehat{\sigma}} \frac{\partial g'}{\partial \sigma} \right).$$

Because the wild bootstrap underestimates the asymptotic variance of the error variance, for statistics of the form $g(\widehat{\alpha}, \widehat{\sigma})$ it produces confidence intervals that are too narrow and have coverage accuracy that is too low asymptotically. It is

worth pointing out that, in practice, the uncertainty about the error covariance matrix quickly dies out as the horizon lengthens, such that the residual-based wild bootstrap adequately captures the uncertainty about structural impulse responses at all but the lowest horizons. Although the residual-based wild bootstrap is a clear improvement on the iid residual-based bootstrap in the presence of conditional heteroskedasticity, it, nevertheless, is unsatisfactory. Brüggemann, Jentsch, and Trenkler (2016) propose an alternative bootstrap method intended to ensure high asymptotic coverage accuracy for structural impulse responses at all horizons that will be discussed in the section on block bootstrap methods.

12.2.4 Bootstrapping Tuples of Regressands and Regressors

An alternative approach that is robust to heteroskedasticity of unknown form involves bootstrapping tuples consisting of the regressand and the regressors. This method is a natural generalization of the pairwise bootstrap for linear regression first suggested by Freedman (1981). Its asymptotic validity for autoregressions is established in Gonçalves and Kilian (2004).

Consider the unknown stationary VAR(p) DGP

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t.$$

It is clear that when the data are dependent over time, one cannot resample each date t observation by drawing with replacement without destroying the dependence structure in the data. This problem may be circumvented by forming tuples

$$x_t \equiv (y'_t, 1, y'_{t-1}, \dots, y'_{t-p}),$$

of dimension $1 \times (Kp + K + 1)$ for $t = 1, \dots, T$ and stacking them into a matrix X . Note that x_t contains a 1 to account for the intercept term in the model. As long as the regression error is not serially correlated, we can draw with replacement from the T rows of the matrix X to form a bootstrap sample matrix X^{*r} with elements

$$x_t^{*r} \equiv (y_t^{*r}, 1, y_{t-1}^{*r}, \dots, y_{t-p}^{*r}), \quad t = 1, \dots, T.$$

The first K columns of the resulting matrix X^{*r} correspond to the $T \times K$ dependent variable $\{y_t^{*r}\}_{t=1}^T$, while the remaining columns form the $T \times (Kp + 1)$ matrix of bootstrap regressors. This is all we need to form the bootstrap LS estimates $[\hat{v}^{*r}, \hat{A}_1^{*r}, \dots, \hat{A}_p^{*r}]$ and $\hat{\Sigma}_u^{*r}$. After repeating this procedure for $r = 1, \dots, R$, we construct the empirical distribution of $\hat{\theta}_{ik,h}^{*r}$, $r = 1, \dots, R$, and the implied bootstrap confidence intervals as before.

Note that we never specified the error term u_t^* in this algorithm. This means that this algorithm will be valid for regression models with white noise errors,

even if the error term is not iid, but conditionally heteroskedastic. Of course, methods that exploit the iid assumption when it is true in population will tend to be more efficient and hence more accurate in small samples. Like the fixed-design bootstrap, this alternative bootstrap is less parametric and hence slightly less accurate when the autoregressive model is the data generating process, as shown in Gonçalves and Kilian (2004). Like the wild bootstrap, however, it fails to capture the full uncertainty in the residual covariance matrix and hence is not designed for inference on structural impulse responses (see Brüggemann, Jentsch, and Trenkler 2016).

12.2.5 Block Bootstrap Methods

If the innovations of the model are serially correlated, none of the preceding bootstrap methods are applicable. An obvious response would be to increase the VAR lag order appropriately before applying conventional bootstrap methods. An alternative approach, designed to protect against serial correlation in the error term, is the use of block bootstrap methods. Block bootstrap methods require the underlying DGP to be $I(0)$. It may seem that the use of the block bootstrap would always be preferred, given its robustness to departures from the iid error assumption, but this is not the case because this robustness comes at a cost. For example, the use of block bootstrap methods would be inefficient if we were concerned with serial correlation of known form in the error term.

There are different types of block bootstrap methods in the literature, only some of which are relevant for working with VAR models. In explaining the block bootstrap and how it may be applied in the context of VAR models, it is useful to start with a much simpler example.

Naive Block Bootstrap. The most common form of the bootstrap is based on the exchangeability of iid observations. Usually, in the case of dependent data, we parameterize the dependence by imposing an autoregressive structure, so that the problem can be reduced to one of resampling residuals that are iid in population.

An alternative strategy is to re-sample blocks of observations rather than individual observations. This simplest block bootstrap approach is asymptotically valid for statistics that are symmetric functions of time series observations generated from covariance stationary processes. Symmetry here means that the statistic does not depend on the order of the observations. A good example is the sample mean:

$$\bar{y} = \frac{1}{T} \sum_{t=1}^T y_t.$$

A good counterexample is the autocorrelation coefficient. To cement ideas, we begin with a simple example. Consider a sample y_1, \dots, y_T generated by a covariance stationary scalar process $\{y_t\}_{t=-\infty}^{\infty}$, which may or may not be an autoregressive process. The procedure involves three simple steps:

1. Suppose that we observe the sample

$$(y_1, y_2, y_3, y_4, y_5, y_6, y_7, y_8, y_9, y_{10}, y_{11}, y_{12}, y_{13}, y_{14}, y_{15})'$$

We begin by defining $b = T - l + 1$ overlapping blocks of length l .¹ Let $l = 5$. Then $b = 15 - 5 + 1 = 11$ and the set of blocks we will draw from is:

$$\begin{aligned}x_1 &= (y_1, y_2, y_3, y_4, y_5) \\x_2 &= (y_2, y_3, y_4, y_5, y_6) \\x_3 &= (y_3, y_4, y_5, y_6, y_7) \\x_4 &= (y_4, y_5, y_6, y_7, y_8) \\\vdots \\x_9 &= (y_9, y_{10}, y_{11}, y_{12}, y_{13}) \\x_{10} &= (y_{10}, y_{11}, y_{12}, y_{13}, y_{14}) \\x_{11} &= (y_{11}, y_{12}, y_{13}, y_{14}, y_{15})\end{aligned}$$

Bootstrap replicates of the original sample may be generated by sampling with replacement from the set $\{x_t\}_{t=1}^b$. We then concatenate the iid draws, denoted by x_t^* , to form a bootstrap replication of the original time series of length $T = sl$ of the form $\{y_t^*\}_{t=1}^T = (x_1^*, x_2^*, \dots, x_s^*)'$, where $s = T/l$. For example,

$$(y_9, y_{10}, y_{11}, y_{12}, y_{13}|y_2, y_3, y_4, y_5, y_6|y_5, y_6, y_7, y_8, y_9)'$$

could be the r^{th} bootstrap replication of the original time series. In the example, $s = 15/5 = 3$, so each bootstrap time series of length T consists of three blocks.

2. Given this bootstrap time series, we construct

$$\bar{y}^{*r} = \frac{1}{T} \sum_{t=1}^T y_t^*.$$

3. This procedure is repeated for $r = 1, \dots, R$, allowing us to build up the bootstrap distribution of the sample mean.

¹ Although we focus on the case of overlapping blocks, which is most common in practice, it should be noted that block bootstrap methods can also be applied to nonoverlapping blocks. The latter approach makes less efficient use of the data, however.

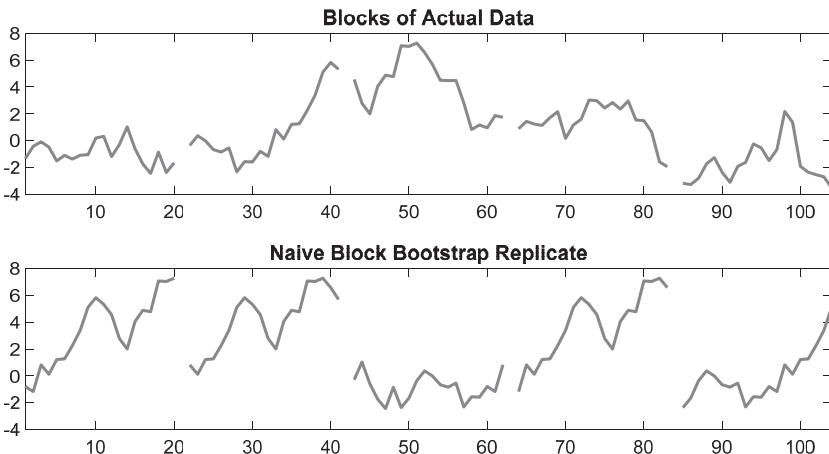


Figure 12.1. Example of how resampling overlapping blocks may destroy the dependence structure in the bootstrap data at the point of transition from one block to the next.

Notes: The data were generated from a stationary AR(1) model. The plot was created by inserting an extra space in between any two blocks to allow visual separation of each block.

This procedure was originally introduced by Künsch (1989). The asymptotic validity of this bootstrap approach for stationary data requires $s \rightarrow \infty$ and $l \rightarrow \infty$ at a suitable rate, as $T \rightarrow \infty$, where s is the number of blocks of length l such that $T = sl$.

Limitations of the Naive Block Bootstrap. It is easy to see why this idea does not generalize easily to statistics such as autoregressive coefficients that are not symmetric functions of the data. For illustrative purposes, consider the time series shown in the upper panel of Figure 12.1, which was generated from an AR(1) process with a parameter of 0.95. The iid innovations of this process were drawn from a $\mathcal{N}(0, 1)$ distribution. The sample size is $T = 100$. For expository purposes the sample has been divided into five blocks of length $l = 20$.

The lower panel of Figure 12.1 displays a representative replicate of bootstrap data obtained by drawing with replacement from the set of overlapping blocks of length $l = 20$ constructed from the original time series. This figure illustrates a major problem with the naive block bootstrap applied to statistics such as autocorrelation coefficients or autoregressive slope coefficients. As we paste one block of data to the next, we destroy the time series dependence. Clearly, at these points there are discontinuous jumps in the time series not found in the original data, so the autocorrelation structure of the bootstrap data differs systematically from that found in the original DGP.

If one lets the block length l increase with the sample size, this problem can be mitigated at least asymptotically. This is of little help, however, when the sample is short and the data are persistent. On the one hand, we need the blocks to be long enough for the original time series dependence to be preserved in the resampled series. On the other hand, in short samples, large l makes for few estimates of the statistic of interest and hence poor approximations to the empirical distribution. Thus the researcher finds himself between a rock and a hard place.

Points of discontinuity are not a problem for statistics such as the sample mean that do not depend on the order of the time series observations, but they call into question applications of the naive block bootstrap to nonsymmetric statistics such as autocorrelations or estimators of autoregressive slope coefficients. There are two modifications of the original block bootstrap idea designed to deal with this problem. One is the blocks-of-blocks bootstrap; the other is the residual based block bootstrap.

Blocks-of-Blocks Bootstrap. Nonsymmetric statistics such as autocorrelations or estimators of autoregressive slope coefficients can be expressed as a symmetric function of m -dimensional blocks of consecutive observations, $m > 1$. This fact suggests that, when dealing with nonsymmetric statistics, we resample blocks of l consecutive m -tuples drawn from the set of all possible m -tuples. This amounts to a blocks-of-blocks bootstrap. The asymptotic validity of this procedure for stationary processes requires letting the block size l increase at a suitable rate, as $T \rightarrow \infty$.

Suppose that the sample data $\{y_t\}_{t=1}^T$ have already been demeaned. We are interested in the first-order autocorrelation coefficient which depends on y_t and y_{t-1} . This fact tells us that $m = 2$. Further suppose that $T = 16$. Construct the (15×2) matrix of m -tuples by stacking y_t next to y_{t-1} :

$$X = \begin{bmatrix} 0.10 & -0.3 \\ -0.5 & 0.1 \\ 0.4 & -0.5 \\ 0 & 0.4 \\ -0.1 & 0 \\ \vdots & \vdots \\ -0.2 & 0.11 \\ 0.05 & -0.2 \\ 0.1 & 0.05 \end{bmatrix}.$$

Rather than drawing with replacement from the rows of this matrix, we draw with replacement from the set of overlapping blocks of rows of this matrix. For

example, for $l = 3$, we might draw a bootstrap replicate:

$$\text{index}^* = \begin{pmatrix} 13 \\ 14 \\ \underline{15} \\ 2 \\ 3 \\ \underline{4} \\ 3 \\ 4 \\ 5 \end{pmatrix} \Rightarrow X^* = \begin{bmatrix} -0.2 & 0.11 \\ 0.05 & -0.2 \\ 0.1 & 0.05 \\ -0.5 & 0.1 \\ 0.4 & -0.5 \\ 0 & 0.4 \\ 0.4 & -0.5 \\ 0 & 0.4 \\ -0.1 & 0 \end{bmatrix}.$$

Having generated X^* , we substitute the first and second column of X^* into the formula for the first-order autocorrelation coefficient to generate the r^{th} replicate of that statistic. Note that when we use the blocks-of-blocks bootstrap, we never actually construct the bootstrap realizations of the data $\{y_t^*\}_{t=1}^T$. Rather, we directly substitute the re-sample into the function that defines the statistic of interest in terms of the m -tuples.

In applications of the blocks-of-blocks bootstrap, the statistic of interest determines the choice of m . In the example of the first-order autocorrelation, $m = 2$. When considering an AR(3) model, on the other hand, we would have $m = 4$. This leaves the task of choosing l . This problem is analogous to the choice of the truncation lag for heteroskedasticity and autocorrelation consistent (HAC) standard errors (see Andrews 1991). There are no good rules of thumb for choosing l in general. In practice, it is advisable to check the sensitivity of the results with respect to l .

The blocks-of-blocks bootstrap idea may be easily generalized to the problem of bootstrapping structural impulse responses in vector autoregressions. In that case

$$X = \begin{bmatrix} y'_1 & 1 & y'_0 & \dots & y'_{-p+1} \\ y'_2 & 1 & y'_1 & \dots & y'_{-p+2} \\ \vdots & \vdots & \vdots & & \vdots \\ y'_T & 1 & y'_{T-1} & \dots & y'_{T-p} \end{bmatrix}.$$

It is convenient to include the intercept in X , such that X is of dimension $T \times (Kp + K + 1)$. The choice of l depends on how much serial correlation we expect in the error term. Typically, in VAR(p) models with large p , there is not much serial correlation left in the error term, so l tends to be small. Having determined l , we generate X^* by drawing with replacement from blocks of l adjacent rows of X . Note that this approach may also be viewed as a generalization of the earlier idea of bootstrapping tuples of regressands and

regressors that allows for serial correlation in the error term of an autoregression.

Given X^{*r} , we recover the regressand and regressors, reestimate the VAR(p) model on the bootstrap data, and compute $\widehat{\theta}_{ik,h}^{*r}$ from $[\widehat{v}^{*r}, \widehat{A}_1^{*r}, \dots, \widehat{A}_p^{*r}]$ and $\widehat{\Sigma}_u^{*r}$. After repeating this procedure for $r = 1, \dots, R$, we construct the empirical distribution of $\widehat{\theta}_{ik,h}^{*r}$, $r = 1, \dots, R$, and the implied bootstrap confidence intervals as before. Although the implementation of this method is straightforward, we are not aware of applications to vector autoregressions.

The block bootstrap is based on blocks of fixed length l . The stationary bootstrap is a variation of the block bootstrap with the block size l modeled as a random variable with a geometric distribution with mean l . The stationary bootstrap in practice does not improve the accuracy of the block bootstrap approximation for autoregressions and may worsen it in small samples (see Berkowitz, Birgean, and Kilian 2000).

Residual-Based Block Bootstrap. An alternative approach is to apply the original block bootstrap not to the data directly, but rather to the residuals of an autoregression. The idea of the residual-based block bootstrap can be traced to Lahiri (1996). This approach has also been referred to as a pre-whitened bootstrap because fitting the autoregression removes much of the persistence of the data that calls into question the application of the original block bootstrap (see, e.g., Berkowitz, Birgean, and Kilian 2000). Although a number of closely related approaches have been discussed in the literature, there is no reference that explicitly discusses this method and derives its asymptotic properties for the stationary VAR(p) model.

This residual block bootstrap procedure may be summarized as follows. Consider the VAR approximation

$$y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t,$$

where u_t is assumed to be a linear stationary process with innovations that are mean zero and satisfy suitable restrictions on their moments. The error u_t is also assumed to satisfy a suitable strong mixing condition (see, e.g., Jentsch, Politis, and Paparoditis 2015).² Having fitted the VAR(p) model to the data and having recovered estimates of the model parameters and of the model residuals

$$\widehat{u}_t = y_t - \widehat{v} - \widehat{A}_1 y_{t-1} - \dots - \widehat{A}_p y_{t-p}$$

² Strong mixing conditions ensure a sufficiently fast rate of decay in the dependence between u_t and u_s as the distance between t and s grows, such that observations sufficiently far apart in time are nearly independent (see Davidson 1994).

for $t = 1, \dots, T$, we arrange the residuals in the form of a matrix

$$\begin{bmatrix} \widehat{u}_1 & \widehat{u}_2 & \dots & \widehat{u}_l \\ \widehat{u}_2 & \widehat{u}_3 & \dots & \widehat{u}_{l+1} \\ \vdots & \vdots & & \vdots \\ \widehat{u}_{T-l+1} & \widehat{u}_{T-l+2} & \dots & \widehat{u}_T \end{bmatrix},$$

where each row denotes a block of l consecutive residuals and the blocks are overlapping. The number of blocks is $s = T - l + 1$. The bootstrap innovations are obtained by drawing at random with replacement from the rows of this matrix, laying these blocks of residuals end to end, and retaining the first T bootstrap innovations.

The resulting bootstrap innovation draws, denoted by $[\widetilde{u}_1^*, \dots, \widetilde{u}_T^*]$, however, must be recentered to ensure that $\mathbb{E}^*(u_t^*) = 0$ for all $t = 1, \dots, T$, where \mathbb{E}^* denotes the expectation under the bootstrap probability measure. The reason is that the bootstrap innovations must have the same population mean as the original population innovations for the bootstrap analogy to hold. This may be accomplished by constructing the bootstrap innovations, $[u_1^*, \dots, u_T^*]'$, as

$$u_{jl+i}^* = \widetilde{u}_{jl+i}^* - \mathbb{E}^*(\widetilde{u}_{jl+i}^*) = \widetilde{u}_{jl+i}^* - \frac{1}{T-l+1} \sum_{r=0}^{T-l} \widehat{u}_{i+r}$$

for $i = 1, 2, \dots, l$ and $j = 0, 1, \dots, s - 1$. The corresponding bootstrap data, $[y_{-p+1}^*, \dots, y_T^*]$, are then generated recursively from

$$y_t^* = \widehat{\nu} + \widehat{A}_1 y_{t-1}^* + \dots + \widehat{A}_p y_{t-p}^* + u_t^*.$$

Bootstrap realizations of the estimators of interest are obtained by fitting a VAR(p) model to this sequence of bootstrap data. Finite-sample evidence for this approach can be found in Berkowitz, Birgean, and Kilian (2000).³

A generalization of the residual-based block bootstrap idea was proposed by Yeh (1998). The residual-based wild block bootstrap allows for the errors of the regression model to be nonstationary. Nonstationarity here does not refer to allowing for unit roots in the process governing u_t , but to allowing for heterogeneity in the innovations across time. An example would be a serially correlated error process subject to unconditional heteroskedasticity. The theoretical results in Yeh (1998) are limited to a fixed regression design and nonoverlapping blocks of residuals. As before, given model residuals

$$\widehat{u}_t = y_t - \widehat{\nu} - \widehat{A}_1 y_{t-1} - \dots - \widehat{A}_p y_{t-p}$$

³ Jentsch, Politis, and Paparoditis (2015) discuss extensions of the block bootstrap approach to multivariate integrated and cointegrated models. There are, at present, no formal results, however, about the validity of conducting inference about structural impulse responses in cointegrated VAR models based on the residual-based block bootstrap.

for $t = 1, \dots, T$, we arrange the residuals in the form of a matrix

$$\begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \dots & \hat{u}_l \\ \hat{u}_{l+1} & \hat{u}_{l+2} & \dots & \hat{u}_{2l} \\ \vdots & \vdots & & \vdots \\ \hat{u}_{T-l+1} & \hat{u}_{T-l+2} & \dots & \hat{u}_T \end{bmatrix},$$

where each row denotes a block of l consecutive residuals, but now the blocks are nonoverlapping and the number of blocks is $s = [T/l]$, where s denotes the largest integer smaller or equal to T/l .

Bootstrap innovations are created by post-multiplying each row of this matrix by an independent draw from an external random variable with mean zero and variance 1 such as a draw from a $\mathcal{N}(0, 1)$ distribution. Let η_1, \dots, η_s denote the sequence of realizations of these $\mathcal{N}(0, 1)$ draws. Then the bootstrap innovations, $[u_1^*, \dots, u_T^*]$, are obtained by laying the rows of the matrix

$$\begin{bmatrix} \hat{u}_1\eta_1 & \hat{u}_2\eta_1 & \dots & \hat{u}_l\eta_1 \\ \hat{u}_{l+1}\eta_2 & \hat{u}_{l+2}\eta_2 & \dots & \hat{u}_{2l}\eta_2 \\ \vdots & \vdots & & \vdots \\ \hat{u}_{T-l+1}\eta_s & \hat{u}_{T-l+2}\eta_s & \dots & \hat{u}_T\eta_s \end{bmatrix},$$

end to end. No centering is required. The corresponding bootstrap data, $[y_{-p+1}^*, \dots, y_T^*]$, are generated from

$$y_t^* = \hat{v} + \hat{A}_1 y_{t-1} + \dots + \hat{A}_p y_{t-p} + u_t^*,$$

treating the regression design as fixed in repeated sampling. Bootstrap realizations of the estimators of interest are obtained by regressing $\{y_t^*\}_{t=1}^T$ on the original regressor matrix. Unlike the residual-based block bootstrap, this procedure is not applicable to structural impulse response analysis. It only captures the uncertainty about the VAR slope parameters (and smooth functions thereof).

The Residual-Based Block Bootstrap in Conditionally Heteroskedastic VAR Models. Brüggemann, Jentsch, and Trenkler (2016) show that the residual-based bootstrap may also be used to achieve accurate inference about structural impulse responses in stationary VAR models with conditionally heteroskedastic errors, but no serial correlation in the errors. A precise statement of these conditions can be found in their paper. Their proposal involves fitting to the data the VAR(p) model

$$y_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t,$$

where u_t is serially uncorrelated, but conditionally heteroskedastic, and to recover estimates of the model parameters and the model residuals

$$\hat{u}_t = y_t - \hat{\nu} - \hat{A}_1 y_{t-1} - \cdots - \hat{A}_p y_{t-p}$$

for $t = 1, \dots, T$. They then choose a block length $l < T$ such that $s = [T/l]$ is the number of blocks where $ls \geq T$, and arrange blocks of length l of the model residuals in the form of the matrix

$$\begin{bmatrix} \hat{u}_1 & \hat{u}_2 & \cdots & \hat{u}_l \\ \hat{u}_2 & \hat{u}_3 & \cdots & \hat{u}_{l+1} \\ \vdots & \vdots & & \vdots \\ \hat{u}_{T-l+1} & \hat{u}_{T-l+2} & \cdots & \hat{u}_T \end{bmatrix}.$$

Bootstrap innovations are generated by drawing s times with replacement from the rows of this matrix. The time series of bootstrap innovations, $[\hat{u}_1^*, \dots, \hat{u}_T^*]$, is obtained by laying these randomly chosen blocks end to end and retaining the first T bootstrap innovations. The resulting bootstrap innovation draws again must be recentered to ensure that $\mathbb{E}^*(\hat{u}_t^*) = 0$ for all $t = 1, \dots, T$. We construct the bootstrap innovations, $[\hat{u}_1^*, \dots, \hat{u}_T^*]$, from

$$u_{jl+i}^* = \hat{u}_{jl+i}^* - \mathbb{E}^*(\hat{u}_{jl+i}^*) = \hat{u}_{jl+i}^* - \frac{1}{T-l+1} \sum_{r=0}^{T-l} \hat{u}_{i+r}$$

for $i = 1, 2, \dots, l$ and $j = 0, 1, \dots, s-1$. Given an appropriate draw for the bootstrap presample observations, one can generate the bootstrap sample, $[y_{-p+1}^*, \dots, y_T^*]$, from

$$y_t^* = \hat{\nu} + \hat{A}_1 y_{t-1}^* + \cdots + \hat{A}_p y_{t-p}^* + u_t^*,$$

and compute the bootstrap estimators of interest after fitting a VAR(p) model to this sequence of data. Simulation evidence suggests that the coverage accuracy of structural impulse response confidence intervals based on the residual-based block bootstrap method is consistently high only in fairly large samples.

12.2.6 Alternative Bootstrap Confidence Intervals

So far we have focused mainly on the question of how to generate bootstrap approximations to the distribution of the statistic of interest. An important question is how to construct bootstrap confidence intervals from this distribution. There are a number of options ranging from bootstrap standard error bands to percentile and percentile- t intervals, all of which could be employed in conjunction with any of the resampling methods discussed so far.

For notational simplicity, in what follows, we refer to the scalar structural VAR impulse response coefficient $\theta_{jk,i}$ as θ , suppressing the subscripts.

$\widehat{\theta}$ denotes a consistent and asymptotically normal estimator of this scalar parameter θ , and $\widehat{\theta}^*$ is the bootstrap version of this estimator.

Intervals Based on Bootstrap Standard Errors. A particularly simple approach is to compute the standard deviation of the bootstrap draws of $\widehat{\theta}^*$ numerically. This allows one to form confidence intervals that are centered on the point estimate of the structural impulse response:

$$\widehat{\theta} \pm z_{\gamma/2} \widehat{\sigma}(\widehat{\theta}^*).$$

Effectively, the only difference to the delta method interval is that we have replaced the closed-form estimator of the standard deviation. The bootstrap estimator and the closed-form estimator are asymptotically equivalent under Gaussian white noise errors. One advantage of the bootstrap estimator is that it allows us to relax the assumption of Gaussian iid innovations underlying the conventional delta method interval, for example, by resampling the residuals with replacement or by using the residual-based block bootstrap method. The other advantage is that the bootstrap estimator may be used even in the absence of a closed-form solution for the asymptotic variance.

Efron's Percentile Interval. The most common form of bootstrap confidence intervals is the percentile interval proposed by Efron (1979), which is also the interval employed in Runkle (1987). Because subsequently, alternative “percentile intervals” have been proposed in the literature, we refer to this interval as Efron's percentile interval. Let $\widehat{\theta}_{\gamma/2}^*$ and $\widehat{\theta}_{1-\gamma/2}^*$ be the critical points defined by the $\gamma/2$ and $1 - \gamma/2$ quantiles of the distribution of $\widehat{\theta}^*$. Then Efron's percentile interval is

$$CI_{PER}^{\text{Efron}} = [\widehat{\theta}_{\gamma/2}^*, \widehat{\theta}_{1-\gamma/2}^*].$$

This interval was designed to be asymptotically valid for inference about consistent estimators with an asymptotic normal approximation of the form

$$\widehat{\theta} \sim \mathcal{N}(\theta, \sigma^2(\widehat{\theta})).$$

It is readily apparent that the interval endpoints of the percentile interval will agree with the interval endpoints of the delta method interval asymptotically, if the statistic of interest is asymptotically Gaussian and centered on its population value. As stressed by Efron, however, this percentile interval (as well as its variants discussed later) remains valid even when the distribution of $\widehat{\theta}$ is not normal, as long as normality can be achieved by applying a monotonic transformation to θ and to its estimator $\widehat{\theta}$.⁴ The percentile interval automatically recovers the appropriate interval endpoints without requiring the user to

⁴ An example would be Fisher's inverse hyperbolic tangent transformation for the correlation coefficient from a bivariate normal sample (see Efron 1982).

know the correct transformation. In this sense, it is more general than conventional asymptotic normal approximations. If there is no suitable transformation that satisfies this condition, in contrast, as would be the case when the statistic of interest is biased, the CI_{PER}^{Efron} -interval will not be valid and may have poor coverage accuracy.

It immediately follows that caution is called for in applying this method when conducting inference about structural impulse responses. The finite-sample distribution of structural impulse response estimators, $\widehat{\theta}_{ik,h}$, may be far from Gaussian (see Kilian 1999b). A particular concern is the small-sample bias of the structural impulse response estimator. When modeling autoregressions, bias is not a concern asymptotically, because the small-sample bias of the least-squares estimator is of order T (see Section 2.3.3), but it is an obvious concern for the finite-sample accuracy of the CI_{PER}^{Efron} -interval, which is based on the premise of an unbiased estimator.

Efron (1982) proposed a generalization of the original percentile interval intended to allow for bias in the estimator, while retaining the assumption of approximate normality (possibly after a monotonic transformation), referred to as the bias-corrected or BC percentile interval. His proposal was to adjust the endpoints of the percentile interval depending on the median bias of the estimator $\widehat{\theta}$. This solution presumes that we are dealing with a pure location shift and that the variance of the estimator is invariant with respect to the value of θ . This assumption is clearly violated when dealing with smooth functions of slope parameters in vector autoregressions, suggesting that such bias corrections will not help in our context. Indeed, Kilian (1998c) showed by simulation that the BC -interval is not reliable in small samples when conducting inference about structural impulse responses.

In related work, Efron (1987) therefore proposed a further refinement of the percentile interval, referred to as the accelerated BC -interval or BC_α interval (to which an analytic approximation was developed in later years under the name of ABC interval). Compared with the BC -interval, the BC_α -interval involves an additional acceleration constant, α , that captures changes in the variance of the estimator, as the location of the distribution changes. The BC_α -interval was designed for inference on statistics of the form

$$\frac{\widehat{\theta} - \theta}{\sigma(\widehat{\theta})} \sim \mathcal{N}(-z_0, 1),$$

where $\sigma(\widehat{\theta}) = \sigma(\theta_0)[1 + \alpha(\widehat{\theta} - \theta_0)/\sigma(\theta_0)]$ and θ_0 is any convenient reference point on the scale of θ values. Let $\Phi(\cdot)$ denote the standard normal cumulative distribution function and z_γ denote the $100\gamma^{\text{th}}$ percentile point of the standard normal distribution. Then the BC_α -interval is defined as

$$CI_{BC_\alpha} = [\widehat{\theta}_{\gamma_1}^*, \widehat{\theta}_{\gamma_2}^*],$$

with

$$\gamma_1 = \Phi \left(\widehat{z}_0 + \frac{\widehat{z}_0 + z_{\gamma/2}}{1 - \widehat{\alpha}(\widehat{z}_0 + z_{\gamma/2})} \right)$$

and

$$\gamma_2 = \Phi \left(\widehat{z}_0 + \frac{\widehat{z}_0 + z_{1-\gamma/2}}{1 - \widehat{\alpha}(\widehat{z}_0 + z_{1-\gamma/2})} \right),$$

where the bias-correction term z_0 may be estimated as

$$\widehat{z}_0 = \Phi^{-1}(f),$$

with f denoting the fraction of bootstrap estimates $\widehat{\theta}^*$ smaller than $\widehat{\theta}$ and $\Phi^{-1}(\cdot)$ denoting the inverse function of $\Phi(\cdot)$ such that, for example, $\Phi^{-1}(0.95) = 1.645$. The acceleration constant α corrects for the skewness of the distribution of the estimator and is typically estimated by the delete-one jackknife method as

$$\widehat{\alpha} = \frac{\sum_i (\widehat{\theta}_{(\cdot)} - \widehat{\theta}_{(i)})^3}{6 \left\{ \sum_i (\widehat{\theta}_{(\cdot)} - \widehat{\theta}_{(i)})^2 \right\}^{3/2}},$$

where $\widehat{\theta}_{(i)}$ denotes the estimator obtained by deleting the i^{th} observation from the original sample and $\widehat{\theta}_{(\cdot)}$ is the average of $\widehat{\theta}_{(i)}$ over all elements i of the sample. If $\alpha = 0$, the BC_α -interval reduces to the BC -interval, and if both $\alpha = 0$ and $z_0 = 0$, the BC_α -interval reduces to the original percentile interval. For further discussion the reader is referred to Efron and Tibshirani (1993).

The delete-one jackknife method utilized by Efron to estimate α is based on the assumption of independent observations. This assumption is violated in autoregressive processes. One alternative may be to estimate $\widehat{\alpha}$ instead based on the block-jackknife method proposed in Künsch (1989) for the analysis of time-dependent stationary observations. Alternatively, the jackknife estimator $\widehat{\theta}_{(i)}$ may be constructed by deleting the i^{th} row of the matrix X obtained by stacking

$$x_t \equiv (y'_t, 1, y'_{t-1}, \dots, y'_{t-p}), \quad t = 1, \dots, T,$$

under the assumption of an autoregressive DGP with independent errors. None of these alternatives have been explored in the literature to date. As of now, the BC_α -interval does not appear to have been used for inference in autoregressive models.

Applied researchers sometimes utilize the original percentile interval ignoring the bias of the impulse response estimator altogether. This approach cannot be recommended in general. Kilian (1998c) demonstrates that the standard residual-based bootstrap based on the percentile interval, as implemented by

Runkle (1987), tends to suffer from severe coverage deficiencies in practice, when the data are highly persistent. Its coverage accuracy is much lower than that of the delta method interval, in fact. It is not uncommon for a nominal 95% interval to have effective coverage as low as 10%. A tell-tale sign of these bias problems in practice is that the impulse response estimate often lies outside of the 95% bootstrap confidence interval.

The reason is that the central premise of Runkle's bootstrap method, namely that the least-squares estimates of the VAR model parameters are not systematically different from their true values, is violated in typical applications. For expository purposes, consider the AR(1) model. Suppose that

$$y_t = v + ay_{t-1} + u_t,$$

where in population $a = 0.98$ and $T = 100$. The bootstrap approximation is based on

$$y_t^* = \hat{v} + \hat{a}y_{t-1}^* + u_t^*,$$

where $\mathbb{E}(\hat{a} - a) < 0$ due to the small-sample bias of the LS estimator. In expectation, $a = 0.98$ is associated with an estimate of $\hat{a} = 0.95$ (see Shaman and Stine 1988). This means that the bootstrap DGP is centered on the wrong location. More importantly, the bootstrap estimates \hat{a}^* conditional on this biased DGP are in turn downward biased relative to \hat{a} , amplifying the overall bias in the bootstrap estimator. As a result, in expectation, the bootstrap estimate \hat{a}^* is only about 0.92.⁵ Put differently, the bootstrap method, when applied inappropriately, pulls down the distribution of the estimator even further away from a , which causes the coverage rate of these bootstrap confidence intervals to become very low in practice. In contrast, the delta method, while also centered on the initial biased estimate of 0.95, imposes symmetry on the confidence intervals, preventing the second-round bias and resulting in higher, if still inadequate, coverage accuracy.

This insight goes a long way toward explaining the coverage deficiencies of the Runkle (1987) bootstrap intervals. In practice, there are two potential remedies for this problem in the stationary model. One approach is the bias-adjusted bootstrap method of Kilian (1998c) applied in conjunction with the traditional percentile interval. Another approach is to rely on alternative bootstrap confidence intervals that allow for bias and asymmetry in the distribution of the estimator without requiring explicit bias estimates. Both approaches are discussed later in this chapter.

Equal-Tailed Percentile-*t* Intervals. There are two types of percentile-*t* intervals. One was proposed by Efron (1982). This equal-tailed percentile-*t* interval

⁵ Given that we are conditioning on 0.95 rather than 0.98 and given that the bias is increasing in the autoregressive parameter, the additional bias is slightly smaller than the bias in \hat{a} .

is based on the same idea as the traditional asymptotic intervals, except that we create our own “bootstrap t ”-table to replace the critical points based on the $\mathcal{N}(0, 1)$ table. We approximate the distribution of the asymptotically pivotal t -statistic

$$\frac{\hat{\theta} - \theta}{\hat{\sigma}(\hat{\theta})}$$

by

$$\frac{\hat{\theta}^* - \hat{\theta}}{\hat{\sigma}(\hat{\theta}^*)},$$

where the estimate $\hat{\theta}$ is treated as a fixed parameter in the bootstrap DGP. Let $\hat{t}_{\gamma/2}^*$ and $\hat{t}_{1-\gamma/2}^*$ be the critical points defined by the $\gamma/2$ and $1 - \gamma/2$ quantiles of the distribution of $\hat{t}^* \equiv (\hat{\theta}^* - \hat{\theta})/\hat{\sigma}(\hat{\theta}^*)$. Then

$$CI_{PER-t} = [\hat{\theta} - \hat{t}_{1-\gamma/2}^* \hat{\sigma}(\hat{\theta}), \hat{\theta} - \hat{t}_{\gamma/2}^* \hat{\sigma}(\hat{\theta})].$$

Superficially, this bootstrap interval looks similar to the usual asymptotic interval, but note that the bootstrap t -values allow for possible asymmetry in the distribution and implicitly correct for bias. This fact is important. It explains why the equal-tailed percentile- t interval tends to be much more accurate in practice without any explicit bias adjustments than the percentile interval without further adjustments.

How reliable this percentile- t interval is in finite samples also depends on the accuracy of $\hat{\sigma}(\hat{\theta})$. Here $\hat{\sigma}(\hat{\theta})$ is the estimated standard error of $\hat{\theta}$ and may be computed analytically or by bootstrap simulation. In the latter case, one requires an additional nested bootstrap loop not only for the estimation of $\hat{\sigma}(\hat{\theta})$ but also for the estimation of each $\hat{\sigma}(\hat{\theta}^*)$. In practice, it is common to rely on closed-form solutions for the asymptotic variance, whenever possible. Even if a few hundred bootstrap iterations are enough for estimating the second moment of the distribution, this increases the computational cost by a factor of a few hundred.

The reliability of this interval also depends on how close to pivotal the finite-sample distribution of the t -statistic is. A statistic is pivotal if its distribution is the same, regardless of the values of the population parameters in the DGP. In our context, we know that the t -statistic is asymptotically $\mathcal{N}(0, 1)$ distributed under stationarity and hence pivotal in the limit, but this approximation need not work well in small samples and, in any case, can be expected to work well only for processes that are not very persistent. Nevertheless, in the AR(1) model the finite-sample approximation provided by the equal-tailed percentile- t interval remains reasonably accurate even fairly close to the unit root. Additional simulation evidence for VAR models shows, however, that the small-sample accuracy of the equal-tailed percentile- t interval can be erratic

in some regions of the parameter space (see, e.g., Kilian 1999b; Benkowitz, Lütkepohl, and Neumann 2000).

Symmetric Percentile-*t* Intervals. The other type of percentile-*t* interval is the symmetric percentile-*t* interval proposed by Hall (1992). Let $\widehat{t}_{1-\gamma}^*$ be the critical point defined by the $1 - \gamma$ quantile of the distribution of $\widehat{t}^* \equiv |\widehat{\theta}^* - \widehat{\theta}| / \widehat{\sigma}(\widehat{\theta}^*)$. Then

$$CI_{PER-t}^{\text{symmetric}} = [\widehat{\theta} - \widehat{t}_{1-\gamma}^* \widehat{\sigma}(\widehat{\theta}), \widehat{\theta} + \widehat{t}_{1-\gamma}^* \widehat{\sigma}(\widehat{\theta})].$$

Note that this approach simply amounts to bootstrapping the two-sided *t*-test statistic as opposed to the one-sided *t*-statistic. The same caveats regarding implementation and computational cost apply that we already discussed for the equal-tailed percentile-*t* interval. In simulation studies, the symmetric percentile-*t* interval for autoregressions is often, but not always more accurate than the equal-tailed percentile-*t* interval (see Kilian 1999b).

Hall's Percentile Interval. Finally, Hall (1992) proposed bootstrapping yet another statistic obtained by replacing the denominator of the *t*-statistic underlying the equal-tailed bootstrap-*t* interval,

$$\frac{\widehat{\theta} - \theta}{\widehat{\sigma}(\widehat{\theta})},$$

by 1. Put differently, he proposed bootstrapping the statistic

$$\widehat{\theta} - \theta.$$

Let $\widehat{\theta}_{\gamma/2}^*$ and $\widehat{\theta}_{1-\gamma/2}^*$ be the critical points defined by the $\gamma/2$ and $1 - \gamma/2$ quantiles of the distribution of $\widehat{\theta}^*$. Then Hall's percentile interval is

$$\begin{aligned} CI_{PER}^{\text{Hall}} &= [\widehat{\theta} - (\widehat{\theta}_{1-\gamma/2}^* - \widehat{\theta}), \widehat{\theta} - (\widehat{\theta}_{\gamma/2}^* - \widehat{\theta})] \\ &= [2\widehat{\theta} - \widehat{\theta}_{1-\gamma/2}^*, 2\widehat{\theta} - \widehat{\theta}_{\gamma/2}^*]. \end{aligned}$$

Hall (1992) somewhat misleadingly refers to the resulting bootstrap interval as the percentile interval, and to the original percentile interval proposed by Efron many years earlier as the “other” percentile interval.

Unlike Efron's percentile interval, Hall's percentile interval implicitly builds in an automatic bias correction, in that it has by construction a long left tail if the underlying distribution has a long right tail. This feature is not enough to ensure accurate inference in autoregressions, however. Although Hall's percentile interval in some cases is quite accurate, simulation evidence suggests that its small-sample coverage accuracy may be quite poor in other cases, especially compared with percentile-*t* intervals. Nor is it systematically more accurate than Efron's percentile interval in its original form. For further discussion the reader is referred to the comparison in Kilian (1999b).

Another interesting feature of Hall's percentile interval is that it has the correct size asymptotically when the statistic of interest is consistent and asymptotically normal, as assumed here, but that it may also be asymptotically valid in non-Gaussian settings, in which the Efron percentile interval is known to be invalid.

12.3 Bootstrap Intervals Based on Bias-Adjusted Estimators

Kilian (1998a, 1998b, 1998c, 1999b) proposed a bootstrap method based on bias-adjusted slope parameters designed to address the concerns about the lack of small-sample accuracy of the recursive bootstrap method in Runkle (1987). Like Runkle, Kilian relies on the percentile interval.⁶

The proposal is to approximate the unknown stationary VAR(p) data generating process

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where $u_t \stackrel{iid}{\sim} F$, conditional on the observed data, not by the bootstrap DGP

$$y_t^* = \hat{v} + \hat{A}_1 y_{t-1}^* + \cdots + \hat{A}_p y_{t-p}^* + u_t^*$$

with $u_t^* \stackrel{iid}{\sim} \hat{F}_T$, but rather by the bootstrap DGP

$$y_t^* = \hat{v} + \hat{A}_1^{BC} y_{t-1}^* + \cdots + \hat{A}_p^{BC} y_{t-p}^* + u_t^*,$$

where $[\hat{A}_1, \dots, \hat{A}_p]$ has been replaced by first-order mean-bias-corrected LS estimates $[\hat{A}_1^{BC}, \dots, \hat{A}_p^{BC}]$ already discussed in Chapter 2 and where $u_t^* \stackrel{iid}{\sim} \hat{F}_T$. Let $\alpha = \text{vec}([A_1, \dots, A_p])$. Then, under some regularity conditions, $\mathbb{E}(\hat{\alpha} - \alpha) = b(\hat{\alpha})/T + O(T^{-3/2})$, so $\hat{\alpha}^{BC} = \hat{\alpha} - b(\hat{\alpha})/T$, where $b(\hat{\alpha})/T$ may be estimated based on the closed-form solutions discussed in Chapter 2 or by bootstrap methods. The motivation for this bias correction is that we want the bootstrap DGP to be centered on parameter values that are not systematically different from their population values.

Sometimes $\hat{\alpha}$ is already close to the unit circle, and adjusting for small-sample bias would push $\hat{\alpha}^{BC}$ outside the stationary region. Such an adjustment

⁶ This method of bias correcting the bootstrap is not related to the BC interval of Efron. In the context of bootstrapping structural impulse responses, the BC interval would adjust the endpoints of the interval for the conventional bootstrap impulse response estimator proposed by Runkle to account for the median bias in the impulse response estimator. This procedure does not involve bias-adjusting either the estimator of the impulse response or that of any of the autoregressive slope parameters. In contrast, Kilian adjusts for mean bias in the original as well as in the bootstrap slope parameter estimates before computing the bootstrap impulse responses, while evaluating the bootstrap distribution using Efron's original percentile interval.

would be inconsistent with the premise of a stationary DGP. To guard against this possibility, Kilian proposes a stationarity correction of the bias estimate (see Chapter 2). One first computes the modulus of the largest root of the companion matrix $\widehat{\mathbf{A}}$. Denote this quantity by $m(\widehat{\boldsymbol{\alpha}})$. If $m(\widehat{\boldsymbol{\alpha}}) \geq 1$, set $\widehat{\boldsymbol{\alpha}}^{BC} = \widehat{\boldsymbol{\alpha}}$. If $m(\widehat{\boldsymbol{\alpha}}) < 1$, however, compute $\mathbf{b} = b(\widehat{\boldsymbol{\alpha}})/T$ and $\widehat{\boldsymbol{\alpha}}^{BC} = \widehat{\boldsymbol{\alpha}} - \mathbf{b}$. If the implied $m(\widehat{\boldsymbol{\alpha}}^{BC}) \geq 1$, iterate on $\widehat{\boldsymbol{\alpha}}_i^{BC} = \widehat{\boldsymbol{\alpha}}^{BC} - \mathbf{b}_i$, $i = 1, 2, \dots$, until $m(\widehat{\boldsymbol{\alpha}}_i^{BC}) < 1$, where \mathbf{b}_i is shrunk by 1% for each i . This stationarity correction has no effect asymptotically.

The most convenient way of estimating the first-order mean bias in the VAR model is to rely on the closed-form solutions provided by Pope (1990). An alternative would be to estimate the bias by a nested bootstrap loop as in Kilian (1998c). If the regression model includes additional deterministic regressors such as a linear time trend, existing closed-form solutions for the LS bias in VAR models do not apply and we have to rely on the bootstrap. Especially when using closed-form solutions, it is possible to iterate the first-order bias estimates for improved accuracy. As discussed in Kilian (1998c), the magnitude of these higher-order bias terms is typically too small to matter in practice.

It may seem that we should also bias correct $\widehat{\sigma}$, because the structural impulse response estimator $\widehat{\theta}_{ik,h}$ depends on both $\widehat{\boldsymbol{\alpha}}$ and $\widehat{\sigma}$. Such a modification would only have second-order effects, however, and would require us to recompute and recenter the residuals as well to preserve the internal consistency of the bootstrap algorithm. Given these complications, the small order of the effects, and the favorable performance of the algorithm even without this correction, it makes sense to retain the original $\widehat{\sigma}$.

A subtle but important point to keep in mind is that bias corrections also affect the mean of the process because $\mathbb{E}(y_t) = (I_K - A_1 - \dots - A_p)^{-1}v$. This problem does not arise when $v = 0$. In practice, however, v may be far from zero, in which case the bootstrap replicates of the data will have a different mean after the bias correction, which affects the stationary distribution and causes transitory dynamics when generating bootstrap data in the conventional manner. These transitory dynamics undermine the accuracy of the bootstrap approximation, unless we discard the transients. The best way to avoid this problem in practice is to demean the data prior to estimating the model.

Having generated a sequence of data $\{y_t^{*r}\}_{t=-p+1}^T$, one fits the VAR(p) model with intercept to obtain the LS estimates $[\widehat{v}^{*r}, \widehat{A}_1^{*r}, \dots, \widehat{A}_p^{*r}]$ and $\widehat{\Sigma}_u^{*r}$, one implements the bias corrections on $\widehat{\boldsymbol{\alpha}}^*$ exactly as discussed earlier, and one constructs the implied bootstrap estimates $\widehat{\theta}_{ik,h}^{*r} = g(\widehat{\boldsymbol{\alpha}}^{BC*r}, \widehat{\sigma}^{*r})$ for $r = 1, \dots, R$. From the bootstrap approximation to the unknown empirical distribution of the structural impulse response estimates one may construct the $100(1 - \gamma)$ percentile confidence interval

$$[\widehat{\theta}_{\gamma/2}^*, \widehat{\theta}_{1-\gamma/2}^*],$$

where $\widehat{\theta}_{\gamma/2}^*$ and $\widehat{\theta}_{1-\gamma/2}^*$ refer to the lower and upper interval endpoints of the distribution of $\widehat{\theta}_{ik,h}^{*r}$.

The asymptotic validity of the bias-adjusted bootstrap follows from standard arguments in the literature by observing that the first-order bias is of order T and does not affect the limiting distribution of $\sqrt{T}(\widehat{\alpha} - \alpha)$ and $\sqrt{T}(\widehat{\sigma} - \sigma)$. Hence, none of the proposed adjustments affect the first-order asymptotic validity of the recursive bootstrap method (see Kilian 1998c).

Originally, Kilian (1998c) proposed a computational shortcut to the algorithm described above, which is referred to as the bootstrap-after-bootstrap method. The proposal was to use the first-order bootstrap bias estimate $b(\widehat{\alpha})/T$ as a proxy for the bias of the bootstrap estimates $b(\widehat{\alpha}^{*r})/T$ for $r = 1, \dots, R$. Kilian (1998c) proves that $b(\widehat{\alpha})/T$ and $b(\widehat{\alpha}^{*r})/T$ differ only by $O_p(T^{-3/2})$. This difference vanishes faster than the bias estimate itself and hence can be ignored in practice. The bootstrap-after-bootstrap speeds up the implementation of the bias-adjusted bootstrap method by a factor roughly corresponding to the number of bootstrap replications used in estimating the bias by bootstrap. This computational shortcut was essential when Kilian (1998c) was written, but had become largely obsolete by the time the article appeared in print, as computers had become much more powerful in the intervening years. Indeed, all subsequent papers on this topic, including Kilian (1998a, 1998b, 1999b), Kilian and Chang (2000), and Inoue and Kilian (2002b), use the unabridged algorithm described above.

An important caveat is that the closed-form bias estimates of Pope (1990) do not allow for conditional or unconditional heteroskedasticity. In fact, there are no analytic results about the mean bias in that case, and the order of the bias is unknown. It is possible, of course, to estimate the bias using the wild bootstrap, for example. The theoretical rationale for applying the bias-adjusted bootstrap approach in this case has yet to be established, however.

12.4 Potential Pitfalls in Impulse Response Inference

As already mentioned, one of the crucial assumptions in applying the Gaussian approximation is that the variance of the structural impulse response estimator is not zero (see Benkowitz, Lütkepohl, and Neumann 2000). This caveat applies not only to the delta method but also to all bootstrap methods. The nature of this problem is best understood in the context of a univariate stationary AR(1) process

$$y_t = ay_{t-1} + u_t,$$

where $|a| < 1$ and u_t is iid white noise. In this model, the response to a unit shock at horizon h is

$$\phi_h = a^h.$$

Under general conditions, the limiting distribution of the least-squares estimator \hat{a} can be expressed as

$$\sqrt{T}(\hat{a} - a) \xrightarrow{d} \mathcal{N}(0, 1 - a^2),$$

so, by the delta method,

$$\sqrt{T}(\hat{a}^h - a^h) \xrightarrow{d} \mathcal{N}(0, h^2 a^{2(h-1)}(1 - a^2))$$

provided $a \neq 0$. If $a = 0$, in contrast, such that all impulse responses are zero in population,

$$\sqrt{T}(\hat{a}^h - a^h) = \sqrt{T}\hat{a}^h \xrightarrow{d} 0$$

for $h > 1$. In other words, $(\hat{a}^h - a^h)$ converges to zero at a rate more rapid than $T^{1/2}$ and the distribution of $\sqrt{T}(\hat{a}^h - a^h)$ is degenerate.

Because the estimate of the limiting variance is nonzero almost everywhere in the parameter space, it may be tempting to use the quantity

$$\sqrt{T}(\hat{a}^2 - a^2) / 2\hat{a}(1 - \hat{a}^2)^{1/2}$$

for constructing a confidence interval for ϕ_2 , for example. However, for $a = 0$, the t -ratio reduces to

$$\sqrt{T}\hat{a}/2\sqrt{(1 - \hat{a}^2)},$$

where $\text{plim } \sqrt{(1 - \hat{a}^2)} = 1$ such that the t -ratio converges to $\mathcal{N}(0, 1/4)$ asymptotically, given that

$$\sqrt{T}\hat{a} \xrightarrow{d} \mathcal{N}(0, 1)$$

for $a = 0$. Hence, a confidence interval based on the asymptotic normal distribution would lack coverage accuracy.

As a result, standard asymptotic inference which ignores this possible singularity in the asymptotic variance-covariance matrix of the impulse response estimator will be misleading. This problem generalizes to structural impulse response analysis in vector autoregressions. Analogous problems arise for related statistics such as structural forecast error variance decompositions (see Benkowitz, Lütkepohl, and Neumann 2000).

The bootstrap approach does not help solve the problem of degenerate asymptotic distributions.⁷ In fact, in some cases bootstrap intervals may be even less accurate than the delta method in the presence of singular covariance structures. For example, the CI_{PER}^{Efron} -interval has zero coverage probability for

⁷ A common view is that changes in the rate of convergence at the point of discontinuity necessarily cause the bootstrap to fail. Inoue and Kilian (2003) prove by counterexample that the bootstrap may be valid even in the presence of such discontinuities. Nevertheless, in the current setting, the bootstrap approximation fails.

even h , when $a = 0$, because in that case the bootstrap impulse response estimator $(\hat{a}^*)^h$ will be strictly positive with probability 1. For odd h , this problem does not arise. To illustrate this point, consider the example of $h = 2$. Given the asymptotic standard normal distribution of $\sqrt{T}\hat{a}$, we know that

$$T\hat{a}^2 \xrightarrow{d} \chi^2(1),$$

which has positive support. Consequently, the lower confidence band of the CI_{PER}^{Efron} -interval for $(\hat{a}^*)^2$ cannot include $a^2 = 0$, resulting in a coverage rate of zero in repeated sampling. This result is unaffected by modifications of the original percentile interval such as small-sample bias adjustments.

Although other bootstrap confidence intervals such as CI_{PER}^{Hall} , CI_{PER-t} , or $CI_{PER-t}^{\text{symmetric}}$ will have higher coverage than the CI_{PER}^{Efron} -interval for even h , they also lack asymptotic justification. Moreover, some of the intervals that perform well for $a = 0$ in finite samples may perform poorly in asymptotically unproblematic parameter regions, making it difficult to offer advice to applied researchers.

Benkowitz, Lütkepohl, and Neumann (2000) explore a number of alternative approaches that remain asymptotically valid in this setting. For example, one possibility is to circumvent the problem by postulating a VAR(∞) process and allowing the lag order of the approximating VAR model to grow with the sample size. Another possibility is the use of subsampling methods. Benkowitz et al. conclude that either the small-sample accuracy of these procedures is not impressive or that it is not obvious how they can be extended to higher-dimensional processes. Yet another possibility would be to eliminate all points of singularities by fitting suitably restricted VAR models. That approach may not be practical, however, unless it is already known which impulse responses are zero. The use of pretests for identifying subset VAR structures also would raise concerns about the validity of asymptotic inference.

How practically important the example in Benkowitz, Lütkepohl, and Neumann (2000) is for structural VAR analysis depends on one's perspective. Benkowitz et al.'s argument is that there are situations where zero impulse responses are of particular interest in applied work. It may seem that zero impulse responses would arise easily in higher-dimensional stationary VAR models, but this is not the case, given that there are no exogenous observables in general equilibrium models of the economy and only exogenous variables would exhibit responses that are literally zero. There are some situations in which economic theory implies a zero response, however, as in Eichenbaum and Evans' (2005) analysis of uncovered interest parity. In the latter situation, Benkowitz et al.'s (2000) point that there may be problems of inference in VAR models due to singularities in the asymptotic distribution clearly applies.

However, there are many situations in which zero impulse responses are implausible. For example, no one would expect the response of real GDP to its

own innovation to be zero at all horizons, given the high persistence of the real GDP data. Thus, the problem identified by Benkwitz et al. does not necessarily invalidate standard methods of inference.

12.5 Finite-Sample Properties of Bootstrap Confidence Intervals

Abstracting from the problem of singular covariance structures, it can be shown that the symmetric percentile- t interval has theoretical properties similar to the infeasible BC_α and ABC intervals. Hall (1992) proves, albeit in a much simpler context, that among all two-sided confidence intervals apart from the BC_α and ABC intervals only the symmetric percentile- t interval achieves improved coverage accuracy asymptotically. All other two-sided confidence intervals are no more accurate asymptotically than the delta method interval. In particular, bootstrapping a studentized estimator alone is not sufficient for improved asymptotic coverage accuracy.

Hall's theoretical results, however, assume that an Edgeworth expansion of the distribution function of the estimator of interest exists, which has yet to be established in many applications. They also are based on the premise that the sample is large enough to invoke this asymptotic approximation. A simulation study in Kilian (1999b) suggests that even in a stylized bivariate VAR(1) model it may take sample sizes as large as $T = 2,000$ for the first-order asymptotics for $\hat{\theta}_{ik,h}$ to provide a good approximation. In contrast, actual sample sizes in empirical macroeconomics tend to be between 160 and 720, not to mention that actual VAR models tend to have more lags and variables. Hence, it is not clear how relevant Hall's results about asymptotic refinements are for practitioners.

A small-scale simulation study based on AR(1) and VAR(p) models in Kilian (1999b) provides no evidence that the symmetric percentile- t interval is systematically more accurate than the equal-tailed percentile- t interval or than the percentile interval based on the bias-adjusted bootstrap method of Kilian (1998b). In fact, there are cases when it is less accurate than either alternative in finite samples. It is also shown that Hall's percentile interval tends to perform poorly in many cases, as does Efron's percentile interval when applied without explicit bias adjustment, confirming the earlier simulation results in Kilian (1998c). Related results for the AR(1) model can also be found in Benkwitz, Lütkepohl, and Neumann (2000).

In related work, Kilian (1998c, 1999b) demonstrates that, when applying the standard residual-based bootstrap method for VAR models with iid errors, small-sample bias adjustments greatly improve the coverage accuracy of Efron's percentile interval for structural impulse responses. Among the bootstrap intervals considered so far, it tends to be the most robust and most reliable approach, as long as the conditions for applying the delta method are satisfied. In models with only an intercept, the effective coverage tends to be reasonably close to the nominal coverage even for borderline stationary processes with

roots close to unity. In models with an added deterministic linear time trend, the improvement in coverage accuracy compared with Runkle's (1987) approach remains striking, but the actual coverage accuracy may be significantly lower than the nominal coverage accuracy. None of the methods discussed so far can be considered satisfactory in the latter class of models. This is a small-sample problem, not a problem with their asymptotic validity.

Coverage accuracy is the most important criterion for selecting between confidence intervals, but there are also other differences to keep in mind. For example, percentile- t intervals are not transformation respecting. This means that we cannot infer the interval endpoints for a statistic $g(\hat{\theta})$, where $g(\cdot)$ denotes a nonlinear transformation, from the endpoints of the interval constructed for $\hat{\theta}$. Moreover, like the delta method confidence interval, the percentile- t interval is not range-respecting. For example, when estimating a proportion, the interval may include values outside of the interval [0, 1]. On the other hand, Efron's percentile interval, while range-respecting, by construction does not include values of 0 or 1, which are on the boundary of the parameter space of the proportions estimator. They are valid only for inference about interior solutions. This point also applies to the problem of bootstrapping structural forecast error variance decompositions, for example.

12.6 Inference for Integrated and Cointegrated VAR Processes

If the variables in the VAR model are known to be integrated and/or cointegrated, it is straightforward to adapt the methods of inference discussed so far. As before, we assume that the structural model is exactly identified.

12.6.1 VAR Models in Differences

In the simplest example of variables that are integrated of order 1, one simply expresses these variables in first differences. If all variables are in first differences, the lag order of the VAR model to be bootstrapped is reduced by 1, resulting in the VAR($p - 1$) model

$$\Delta y_t = v + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t.$$

In this case, inference would be based on the cumulative impulse responses of Δy_t , which correspond to the responses of the level of y_t . Lütkepohl (1990) provides closed-form solutions of the asymptotic variance-covariance matrices of the cumulative structural impulse responses that facilitate the application of the delta method.

Bootstrap inference in this model is almost identical to the methods outlined earlier, except that we bootstrap the cumulative sum of the impulse responses

of Δy_t . We approximate the DGP by

$$\Delta y_t^* = \widehat{\nu} + \widehat{\Gamma}_1 \Delta y_{t-1}^* + \cdots + \widehat{\Gamma}_{p-1} \Delta y_{t-p+1}^* + u_t^*$$

conditional on the data. Bias adjustments of the type discussed in Kilian (1998c, 1999b) tend to be of lesser importance after imposing the unit roots in estimation.

12.6.2 Vector Error Correction Models

As shown in Chapter 3, there are three different representations of VECMs used in empirical work.

VECMs with Known or Consistently Estimated Cointegrating Vectors. If the cointegrating vector is known, we can rewrite the VAR model as a VECM:

$$\Delta y_t = \alpha z_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t,$$

where $z_{t-1} = \beta' y_{t-1}$, $\Gamma_j = -[A_{j+1} + A_{j+2} + \cdots + A_p]$ for $j = 1, 2, \dots, p-1$, $\alpha \beta' = - (I_K - A_1 - A_2 - \cdots - A_p)$, and where we have assumed that there are no deterministic terms in the DGP for expository purposes. This VECM can be estimated by multivariate LS using a suitably specified regressor matrix. Because $z_t \sim I(0)$ and $\Delta y_{t-i} \sim I(0)$, $i = 1, \dots, p-1$, all regressors can be written as stationary regressors and the slope estimator is \sqrt{T} -consistent and jointly asymptotically normal.

Alternatively, if the cointegrating vector is unknown, we can estimate the VECM

$$\Delta y_t = \alpha \widehat{z}_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t,$$

where $\widehat{z}_{t-1} = \widehat{\beta}' y_{t-1}$ may be obtained prior to estimating the VECM by one of the estimators described in Chapter 3. Because the estimator $\widehat{\beta}$ is super-consistent in that it converges at rate T rather than \sqrt{T} , we can condition on \widehat{z}_{t-1} in estimating the VECM as though it were known. Whether β is known or estimated makes no difference for the asymptotic distribution of the implied structural impulse responses.

For asymptotic inference on the structural impulse responses in the VECM, we first need to express the VECM as a VAR model in levels:

$$y_t = A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + u_t,$$

where $A_1 = \alpha \beta' + I_K + \Gamma_1$, $A_i = \Gamma_i - \Gamma_{i-1}$, $i = 2, \dots, p-1$, and $A_p = \Gamma_{p-1}$. Let

$$A = [A_1, \dots, A_p] = [\Pi, \Gamma] W + J, \quad (12.6.1)$$

where

$$\boldsymbol{\Pi} = - (I_K - A_1 - \cdots - A_p),$$

$$J_{K \times Kp} = [I_K, 0_{K \times K(p-1)}],$$

and

$$W_{Kp \times Kp} = \begin{bmatrix} I_K & 0 & 0 & \dots & 0 & 0 \\ I_K & -I_K & 0 & \dots & 0 & 0 \\ 0 & I_K & -I_K & & 0 & 0 \\ \vdots & & \ddots & \ddots & & \vdots \\ \vdots & & & \ddots & \ddots & \vdots \\ 0 & 0 & \dots & \dots & I_K & -I_K \end{bmatrix}.$$

Lütkepohl (2005, section 7.2) shows that $\widehat{\boldsymbol{\Gamma}} = [\widehat{\boldsymbol{\Gamma}}_1, \dots, \widehat{\boldsymbol{\Gamma}}_{p-1}]$ and $\widehat{\boldsymbol{\Pi}}$ in the VECM have a joint asymptotic normal distribution, as has the least-squares estimator \widehat{A} of A :

$$\sqrt{T} \text{vec}(\widehat{A} - A) \xrightarrow{d} \mathcal{N}(0, \Sigma_\alpha^{co}).$$

The variance-covariance matrix Σ_α^{co} may be consistently estimated by

$$\widehat{\Sigma}_\alpha^{co} = (ZZ')^{-1} \otimes [(Y - \widehat{A}Z)(Y - \widehat{A}Z)'],$$

where $Y = [y_1, \dots, y_T]$ and $Z \equiv [Y_0, \dots, Y_{T-1}]$ with $Y_{t-1} \equiv [y'_{t-1}, \dots, y'_{t-p}]'$. Moreover, in the Gaussian model,

$$\sqrt{T} \text{vech}(\widehat{\Sigma}_u - \Sigma_u) \rightarrow \mathcal{N}(0, 2\mathbf{D}_K^+(\Sigma_u \otimes \Sigma_u)\mathbf{D}_K^{+'}).$$

Asymptotically, $\widehat{\Sigma}_u$ is independent of \widehat{A} . Given these estimates, inference about the structural impulse responses based on the delta method can be conducted analogously to the stationary Gaussian VAR model.

Bootstrap inference for known β involves approximating the VECM DGP by

$$\Delta y_t^* = \widehat{\alpha} \beta' y_{t-1}^* + \widehat{\boldsymbol{\Gamma}}_1 \Delta y_{t-1}^* + \cdots + \widehat{\boldsymbol{\Gamma}}_{p-1} \Delta y_{t-p+1}^* + u_t^*$$

conditional on the data. For each bootstrap replication of the VECM estimates, the bootstrap estimates $\widehat{\boldsymbol{\Gamma}}^*$ and $\widehat{\boldsymbol{\Pi}}^*$ are converted to bootstrap estimates \widehat{A}^* using (12.6.1), allowing us to build up the bootstrap approximation to the empirical distribution of the structural impulse responses exactly as in the stationary case. In the absence of an intercept, care must be taken to recenter the residuals.

If the VECM relies on a first-stage estimate of the cointegrating vector β , one option is to first cumulate the Δy_t^* realizations implied by the bootstrap

DGP

$$\Delta y_t^* = \widehat{\alpha} \widehat{\beta}' y_{t-1}^* + \widehat{\Gamma}_1 \Delta y_{t-1}^* + \cdots + \widehat{\Gamma}_{p-1} \Delta y_{t-p+1}^* + u_t^*$$

to obtain $\{y_t^*\}_2^T$ and then estimate $\widehat{\beta}^*$ from these bootstrap data before reestimating the VECM conditional on $\widehat{\beta}^*$. The other option is to condition on the initial $\widehat{\beta}$ throughout as though β were known.

VECMs with Known Cointegrating Rank. If we know only the number of cointegrating relationships in the data, the VECM may equivalently be written as

$$\Delta y_t = \Pi y_{t-1} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t.$$

As discussed in Chapter 3, this model may be estimated efficiently by Gaussian maximum likelihood estimation if we have a priori information about the number of cointegrating relationships (or equivalently about the rank of Π). Johansen (1995) proves that the maximum likelihood estimators $\widetilde{\Gamma}$, $\widetilde{\Pi}$, and $\widetilde{\Sigma}_u$ in this VECM have the same joint asymptotic normal distribution as the Gaussian ML estimator in the restricted VECM we discussed earlier.⁸

Bootstrap inference may be based on the DGP,

$$\Delta y_t^* = \widetilde{\Pi} y_{t-1}^* + \widetilde{\Gamma}_1 \Delta y_{t-1}^* + \cdots + \widetilde{\Gamma}_{p-1} \Delta y_{t-p+1}^* + u_t^*,$$

where the known rank of $\widetilde{\Pi}^*$ is used in each bootstrap replication to construct the estimates $\widetilde{\Gamma}^*$ and $\widetilde{\Pi}^*$, which are converted to \widetilde{A}^* in order to construct the bootstrap estimates of the structural impulse responses.

The VECM can also be used if the cointegrating rank is unknown. In that case, statistical tests are used to determine the cointegrating rank first, and the bootstrap is conditioned on the cointegrating rank obtained in this way (see Chapter 3). Unlike in the case of a known cointegrating rank, the asymptotic properties of this procedure are not known.

VECMs with Known Cointegrating Vector in Triangular Representation. For cointegrated models with only two variables and a known cointegrating vector $(1, -1)$, there is yet another approach, called the triangular VAR representation of the VECM. For expository purposes, consider a VECM for spot exchange rates (s_t) and economic fundamentals (f_t). The presumption is that in the long run, s_t and f_t move together such that $s_t - f_t \sim I(0)$. It can be shown that this

⁸ Equivalently, one could have written the VECM as $\Delta y_t = \Pi y_{t-p} + \Gamma_1 \Delta y_{t-1} + \cdots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t$, where the first term involves y_{t-p} rather than y_{t-1} . This does not affect the asymptotic results except for some slight changes in notation. For details, see Lütkepohl and Reimers (1992a).

VECM is algebraically equivalent to a VAR model of the form:

$$\begin{pmatrix} \Delta s_t \\ s_t - f_t \end{pmatrix} = C_1 \begin{pmatrix} \Delta s_{t-1} \\ s_{t-p} - f_{t-p} \end{pmatrix} + \cdots + C_p \begin{pmatrix} \Delta s_{t-p} \\ s_{t-p} - f_{t-p} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}$$

with some restrictions on the lagged coefficients that in practice tend to be ignored. The latter model directly incorporates the known cointegrating vector (see, e.g., Campbell and Shiller 1987; Kilian 1999a). Estimation is by unrestricted LS. This example illustrates that, in many cases, known cointegrating vectors can be embedded as ratio variables in regular VAR models. The advantage is that standard residual-based bootstrap or asymptotic methods of inference may be applied to this stationary model. The responses of s_t correspond to the cumulative response of Δs_t , and the response of f_t may be inferred from the difference of the responses of s_t and $s_t - f_t$. Thus, asymptotic inference about the response of f_t requires knowledge of the covariance between the responses. In practice, inference is usually based on bootstrap methods.

12.6.3 Integrated and/or Cointegrated VAR Models in Levels

Although imposing unit roots and cointegration restrictions improves the efficiency of impulse response estimators, in many cases consistent estimators with conventional asymptotic distributions may also be constructed from the unrestricted levels specification of the VAR model.

Gaussian Asymptotic Approximations. As already discussed in Chapters 2 and 3, any VAR model may be consistently estimated in levels, regardless of the possible presence of unit roots or cointegration, provided the model includes an intercept and enough lags such that the slope coefficients can be written as coefficients on zero mean stationary regressors (see Sims, Stock, and Watson 1990). There are two situations in which the use of this level specification may be preferred. One situation is that we are not sure about the presence of unit roots in the model variables. In this case, a suitably specified VAR model in levels with intercept encompasses both the integrated and possibly cointegrated VAR model and the stationary VAR model without trend. It does not allow for trending behavior in the stationary variables, however. Such trending behavior would call for the inclusion of an additional deterministic time trend in the regression in the event the model is stationary. The inclusion of a deterministic time trend does not affect the asymptotic results, but it magnifies the small-sample bias in the slope parameters of the VAR model and hence may seriously undermine the accuracy of the asymptotic approximation in small samples. Perhaps for this reason, in applied work, users of VAR models typically exclude the deterministic time trend even for trending variables such as real output. In the frequentist framework, this amounts to treating real output as a near-unit

root (and possibly near-cointegrated) process in which the intercept acts as a near-drift.

Another important situation is a VAR model in which there may or may not be cointegration between the unit root variables. Rather than relying on pretests for the cointegration rank with the pretests outcome being subject to error, or simply assuming the existence of a cointegration relationship, it makes sense in this case to estimate the VAR without these restrictions, even if we are sure that there are unit roots in some of the variables.

As discussed in Chapters 2 and 3, asymptotic inference in level $\text{VAR}(p)$ models is complicated by the fact that the presence of integrated variables may render the asymptotic variance of the statistic of interest singular. Although the individual slope parameters retain their marginal Gaussian asymptotic distribution after \sqrt{T} -scaling provided $p > 1$, their joint variance-covariance matrix is singular and inference about certain linear combinations of slope parameters is nonstandard (see Chapter 3). This fact also has implications for inference about smooth nonlinear functions of $\text{VAR}(p)$ slope parameters, $g(\alpha)$. Under certain conditions, it may render the asymptotic covariance matrix of the estimator singular, invalidating standard inference.

This problem is best illustrated in the simpler context of a univariate autoregression. Let the scalar time series y_t be generated by the autoregressive process

$$y_t = \nu + a_1 y_{t-1} + \cdots + a_p y_{t-p} + u_t,$$

where $u_t \stackrel{iid}{\sim} F$, and $\rho \equiv a_1 + \cdots + a_p = 1$. Define $a = (a_1, \dots, a_p)'$. Consider the first-order linear approximation to the response of y_{t+h} to a unit shock in u_t at horizon $h \geq p$, given by

$$\phi_h = g(a_1, \dots, a_p) \approx \frac{\partial g}{\partial a'} (\widehat{a} - a).$$

Then it is immediately apparent that when this linear approximation is proportionate to the parameter ρ , i.e.,

$$\frac{\partial g}{\partial a_1} a_1 + \cdots + \frac{\partial g}{\partial a_p} a_p \propto a_1 + \cdots + a_p,$$

which happens when all elements of the derivative vector $\partial g / \partial a'$ are identical in population, the distribution of the impulse response estimator $\widehat{\phi}_h$ becomes nonstandard with the estimator converging at rate T (like the estimator of ρ in a Dickey-Fuller regression) rather than \sqrt{T} . In this case, the limiting variance of $\sqrt{T}(\widehat{\phi}_h - \phi_h)$ is zero and standard asymptotic and bootstrap inference fails even when $p > 1$. Analogous results can be derived for vector-valued unit root processes with and without cointegration.

How much of a concern this situation is in practice is not clear. One response, therefore, is to simply ignore this possibility, when constructing

pointwise confidence intervals for smooth functions $g(a)$. An alternative response is to adapt the approach of Toda and Yamamoto (1995) and Dolado and Lütkepohl (1996), as discussed in Chapter 3. If y_t consists of $I(0)$ and $I(1)$ variables only, it suffices to add an extra lag to the VAR process fitted to the data to obtain a nonsingular covariance matrix for the parameters associated with the first p lags. In other words, if the DGP is a VAR(p) process and a lag-augmented VAR($p + 1$) model,

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + A_{p+1} y_{t-p-1} + u_t,$$

is fitted by LS, then the estimator of $[A_1, \dots, A_p]$ has a nonsingular variance-covariance matrix. Hence, inference on smooth functions of the slope parameters such as reduced-form impulse responses may be conducted by the delta method based on $[\hat{A}_1, \dots, \hat{A}_p]$. More generally, if $I(d)$ variables are present the singularity problem of the covariance matrix can be resolved by augmenting the VAR model by d extra lags for estimation. This approach not only suffices to rule out singularities in the pointwise asymptotic variance of $g(a)$; it is, in fact, required if we are interested in conducting joint rather than pointwise inference on statistics that are smooth functions of the form $g(a)$.

Of course, most statistics of interest in structural VAR analysis involve smooth functions of the form $g(\alpha, \sigma)$ rather than $g(\alpha)$. A case in point are structural impulse responses and structural forecast error variance decompositions. In this case, the concern regarding forbidden linear combinations typically does not arise. Recall that, under suitable conditions,

$$\sqrt{T} (g(\hat{\alpha}, \hat{\sigma}) - g(\alpha, \sigma)) \xrightarrow{d} \mathcal{N} \left(0, \frac{\partial g}{\partial \alpha'} \Sigma_{\hat{\alpha}} \frac{\partial g'}{\partial \alpha} + \frac{\partial g}{\partial \sigma'} \Sigma_{\hat{\sigma}} \frac{\partial g'}{\partial \sigma} \right).$$

Even if the leading term of the limiting variance,

$$\frac{\partial g}{\partial \alpha'} \Sigma_{\hat{\alpha}} \frac{\partial g'}{\partial \alpha},$$

were zero due to some singularity in $\Sigma_{\hat{\alpha}}$, the other term,

$$\frac{\partial g}{\partial \sigma'} \Sigma_{\hat{\sigma}} \frac{\partial g'}{\partial \sigma},$$

will always be nonsingular, ensuring that the limiting variance of the structural impulse response estimator is nonsingular, unless the value of the reduced-form impulse response is zero.

This result provides the basis for conducting asymptotically valid pointwise inference on structural impulse responses in VAR models using the delta method. Provided the structural forecast error variance components are bounded away from the limiting values of 0 or 1, this result also allows asymptotic inference on structural forecast error variance decompositions. Of course,

the limit of this decomposition as $h \rightarrow \infty$ is not well defined in the presence of a unit root because the Wald MA representation of y_t does not exist.

Only if we are interested in conducting joint inference for sets of structural impulse responses will the use of the lag augmentation approach be required. It should be noted that the lag augmentation approach involves a loss of efficiency in that we include a redundant lagged regressor in the VAR model. In practice, however, this effect is likely to be negligible, as the width of confidence intervals of structural impulse responses tends to be insensitive to small changes in the lag order, especially if a reasonably large number of lags is used in the first place.

Bootstrap Approximations. Results analogous to those in Sims, Stock, and Watson (1990) also apply to conventional residual-based bootstrap methods for VAR models in levels, as shown in Inoue and Kilian (2002a). For expository purposes, consider first the simpler univariate context. Suppose the DGP is a random walk,

$$y_t = ay_{t-1} + u_t,$$

where $a = 1$, and we do not impose the unit root in estimation. Then the standard bootstrap estimator of a based on the AR(1) model (with or without deterministic regressors) will not recover the correct asymptotic Dickey-Fuller (DF) distribution. Rather, it will converge to a random distribution. In other words, the standard bootstrap based on the unrestricted AR(1) regression model fails (see Basawa, Mallik, Cormick, Reeves, and Taylor 1991; Datta 1996).⁹

By augmenting this AR(1) model with additional lags, however, say by fitting an AR(2) model rather than an AR(1) model, we can overcome the inconsistency of the bootstrap, at least for statistics such as individual slope parameters. It is irrelevant for this argument that the DGP is an AR(1) model. What matters is the number of lags in the fitted model. This approach does not ensure high accuracy of bootstrap confidence intervals in small samples, but it ensures their first-order asymptotic validity.¹⁰

More generally, the results in Inoue and Kilian (2002a) apply to all higher-order AR(p) models, even when a deterministic time trend is included in the regression in addition to the intercept. The reason is that the bootstrap estimate \hat{a}^* , where a is the sum of the autoregressive coefficients, converges at a rate

⁹ Interestingly, the unrestricted bootstrap remains asymptotically valid when we fit an AR(1) with intercept to data from a random walk with drift, although there is a similar discontinuity in the asymptotic distribution at the unit circle (see Inoue and Kilian 2003).

¹⁰ One practical concern is that residual-based bootstrap methods for structural impulse responses under these conditions may have poor coverage accuracy without small-sample bias adjustments of the type discussed in Kilian (1998c, 1999b). Although first-order bootstrap mean-bias adjustments in simulation studies tend to improve the small-sample accuracy of bootstrap intervals in the presence of unit roots, their theoretical validity so far has been established only for stationary VAR models (see Kilian 1999b).

faster than \sqrt{T} . The fact that \widehat{a}^* converges to a random distribution is still true even in AR(p) models. This fact does not matter, however, because \widehat{a}^* does not enter the limiting distribution of the level slope parameter estimates after \sqrt{T} -scaling. This means that we can conduct valid bootstrap inference on individual slope parameters and linear combinations of individual slope parameters except for combinations that are proportionate to a (see Inoue and Kilian 2002a).

This reasoning also generalizes to vector autoregressive processes of order $p > 1$, whether the data are $I(0)$, $I(1)$, or cointegrated, provided an intercept is included in estimation. More generally, residual-based bootstrap inference remains first-order asymptotically valid for smooth nonlinear functions $g(\alpha, \sigma)$, as long as the limiting variance of the impulse response is not degenerate. As in the case of the delta method, singularities in the joint variance-covariance matrix of the slope parameters may be overcome by relying on the lag-augmented VAR model if necessary.

An important assumption underlying the derivation of the asymptotic validity of the bootstrap for structural impulse responses in the presence of unit roots is that the horizon of the impulse response remains fixed as the sample size increases. Without that assumption, which is also used by the delta method intervals of Lütkepohl (1990), bootstrap inference for structural impulse responses based on the VAR model in levels would be invalid (see Phillips 1998). In practice, this assumption is commonly interpreted as requiring the horizon to be small relative to the sample size. In this regard, Kilian and Chang (2000) document by simulation that the coverage accuracy of conventional asymptotic and bootstrap intervals for VAR impulse responses tends to deteriorate at longer horizons when the data are highly persistent. For typical large-dimensional VAR models estimated in levels, the usual pointwise confidence intervals often remain accurate only for a horizon of about one year. This finite-sample result is not surprising, because the data have roots close to unity and these intervals are based on asymptotic approximations derived under the assumption of a fixed horizon.

12.7 Inference in Local-to-Unity VAR Processes

Many macroeconomic time series are highly persistent. An obvious concern for applied work is that we do not know whether these time series are integrated or not. This is not a question that can be resolved empirically. Even the most powerful unit root tests lack the power to help us discriminate between these hypotheses for empirically relevant sample sizes.

So-called confirmatory analysis based on tests of the $I(0)$ null hypothesis does not remedy this problem and has been shown to be misleading (see, e.g., Caner and Kilian 2001; Müller 2005). Confirmatory analysis refers to the

practice of testing both the null hypothesis that a given process is $I(1)$ and that it is $I(0)$, and interpreting a failure to reject the $I(1)$ null and a simultaneous rejection of the $I(0)$ null as confirmation of the $I(1)$ hypothesis. This practice tends to be misleading, because tests of the $I(1)$ hypothesis lack power against persistent alternatives in small samples, typically resulting in a non-rejection, while a spurious rejection of the $I(0)$ hypothesis is made likely by the fact that the standard critical values of tests of the $I(0)$ hypothesis do not allow for persistent processes under the null hypothesis, causing these tests to suffer from severe size distortions in small samples, with rejection rates approaching 70% at the nominal 5% level in some realistic settings. Moreover, even if the data were known to be integrated of order 1, we would not know whether they are cointegrated, and if they were known to be cointegrated, we would need to know at least the cointegrating rank, if not also the cointegration vectors. Statistical tests for cointegration often lack power, and identifying the correct cointegrating vectors in higher-dimensional VAR systems remains a challenge in practice.

It may be tempting simply to impose unit roots and/or cointegration restrictions in estimation, but imposing these restrictions when they do not hold will render the VAR estimates inconsistent and inference invalid. In contrast, the failure to impose unit root and cointegration restrictions when they do hold will render the model estimates less efficient, but the estimates will remain consistent. Moreover, as we have discussed, impulse response inference in suitably specified VAR models estimated in levels remains asymptotically valid in most practically relevant situations even without imposing these restrictions, and lag augmentation may be used to overcome the remaining problems.

These asymptotic arguments suggest that in case of doubt we should err on the side of not imposing unit roots and cointegration on higher-order VAR models. In finite samples, however, that strategy is not without its own drawbacks because the high persistence of the levels data increases the small-sample bias of the VAR slope parameters and may undermine the accuracy of conventional asymptotic and bootstrap approximations even in higher-order VAR models.

12.7.1 Local-to-Unity Asymptotics

The questionable accuracy of conventional delta method and bootstrap intervals near the unit circle has motivated the development of an alternative asymptotic approximation for estimators based on VAR models in levels, referred to as local-to-unity asymptotics (see Chapter 3). This method is designed to be robust to the possible presence of unit roots in VAR(1) models as well as higher-order VAR models. By construction, it is intended for VAR models based on short-run identifying restrictions only, as departures from exact unit roots would immediately invalidate the use of long-run identifying restrictions.

Early applications of this modeling framework to the construction of impulse response confidence intervals for univariate autoregressions include Wright (2000) and Gospodinov (2004). Here we focus on the vector model. The standard procedure of generating confidence intervals for structural VAR impulse responses within the local-to-unity framework was proposed by Pesavento and Rossi (2006). Consider the K -dimensional reduced-form model

$$(I_K - CL)y_t = \Psi(L)u_t, \quad (12.7.1)$$

where $C = I_K + \Lambda/T$, $\Lambda = \text{diag}(\lambda_1, \lambda_2, \dots, \lambda_K)$, and $\Psi(L) = \sum_{i=0}^{\infty} \Psi_i L^i$ is an MA operator. This assumption means that the diagonal elements of C are in the neighborhood of 1, where that neighborhood depends on a constant that shrinks toward zero at rate T . That rate T is chosen to match the rate of convergence of the least-squares estimator of C . If λ_j were 0, y_{jt} would follow an exact unit root process. Allowing $\lambda_j < 0$ means that the root is near unity, but below unity (see Chapter 3).

Our interest is in inference on the structural impulse responses. Unlike in conventional asymptotics, we parameterize the horizon h of the response as an increasing function of the sample size. Specifically, let $h = [\delta T]$ for some fixed $\delta > 0$, where $[\cdot]$ denotes the integer part of the number in brackets. Then, given that

$$\lim_{h \rightarrow \infty} \left(1 + \frac{\delta \lambda_k}{h} \right)^h = e^{\delta \lambda_k},$$

we have

$$C^h \rightarrow e^{\delta \Lambda}$$

as $h \rightarrow \infty$, where $e^{\delta \Lambda}$ is a diagonal matrix with the vector $(e^{\delta \lambda_1}, e^{\delta \lambda_2}, \dots, e^{\delta \lambda_K})$ on the main diagonal. The reduced-form impulse response at horizon h is:

$$\frac{\partial y_{t+h}}{\partial u'_t} = \Phi_h = C^h (I_K + C^{-1} \Psi_1 + C^{-1} \Psi_2 + \dots) + o(1) \approx C^h \Psi(1), \quad (12.7.2)$$

and the corresponding structural response of variable j to shock k is:

$$\frac{\partial y_{j,t+h}}{\partial w_{k,t}} \approx e'_j C^h \Psi(1) B_0^{-1} e_k \rightarrow e'_j e^{\delta \Lambda} \Psi(1) B_0^{-1} e_k \quad (12.7.3)$$

as $T \rightarrow \infty$, where e_j and e_k are columns j and k of the matrix I_K , and B_0^{-1} is the structural impact multiplier matrix. Note that this limiting response depends on the largest roots in Λ as well as the cumulative short-run dynamics $\Psi(1)$. At long horizons, the uncertainty associated with $\Psi(1)$ and B_0^{-1} will be negligible compared with the uncertainty about Λ , so we can simply replace these terms

by consistent estimates $\widehat{\Psi}(1)$ and \widehat{B}_0^{-1} . Although Λ cannot be estimated consistently, the confidence limits for the localizing constants λ_j can be obtained by inverting a unit root test such as the augmented Dickey-Fuller test (see, e.g., Stock 1991).¹¹ Let $(\lambda_{L,j}, \lambda_{U,j})$ denote the lower and upper bounds of this confidence interval for λ_j , $j = 1, \dots, K$. Because the components of the limiting response are monotone functions of λ_j , the implied confidence interval for the structural impulse response in (12.7.3) is

$$\left(e^{\delta\lambda_{L,j}} e'_j \widehat{\Psi}(1) \widehat{B}_0^{-1} e_k, \quad e^{\delta\lambda_{U,j}} e'_j \widehat{\Psi}(1) \widehat{B}_0^{-1} e_k \right).$$

The difference from the pointwise impulse response confidence intervals discussed earlier is that coverage is not pointwise, but joint for all long horizons. Simulations suggest that at horizons shorter than about 10% of the sample size, the asymptotic approximation breaks down. One caveat regarding this interval is that the assumption of a diagonal C , while sometimes reasonable in macroeconomic applications, may not hold in all empirical applications. Relaxing this assumption would considerably complicate inference. The other obvious drawback of this procedure is that it provides accurate approximations only for responses at long horizons.

The procedure in Pesavento and Rossi (2006) is a two-step procedure in that we first construct a confidence interval for λ_j and then form the confidence interval for the structural impulse response based on (12.7.3). In related work, Gospodinov (2004) proposes an alternative, more computationally intensive approach that involves only one step. His proposal is to construct the interval based on a one-step inversion of the LR test statistic for the impulse response. Here we consider a generalization of this approach to structural impulse responses discussed in Gospodinov (2003). As in Chapter 3, let

$$y_t = Cy_{t-1} + \Gamma_1 \Delta y_{t-1} + \dots + \Gamma_{p-1} \Delta y_{t-p+1} + u_t,$$

where $C = I_K + \Lambda/T$, and define $\chi \equiv (\text{vec}(C)', \text{vec}(\Gamma_1)', \text{vec}(\Gamma_2)', \dots, \text{vech}(\Sigma_u)')'$. Consider the null hypothesis $\mathbb{H}_0 : \Theta_h(\chi) = \Theta_{0,h}$ and let

$$LR_T = T \log [\det(\widetilde{\Sigma}_u^{-1}) \det(\widetilde{\Sigma}_u^r)]$$

denote the likelihood ratio statistic, where $\widetilde{\Sigma}_u^r$ and $\widetilde{\Sigma}_u$ denote the variance-covariance matrices of the restricted and unrestricted residuals, respectively, in

¹¹ Phillips (2014) shows that the inversion of the usual ADF statistic leads to invalid inference when the DGP is stationary. This observation suggests that caution is called for when constructing the first-stage confidence interval for the local-to-unity parameter. Phillips (2014) discusses possible solutions based on the switching approach of Elliott, Müller, and Watson (2015) and based on the approaches of Hansen (1999) and Mikusheva (2007).

the approximating VAR model. Then, given $h = [\delta T]$, under \mathbb{H}_0

$$LR_T \xrightarrow{d} \text{tr} \left[\left(\int_0^1 \mathbf{J}_c(s) d\mathbf{W}(s)' \right)' \left(\int_0^1 \mathbf{J}_c(s) \mathbf{J}_c(s)' ds \right)^{-1} \left(\int_0^1 \mathbf{J}_c(s) d\mathbf{W}(s)' \right) \right],$$

where $\mathbf{J}_c(s)$ is an Ornstein-Uhlenbeck vector process and \mathbf{W} is a vector of standard Brownian motions. Joint confidence sets for horizon h are obtained by inverting the LR_T test on a grid of possible values of $\Theta_{0,h}$. Pointwise confidence intervals at horizon h are obtained using the projection method (see, e.g., Dufour and Taamouti 2005; Chaudhuri and Zivot 2011).

There is no simulation evidence on the accuracy of Gospodinov's approach for VAR models, but evidence for univariate models suggests consistently high coverage at all but the shortest horizons. Compared with the two-step procedure of Pesavento and Rossi (2006), the inversion of the LR test statistic is more computationally intensive, which is why this method does not appear to have been used for VAR models in practice.

12.7.2 Inference in Levels for Local-to-Unity VAR Models

It can be shown that the results of Sims, Stock, and Watson (1990) and Inoue and Kilian (2002a) on the asymptotic validity of inference in higher-order vector autoregressions with exact unit roots also extend to data generating processes with roots local-to-unity. As long as the delta method can be applied, inference about structural impulse responses remains standard. Likewise, lag augmentation may be used to render asymptotic inference invariant to roots local to unity (see Bauer and Maynard 2012).

One caveat is that these results are conditional on the horizon being fixed with respect to the sample size. In practice, one would expect conventional bootstrap or delta method intervals to be accurate only at reasonably short horizons relative to the sample size. In this sense, the conventional approach is complementary to the nonstandard asymptotic analysis in Pesavento and Rossi (2006) for longer horizons, and one might consider ways of combining these asymptotic thought experiments. We will return to this point later in this chapter.

A second and less obvious caveat is that the robustness results in Inoue and Kilian (2002a) do not extend to all VAR models based on long-run identifying restrictions. Consider the bivariate model of Blanchard and Quah (1989), for example. Their identifying assumptions require that real GDP be treated as $I(1)$ and hence expressed in differences, whereas the unemployment rate must be $I(0)$ and hence expressed in levels. In practice, an obvious concern is that the unemployment rate itself is quite persistent. Gospodinov (2010) proves that

if the unemployment rate is parameterized as a local-to-unity process rather than an $I(0)$ process, reflecting our uncertainty about its true integration status, the structural impulse responses can no longer be consistently estimated, even assuming that the horizon is fixed. Likewise, more conventional asymptotic and bootstrap confidence intervals for the structural impulse responses are invalid. This result suggests caution in interpreting estimates of Blanchard-Quah style models, unless the second variable is well inside the stationary region.

12.7.3 The Grid Bootstrap Method

So far we have considered the local-to-unity framework from an asymptotic point of view. We now turn to bootstrap methods for local-to-unity processes. While conventional bootstrap methods remain valid asymptotically for approximating the distribution of individual slope parameters in AR(p) models as long as $p > 1$, they are unable to capture the distribution of the sum of the autoregressive slope coefficients (see Inoue and Kilian 2002a). Indeed, that problem is formally equivalent to the problem of bootstrapping the slope parameter of the AR(1) model with a unit root. As mentioned earlier, Basawa, Mallik, Cormick, Reeves, and Taylor (1991) proved that the conventional bootstrap for autoregressions fails in that situation.¹² While this result also extends to the local-to-unity AR(1) model, such proofs do not rule out the possibility that nonstandard bootstrap methods remain valid in this case.

Hansen (1999) proposed such a nonstandard bootstrap approach, called the grid bootstrap, for univariate time series. His objective was to construct confidence intervals for the sum of the autoregressive roots in the AR(p) local-to-unity model. He showed that confidence intervals constructed with this method achieve asymptotically first-order correct coverage for both stationary and local-to-unity autoregressive models.

The sum of the autoregressive coefficients corresponds to the parameter ρ in the augmented Dickey-Fuller (ADF) representation of y_t ,

$$y_t = \rho y_{t-1} + \zeta_1 \Delta y_{t-1} + \cdots + \zeta_{p-1} \Delta y_{t-p+1} + u_t.$$

The grid bootstrap exploits the fact that we can estimate the remaining parameters of this ADF model conditional on the null hypothesis about ρ . The critical values for the t -statistic for ρ depend on these restricted parameter estimates and will change with the specification of the null hypothesis. Because there are many possible null hypotheses, Hansen considers a grid of null values for the statistic of interest and composes a confidence set for ρ from the ρ values that are not rejected.

¹² For further discussion of the causes of that bootstrap failure, see also Inoue and Kilian (2003).

In Section 12.4, we already alluded to the fact that pointwise asymptotic approximations do not guarantee confidence intervals that have correct coverage uniformly across the parameter space. Such problems often arise when the convergence rate of the estimators to their asymptotic distribution changes across the parameter space. This problem has also been studied extensively in the context of processes with roots near unity. Mikusheva (2007) proves that the Hansen (1999) grid bootstrap procedure produces uniformly asymptotically correct confidence intervals for $|\rho| \leq 1$. Simulation evidence in Hansen (1999) and Kilian (1999b) confirms that this grid bootstrap is quite accurate in finite samples for a wide range of degrees of persistence including the unit root limit. It remains highly accurate even for values of ρ close to zero.

The original grid bootstrap is not designed for structural impulse response inference in VAR models, however. A substantial generalization of the grid bootstrap is provided in Mikusheva (2012). Mikusheva observes that under the assumption that the impulse response horizon is fixed, one indeed obtains impulse response estimates that typically converge to a normal distribution, allowing the use of the standard delta method and of conventional bootstrap methods as shown in Inoue and Kilian (2002a). Under the assumption that the horizon increases with the sample size and that the roots are local to unity, however, the asymptotic distribution will be nonstandard, calling for alternative methods such as the method of Pesavento and Rossi (2006) that we discussed earlier. The incompatibility of these two approaches requires practitioners to choose one or the other. Mikusheva's objective is to provide an approach that works uniformly well for both short horizons and long horizons independently of the presence of a unit root. She accomplishes this objective by applying the grid bootstrap not to the sum of the autoregressive coefficients as in Hansen (1999), but to the LR test statistic of Gospodinov (2004). Her grid bootstrap approach delivers a uniform approximation of the distribution of the reduced-form impulse response statistic over the entire parameter space. For short horizons the approximating distribution is close to normal, while for long horizons it approaches the local-to-unity limit.

These theoretical results help explain the performance of currently used methods of impulse response inference in simulation studies. They are not without important limitations, however. First, like Gospodinov's (2004) analysis, Mikusheva's results apply to the reduced-form impulse responses Φ_h , rather than the structural VAR responses Θ_h , and hence are not immediately relevant for applied work using structural VAR models. While one should be able to relax this simplifying assumption following Gospodinov (2003), this has not yet been done.

Second, unlike the results in Pesavento and Rossi (2006), Mikusheva's theoretical analysis only covers VAR(p) models with at most one large root. It is not uncommon in applied work to deal with multiple large roots. For example, in a

VAR model of the global oil market, both the real price of oil and oil production are highly persistent time series, but there is no reason for these variables to be near-cointegrated. Extending Mikusheva's results to allow for additional large roots, while conceptually straightforward, in practice would require evaluating the LR test statistic on a high-dimensional grid, which would render the construction of the grid bootstrap intervals even more computationally demanding. This is a problem because we have to evaluate this grid repeatedly, as we bootstrap the model, compounding the computational cost of Gospodinov's (2003) method, which itself is already computationally infeasible in realistic VAR models. As a result, Mikusheva's method will remain computationally infeasible in applied work for some time to come. Indeed, she does not demonstrate the feasibility of her grid bootstrap method for models more complex than a univariate AR(2) model.

Finally, we have to ask how well this grid bootstrap works in finite samples. Simulation results for the AR(2) DGPs in Mikusheva (2012) show that the LR grid bootstrap appears to control the coverage accuracy at short and long horizons, as intended, but fails to control coverage accuracy at intermediate horizons. The simulation examples suggest coverage deficiencies of up to 10% for a nominal 95% confidence interval. These coverage deficiencies appear modest, but additional evidence for more realistic models would be useful, as would be a direct comparison to the Bonferroni-based method of Pesavento and Rossi (2006) discussed next.

12.7.4 A Hybrid Method

In closely related work, Pesavento and Rossi (2006) proposed another method of constructing confidence intervals for structural impulse responses in VAR models based on local-to-unity asymptotics. Building on insights in Wright (2000), this hybrid method combines elements of local-to-unity long-run asymptotics and of conventional Gaussian short-run asymptotics based on Bonferroni bounds. There are three steps involved. Under the same assumptions as before, Pesavento and Rossi first compute a $(1 - \gamma_1)100\%$ confidence interval $(\lambda_{L,j}, \lambda_{U,j})$ for λ_j , $j = 1, \dots, K$, by inverting augmented Dickey-Fuller (ADF) tests. They use these confidence intervals to compute $e^{\delta\Lambda_L}$ and $e^{\delta\Lambda_U}$, where $\delta > 0$ is implied by $h = [\delta T]$. Second, they estimate the VAR model (12.7.1) in quasi-differences. This may be accomplished either by approximating Ψ_i at horizons $i = 1, \dots, H$, based on the cumulative impulse responses from a VAR in first differences or from the residuals of a VAR(1) in levels. Pesavento and Rossi then construct $(1 - \gamma_2)100\%$ confidence intervals for the elements of Ψ_i at horizons $i = 1, \dots, H$, using Lütkepohl's (1990) delta method. Third, for the lower and upper bound of the confidence interval for the elements of Ψ_i , Pesavento and Rossi compute $e^{\delta\lambda_{L,j}}\psi_{jk,i}$ and $e^{\delta\lambda_{U,j}}\psi_{jk,i}$,

where $\psi_{jk,i}$ denotes the jk^{th} element of Ψ_i . Then the confidence intervals for the elements of Ψ_i are given by

$$\left(\min_i e^{\delta\lambda_{L,j}} \psi_{jk,i}, \max_i e^{\delta\lambda_{U,j}} \psi_{jk,i} \right).$$

By the Bonferroni inequality, the coverage of the pointwise intervals for the elements of Ψ_i should be at least $(1 - \gamma_1 - \gamma_2)100\%$ at each horizon $i = 1, \dots, H$.

A simulation study in Pesavento and Rossi (2006) suggests that this hybrid method is conservative in that its coverage probability tends to exceed the nominal coverage rate at intermediate horizons. These intervals tend to be substantially more accurate at short horizons than the intervals implied by Pesavento and Rossi's baseline method discussed in Section 12.7.1.

Pesavento and Rossi's hybrid method was explicitly intended for inference on responses of persistent variables. Extensions to VAR models with multiple large roots are straightforward, given the assumption that Λ is diagonal. An open question is how well the proposed approximation works when conducting inference about responses of variables that are not very persistent. If the roots are known to be small, it may make sense to rely on conventional asymptotics for these responses instead. Alternatively, it may be helpful to express these model variables in log-levels rather than differences. Another open issue is how to deal with near-cointegration in the VAR model. In their appendix, Pesavento and Rossi (2006) consider the case of a known cointegration vector but do not discuss how to deal with a setting in which two (or more) model variables may or may not share a common root.

12.7.5 Implications for Second-Stage Inference after Pretesting

It is common among applied users of VAR models to rely on pretests for unit roots and cointegration to determine the appropriate transformation of the time series to be included in the VAR model. Inference is then conducted conditional on the model specification chosen. This approach is questionable. It has been shown that, when applying this procedure to data from a local-to-unity model, even arbitrarily small deviations from exact unit roots may produce impulse response estimates that are severely distorted and confidence intervals that are highly inaccurate. In contrast, the levels specification of the VAR model is robust to these pretest biases.

The Elliott Critique of Pretesting for Unit Roots and Cointegration. Exact unit roots are not an implication of economic theory (see, e.g., Sims 1988). The common perception among economists that there are exact unit roots in

many economic time series can be attributed in no small measure to the inability of unit root tests to reject the unit root null hypothesis (see Nelson and Plosser 1982). In fact, it is common in applied work to impose unit roots even on variables such as interest rates that in conventional economic theory do not exhibit unit root behavior. The problem is that a failure to reject the null hypothesis does not necessarily imply that this null hypothesis is true because unit root tests lack power against local alternatives. Imposing unit roots on the specification of the VAR model based on the results of pretests therefore may do more harm than good.

This problem was first discussed in Elliott (1998), who investigated the effect of pretest for unit roots and rank tests for cointegration in a local-to-unity VAR framework. As before, this framework serves as a device to model situations in which we are unable to determine from the data with any degree of reliability the presence of unit roots or cointegration. Elliott's concern is that in vector autoregressive analysis we are not interested in the presence of unit roots or cointegration for their own sake. Their presence matters only because it affects the specification of the VAR model and hence the distribution of estimators of smooth functions of the parameters of the VAR model. Elliott showed that commonly applied hypothesis tests for the latter objects will typically no longer have their usual χ^2 asymptotic distribution after pretesting. He also showed that confidence intervals for these statistics will lack coverage accuracy after conditioning on the results of pretests.

Implications for Structural Impulse Responses: Simulation Evidence. Elliott's warnings have been ignored by many applied researchers, perhaps because of the perception that estimates of structural impulse responses based on the levels specification are not without their own drawbacks. A systematic simulation analysis of this question is presented in Gospodinov, Herrera, and Pesavento (2013) who focus on the robustness of structural impulse response estimates to pretesting. Their evidence is based on a range of vector autoregressive DGPs with roots of 0.92, 0.95, 0.98, and 1. They show that structural impulse response estimators based on the levels specification have systematically (and often substantially) lower mean-squared error than estimators based on the pretest model, except when the dominant autoregressive root in the DGP is exactly unity. In the latter case, the results are mixed.

As to the coverage accuracy of confidence bands for structural impulse responses, Gospodinov et al. distinguish between models based on short-run identifying restrictions and long-run identifying restrictions. In both cases, intervals based on the bias-adjusted bootstrap method of Kilian (1999b) applied to the levels specification are substantially more accurate than for the pretest model. Related simulation results are also presented in Pesavento and Rossi (2006), who show that pretests undermine the coverage accuracy of

conventional delta method intervals for structural impulse responses, especially for processes with roots close to unity.

Implications for Structural Impulse Responses: Theory. Gospodinov, Maynard, and Pesavento (2011) illustrate why pretesting is particularly problematic in VAR models with near unit roots when using long-run identifying restrictions. The problem is best illustrated in the context of the Blanchard and Quah (1989) model. Recall that $z_t = (\Delta gdp_t, ur_t)'$, where by assumption $z_t \sim I(0)$, but $gdp_t \sim I(1)$. Analogously to model (12.7.1) we may express the bivariate reduced-form model as

$$(I_2 - CL)z_t = \Psi(L)u_t.$$

Building on results in Pesavento and Rossi (2006), Gospodinov et al. show that in this model the matrix C may be parameterized as

$$C = \begin{bmatrix} 1 & \delta \\ 0 & \rho \end{bmatrix},$$

where $\rho = 1 + \lambda/T$, $\lambda < 0$, denotes the dominant root of the unemployment process, and $\delta = -\gamma(1 - \rho)$ is a parameter that determines the low-frequency comovement between the unemployment rate and real GDP growth. Equivalently, we may write $A(L)z_t = u_t$, where

$$A(L) = \Psi(L)^{-1} \begin{bmatrix} 1 & \gamma(1 - \rho)L \\ 0 & 1 - \rho L \end{bmatrix}.$$

Recall that the structural representation of the VAR model is $B(L)z_t = w_t$ such that $w_t = B_0 u_t$. If Σ_w is diagonal and

$$B_0 = \begin{bmatrix} 1 & -b_{12} \\ -b_{21} & 1 \end{bmatrix},$$

then the long-run multiplier matrix $B(1) = B_0 A(1)$ can be expressed as

$$B(1) = \begin{bmatrix} \psi_{11}(1) - b_{12}\psi_{21}(1) & (1 - \rho)([\gamma\psi_{11}(1) + \psi_{12}(1)]) \\ \psi_{21}(1) - b_{21}\psi_{11}(1) & (1 - \rho)([\gamma\psi_{21}(1) + \psi_{22}(1)]) \end{bmatrix},$$

where $\psi_{ij}(1)$ denotes the ij^{th} element of $\Psi(1)^{-1}$. As discussed in Chapter 10, imposing Blanchard and Quah's long-run identifying restriction renders $B(1)$ lower triangular, which implies that for $\rho < 0$

$$b_{12} = [\gamma\psi_{11}(1) + \psi_{12}(1)] / [\gamma\psi_{21}(1) + \psi_{22}(1)].$$

Note the dependence of this solution on the parameter γ , which in turn determines the low-frequency comovement between the unemployment rate and real GDP growth via the parameter δ .

In contrast, if we changed the Blanchard and Quah model by differencing both real GDP and the unemployment rate, the corresponding VAR long-run multiplier matrix would be

$$B(1) = \begin{bmatrix} \psi_{11}(1) - b_{12}\psi_{21}(1) & \psi_{12}(1) - b_{12}\psi_{22}(1) \\ \psi_{21}(1) - b_{21}\psi_{11}(1) & \psi_{22}(1) - b_{21}\psi_{12}(1) \end{bmatrix}$$

with

$$b_{12} = \psi_{12}(1)/\psi_{22}(1)$$

after imposing that the upper right element of $B(1)$ is zero. Clearly, these two solutions for b_{12} differ, and so do the implied structural impulse responses. Moreover, these differences can be large even for $(1 - \rho)$ arbitrarily close to zero because the value of γ does not depend on ρ .

We conclude that the model with both variables in differences by construction ignores the potential low-frequency comovement between unemployment and real GDP growth. This is not a problem if indeed $\rho = 1$, but even for small departures from $\rho = 1$, such comovement in the data could cause the impulse response estimates to become severely biased. It may seem that it would be safer to express the unemployment rate in levels instead, but, while the latter specification preserves the information about the low-frequency comovement, the implied impulse response estimates obtained by conventional methods are known to be inconsistent if the unemployment rate follows a local-to-unity process (see Gospodinov 2010). They would have to be estimated by an alternative estimator proposed by Gospodinov (2010).

The analysis in Gospodinov, Maynard, and Pesavento (2011) also has important implications for the use of unit root pretests. Given that a unit root pretest for the unemployment rate will fail to reject $\rho = 1$ with high probability, even when the process is only near integrated, the use of pretests effectively amounts to ignoring potential low-frequency comovement, resulting in impulse response estimates that may be severely biased.

The root cause of all these problems, of course, is the fact that the unemployment rate is a fairly persistent process. If we replaced the unemployment rate in Blanchard and Quah's model by a variable with low persistence such as the U.S. economy's capacity utilization rate, for example, standard asymptotic approximations would be adequate.

The local-to-unity approach is but one way of allowing for departures from integer orders of integration. An alternative is the use of fractionally integrated processes, as discussed in Chapters 3 and 11. Tschernig, Weber, and Weigand (2013) also provide evidence that assumptions about variables being either $I(0)$ or $I(1)$ in Blanchard-Quah style models may distort structural impulse responses and inference. They propose to conduct inference about the structural impulse responses in a fractionally integrated VAR model identified

based on long-run exclusion restrictions using a bootstrap approach. Nothing is known about the theoretical validity of this bootstrap approach or its finite-sample accuracy at this point.

12.8 Local Projections

Jordà (2005) proposes an alternative method for estimating structural impulse responses based on so-called local projections. His point of departure is a representation of the standard linear VAR model that we already utilized in Section 4.1, where we used the companion form of a VAR(p) model to derive the representation

$$Y_{t+i} = \mathbf{A}^{i+1} Y_{t-1} + \sum_{j=0}^i \mathbf{A}^j U_{t+i-j},$$

where

$$Y_t \equiv \begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix}, \quad \mathbf{A} \equiv \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_K & 0 & & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{bmatrix}, \quad \text{and}$$

$$U_t \equiv \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix},$$

and, for simplicity, the intercept has been suppressed. Left-multiplying the equation by $J = [I_K, 0_{K \times K(p-1)}]$ yields

$$y_{t+i} = A^{(i+1)} Y_{t-1} + \sum_{j=0}^i \Phi_j u_{t+i-j} = A^{(i+1)} Y_{t-1} + v_{t+i}, \quad (12.8.1)$$

where $i = 0, 1, \dots, H - 1$, $A^{(j)} = J\mathbf{A}^j$ is a $K \times Kp$ matrix, $\Phi_0 = I_K$ by definition, $\Phi_j = J\mathbf{A}^j J'$ for $j = 1, \dots, H$, is the j^{th} $K \times K$ coefficient matrix of the Wold MA representation of y_t , and $v_{t+i} = \sum_{j=0}^i \Phi_j u_{t+i-j}$. Note that the first $K \times K$ block of $A^{(j)}$ is just the j^{th} reduced-form impulse response matrix Φ_j from which the structural impulse responses can be computed as $\Theta_j = \Phi_j B_0^{-1}$.

Jordà observes that the errors in the multivariate linear regression model (12.8.1) are uncorrelated with the regressors. Hence, each equation of the system may be estimated individually by LS. Since Φ_{i+1} is part of the parameter matrix $A^{(i+1)}$, a consistent estimator of Φ_{i+1} may be constructed by LS and a

heteroskedasticity and autocorrelation consistent (HAC) estimator of the error covariance matrix may be employed for inference on Φ_{i+1} .

Because the correlation structure of the error term v_{t+i} is known under the maintained assumption of the DGP being a VAR(p) model, GLS estimation may be used to improve the asymptotic efficiency of the estimator. Thus, an alternative approach is to jointly estimate the system of equations

$$\begin{pmatrix} y_t \\ \vdots \\ y_{t+H-1} \end{pmatrix} = \begin{bmatrix} A^{(1)} \\ \vdots \\ A^{(H)} \end{bmatrix} Y_{t-1} + \begin{pmatrix} v_t \\ \vdots \\ v_{t+H-1} \end{pmatrix}, \quad (12.8.2)$$

which contains all impulse response matrices up to horizon H , Φ_1, \dots, Φ_H . Because the autocorrelation structure of the error term of model (12.8.2) can be worked out easily if the true DGP is a VAR(p) process, that structure may be taken into account in GLS estimation.

Given estimates of the Φ_j matrices, the construction of the structural impulse responses is straightforward, provided the structural impact multiplier B_0^{-1} is known. The problem is that the local projection (LP) procedure does not yield an estimate of B_0^{-1} . In this sense, the linear LP estimation procedure is less general than standard linear VAR model estimators, unless the structural impulse responses coincide with the reduced-from impulse responses, as would be the case when considering responses to exogenous variables. In his empirical applications, Jordà deals with this problem by estimating B_0^{-1} from a conventional linear VAR model and substituting this estimate for B_0^{-1} when constructing the LP estimator of the structural impulse responses (see Jordà 2009).

Jordà (2005) suggests that the LP estimator of structural impulse responses has three potential advantages relative to the conventional VAR-based estimator:

1. Impulse responses may be estimated directly by linear LS regressions, simplifying the analysis.
2. Pointwise and joint inference for the impulse responses is straightforward and does not require appealing to the delta method.
3. Local projection estimates of the structural impulse responses are more robust to model misspecification than conventional VAR estimates.

Upon closer examination, none of these alleged advantages hold up to scrutiny. First, using linear LS estimation is indeed simple, but it is also inefficient if the DGP corresponds to the assumed VAR(p). To improve the efficiency of the estimator, one could account for the dependence structure in the error term and use feasible GLS instead of LS. Even feasible GLS

estimation, however, cannot compensate for the much larger number of parameters in the model. Note that there are HpK^2 parameters in the model (12.8.2), many of which are not of direct interest. Only the H impulse response matrices Φ_1, \dots, Φ_H , which involve only HK^2 of the HpK^2 estimated parameters, are of direct interest for impulse response analysis. When long horizons are considered, the number HpK^2 can be quite large. In contrast, a VAR(p) model has pK^2 slope parameters only, and the number of parameters to be estimated does not increase with the impulse response horizon. Thus, ultimately, the relative accuracy of the LP estimator and the VAR estimator is an empirical question. Kilian and Kim (2011) show by Monte Carlo simulation that, when the DGP is a linear VAR model, the LP estimator of the structural impulse responses tends to have higher bias as well as higher variance in small samples than the corresponding conventional VAR estimator. Thus, the simplicity of the LP estimator comes at a considerable cost.

Second, it is correct that LP regressions facilitate pointwise inference on the reduced-form impulse responses in that the LS estimates will be asymptotically normal under mild conditions (see Jordà 2009). This result is of no help for conducting inference about structural impulse responses, however. Valid inference in the latter case relies on the same assumptions as required for the conventional VAR-based estimators because of the dependence of the LP impulse response estimator on the VAR estimate of the structural impact multiplier matrix, and it requires appealing to the delta method for inference on structural impulse responses (see also Jordà 2009). Thus, the LP estimator does not relax any of the assumptions of the VAR-based estimator. Nor is the practical performance of the LP estimator of the structural impulse responses encouraging. Simulation evidence in Kilian and Kim (2011) shows that the coverage accuracy of pointwise LP confidence intervals for structural impulse responses tends to be lower than the accuracy of the corresponding VAR confidence intervals, regardless of whether we appeal to first-order asymptotics for the LP estimator or construct LP bootstrap confidence intervals.

Jordà also observes that the joint asymptotic distribution of the reduced-form impulse responses is easily obtained from estimating the LP model (12.8.2). This idea is developed in full in Jordà (2009). The problem, of course, is not deriving the joint asymptotic distribution. In fact, that distribution is also easily derived for the conventional VAR-based estimator under exactly the same assumptions, as we saw earlier in this chapter. The real challenge instead is how to translate this joint asymptotic distribution into valid confidence bands around the structural impulse responses. Jordà (2009) proposes constructing joint confidence bands for a given impulse response function by Scheffé's S-method of simultaneous inference, building on the asymptotic normality of the joint distribution of the structural impulse response coefficients in stationary VAR models (see Scheffé 1953).

It can be shown, however, that the inequalities underlying the construction of this joint confidence band are only approximate and that this confidence band may not have correct coverage probabilities even under ideal conditions, as was pointed out in Wolf and Wunderli (2015) in the closely related context of constructing joint forecast bands. The same reasoning also applies to impulse responses. Thus, it is perhaps not surprising that the simulation evidence in Kilian and Kim (2011) and Lütkepohl, Staszewska-Bystrova, and Winker (2015a) indicates that the joint LP confidence bands proposed by Jordà (2009) and Jordà and Marcellino (2010) tend to have inadequate coverage accuracy.

Third, it is not the case in general that LP estimators are more robust to misspecification than conventional VAR-based estimators. Recall that in linear VAR models reduced-form impulse responses are nonlinear transformations of the VAR slope parameters. Local projections may be viewed as linear approximations to these nonlinear functions. Given that the LP estimator can be constructed from single-equation LS estimation, it might seem at first sight that the LP estimator might be more robust to possible misspecifications of the reduced-form VAR model. This is not necessarily the case. The population impulse response functions by construction depend on the full system of equations because they reflect the interdependencies between all the variables contained in a given model. If that VAR model is misspecified because there are omitted variables, for example, then the impulse responses of the misspecified model do not properly reflect the actual responses of the system. Because the LP estimator is constructed as an approximation to these misspecified impulse responses, it suffers from the same omitted variable problem.

Nor is it necessarily the case that LP estimators are better suited for approximating VARMA models, because VAR models may be viewed as autoregressive sieves, as discussed in Chapter 2. Indeed, the simulation results in Kilian and Kim (2011) in favor of the conventional VAR-based estimator extend to both VAR and VARMA DGPs.

Finally, it is theoretically possible that the LP estimator of the reduced-form response might be more accurate than a linear VAR model when the DGP is nonlinear, as conjectured by Jordà (2005), but a number of related studies including Marcellino, Stock, and Watson (2006) and Pesaran, Pick, and Timmermann (2011) have provided evidence that the predictive accuracy of linear VAR models for macroeconomic data is higher than that of direct forecasts, provided the VAR model includes enough lags, casting doubt on this conjecture. In any case, if the linearity assumption underlying the VAR model were seriously violated, the LP estimator of the structural impulse response would be invalid by construction because it relies on the linear VAR based estimator of B_0^{-1} . Hence, there is no theoretical basis for Jordà's conjecture that the LP approach would be more robust in this case.

We conclude that there is no evidence to suggest that local linear projections should replace conventional linear VAR models. Given the practical and conceptual limitations of local projections, we do not consider this approach in the remainder of this book.

Although the focus in this chapter is on linear models, it is worth noting that the idea of local projections may be extended to nonlinear models (see Chapter 18). Jordà (2005) discusses a stylized example of a nonlinear local projection. One complication in this case is that accurate nonlinear approximations may require a large number of terms, which tends to undermine estimation efficiency. Another complication is that computing impulse responses from nonlinear local projections requires Monte Carlo integration, because the responses are dependent on the history of the data and the magnitude of the structural shock (see Chapter 18). A third complication is that it is not clear how to obtain an external estimate of the structural impact multiplier matrix in that case.

12.9 Synthesis

We conclude that there is no method of pointwise inference on structural impulse responses (and related objects such as structural forecast error variance decompositions) that works well in all circumstances. It matters how the VAR model is specified and how confidence intervals are constructed. As to the first point, unless we are sure of the nature of the integration and cointegration constraints, a good case can be made for working with a VAR(p) model in levels, provided the model includes an intercept and $p > 1$. Based on recent simulation studies, working with the levels specification tends to produce more accurate results than relying on pretests. In some cases, it will be necessary to use the lag-augmentation approach to ensure asymptotically valid inference. There is at present no simulation evidence on the finite-sample accuracy of structural impulse response inference based on the lag-augmented levels models, but in models with reasonably high lag orders, the inclusion of an extra lag is unlikely to make much of a difference.

As to the second point, bootstrap confidence intervals may improve on the accuracy of the delta method interval for structural impulse responses, but much depends on how these intervals are constructed. Unless the degree of persistence in the process is low, Efron's percentile interval cannot be recommended. For persistent data, Efron's percentile interval based on the bias-adjusted bootstrap method of Kilian (1998c, 1999b) stands out as providing reasonably accurate confidence intervals in most circumstances, at least for VAR models without deterministic time trends. An alternative is the use of the symmetric percentile- t bootstrap interval. The latter is typically preferred over

Hall's percentile interval, provided the statistic of interest is asymptotically normal. It also tends to be more accurate than the equal-tailed percentile-*t* interval in the VAR setting, to the extent that these intervals have been compared by simulation. None of these intervals can be expected to remain accurate at longer horizons, however.

One alternative is the hybrid method of Pesavento and Rossi (2006), which was derived under the assumption of one or more autoregressive roots local to unity. Simulation evidence for this method in the VAR context, however, is limited, and the applicability of this method in the near-cointegrated case remains unclear. Alternative bootstrap methods that are uniformly valid in the VAR parameter space such as the method of Mikusheva (2012) are likely to remain computationally infeasible for some time to come. Local projection estimators, in contrast, although increasingly popular in empirical macroeconomics, are not a practical alternative to VAR-based estimators of structural impulse responses even under stationarity.

12.10 Bayesian Regions of Highest Posterior Density

All methods discussed so far have been motivated from a frequentist point of view. There is a parallel literature on inference about structural impulse responses from a Bayesian point of view (see, e.g., Uhlig 1994, 2005; Sims and Zha 1999). We already reviewed Bayesian inference in some detail in Chapter 5. It is straightforward to extend the algorithms provided there to the problem of constructing the posterior distribution of structural impulse responses and other smooth functions of VAR model parameters. Our discussion, for the time being, assumes that the structural VAR model is exactly identified based on short-run or long-run restrictions. Inference in overidentified models is discussed in Section 12.13.

The Bayesian approach to inference is not only of interest to Bayesian users of VAR models, but Bayesian error bands have also been shown to work reasonably well when evaluated by their coverage properties, although not as well as the bias-adjusted bootstrap (see Kilian 1998c). In fact, under conventional priors, they produce pointwise intervals that asymptotically coincide with frequentist confidence intervals for structural impulse responses in stationary VAR models. This numerical equivalence breaks down, however, in the presence of unit roots in the data.

Our focus in this section is not on providing a comprehensive review of the Bayesian approach to evaluating structural VAR models, but on illustrating – in the context of a widely used Bayesian algorithm – how the Bayesian approach deals with issues such as trends, unit roots, and cointegration that affect frequentist inference about structural impulse responses.

12.10.1 Pointwise Inference on Structural Impulse Responses

It is useful to start with the Bayesian approach most commonly used in applied work. Following Uhlig (2005), consider the VAR(p) model

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where $u_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_u)$. Let $Y \equiv [y_1, \dots, y_T]$, $A \equiv [v, A_1, \dots, A_p]$, $Z \equiv [Z_0, \dots, Z_{T-1}]$, where $Z_t \equiv (1, y'_t, \dots, y'_{t-p+1})'$. Then the Gaussian ML estimator of (A, Σ_u) is $\hat{A} = YZ'(ZZ')^{-1}$ and $\hat{\Sigma}_u = (Y - \hat{A}Z)(Y - \hat{A}Z)' / T$.

Let the prior of the model parameters (A, Σ_u) follow a Gaussian-inverse Wishart distribution of the form

$$\Sigma_u \sim \mathcal{IW}_K(S_*, n), \quad \text{vec}(A) | \Sigma_u \sim \mathcal{N}(\text{vec}(A^*), V_A = V \otimes \Sigma_u),$$

where S_* is a $K \times K$ positive definite covariance matrix and the scalar $n \geq 0$ is a degrees-of-freedom parameter, A^* is the $K \times (Kp + 1)$ prior mean of A , and V is a $(Kp + 1) \times (Kp + 1)$ positive definite matrix.

As derived in Chapter 5, the conjugate posterior distribution is described by

$$\text{vec}(A) | \Sigma_u, Y \sim \mathcal{N}(\text{vec}(\bar{A}), \bar{\Sigma}_A), \quad \Sigma_u | Y \sim \mathcal{IW}_K(S, \tau),$$

where

$$\bar{A} = (A^* V^{-1} + YZ')(V^{-1} + ZZ')^{-1},$$

$$\bar{\Sigma}_A = (V^{-1} + ZZ')^{-1} \otimes \Sigma_u,$$

$$S = T \tilde{\Sigma}_u + S_* + \hat{A} Z Z' \hat{A}' + A^* V^{-1} A^{*\prime} - \bar{A} (V^{-1} + ZZ') \bar{A}',$$

$$\hat{A} = YZ'(ZZ')^{-1},$$

and

$$\tau = T + n.$$

Within this framework, a diffuse prior would involve setting

$$n = 0,$$

$$V^{-1} = 0,$$

$$A^* = 0,$$

$$S_* = 0.$$

Substituting these prior parameters into the expressions for the posterior distribution yields

$$\tau = T,$$

$$\bar{\Sigma}_A = (ZZ')^{-1} \otimes \Sigma_u,$$

$$\bar{A} = \hat{A},$$

$$S = T \tilde{\Sigma}_u.$$

Note that this derivation technically violates the maintained assumption of a positive definite S_* and a positive definite V . It is understood, however, that virtually the same results would be obtained if V were chosen to be very large but finite and if S_* were chosen very close to zero, which helps explain why the shortcut of setting $V^{-1} = 0$ and $S_* = 0$ is commonly used in applied work.

Posterior draws from the inverse Wishart distribution are generated as

$$\Sigma_u^{*r} | Y = (\mathcal{R}\mathcal{R}')^{-1},$$

where \mathcal{R} is a $K \times \tau$ matrix consisting of column vectors independently drawn from $\mathcal{N}(0, S^{-1})$. We use $*$ to denote posterior draws. The draws for Σ_u^{*r} are positive definite by construction. Let $\alpha^* = \text{vec}(A^*)$. Given Σ_u^{*r} , we draw

$$\alpha^{*r} | \Sigma_u^{*r}, Y \sim \mathcal{N}(\text{vec}(\bar{A}), (V^{-1} + ZZ')^{-1} \otimes \Sigma_u^{*r}).$$

Given posterior draws $\sigma^{*r} = \text{vech}(\Sigma_u^{*r})$ and α^{*r} for $r = 1, \dots, R$, it is straightforward to simulate the posterior distribution of the structural impulse response estimator $\hat{\theta}_{ik,h}^{*r} = g(\alpha^{*r}, \sigma^{*r})$. A credible set is often constructed as the interval

$$[\theta_{ik,h,\gamma/2}^*, \theta_{ik,h,1-\gamma/2}^*],$$

where $\theta_{ik,h,\gamma/2}^*$ and $\theta_{ik,h,1-\gamma/2}^*$ are the critical points defined by the $\gamma/2$ and $1 - \gamma/2$ quantiles of the posterior distribution of $\hat{\theta}_{ik,h}^{*r}$. Alternatively, one could form posterior standard error bands

$$\hat{\theta}_{ik,h}^{*r} \pm z_{\gamma/2} \hat{\sigma}(\hat{\theta}_{ik,h}^{*r}), \quad i = 0, 1, 2, \dots,$$

based on the standard deviation of the simulated $\hat{\theta}_{ik,h}^{*r}$ draws. The latter approach, however, would at best be valid asymptotically because the finite-sample distribution of the structural impulse responses tends to be highly non-Gaussian (see, e.g, Phillips 1991; Kilian 1999b). Credible sets for structural forecast error variance decompositions may be constructed along the same lines.

Pointwise Inference on Historical Decompositions. An important difference between frequentist and Bayesian methods of inference is that the latter may also be used to construct pointwise credible sets for historical decompositions if the model is estimated in its structural form (see Baumeister and Hamilton 2015b). Recall the structural VAR representation

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t,$$

where w_t is the vector of mutually uncorrelated structural shocks. Because the posterior of the structural model parameters reflects our subjective uncertainty about these parameters conditional on the data, it is straightforward to assess the posterior uncertainty about the historical decompositions by drawing from the posterior of the structural model parameters, $[B_0, \dots, B_p]$, recomputing the

structural shocks for each draw as

$$w_t = B_0 y_t - B_1 y_{t-1} - \cdots - B_p y_{t-p},$$

where y_1, \dots, y_T denotes the original set of data, and reconstructing the historical decomposition as discussed in Chapter 4. By construction, the cumulative effects of each shock on a given variable will differ across posterior draws, but will add up to the actual demeaned data, allowing us to construct pointwise error bands by simulation.

A similar approach does not work in the frequentist setting because in that setting the only source of uncertainty in $[B_0, \dots, B_p]$ is sampling uncertainty. A historical decomposition conditions on the historical path of the data. Allowing the data to vary in repeated sampling would defeat the purpose of constructing a historical decomposition. If we wish to understand the determinants of the 1982 recession, for example, considering alternative hypothetical histories, in which there is no recession in 1982, is not helpful. Thus, the Bayesian algorithm discussed in Baumeister and Hamilton (2015b) cannot be adapted to a frequentist setting. There are at present no frequentist methods for conducting inference on historical decompositions.

Limitations of Conventional Methods of Bayesian Inference. The widely used Bayesian algorithm based on a conjugate Gaussian-inverse Wishart prior is not without limitations. While we may relax the assumption of a diffuse prior, the derivation of the posterior remains valid only as long as the prior is symmetric across the equations of the VAR system. This requirement is quite restrictive and violated even when using standard Minnesota priors. As a result, prior to the work of Kadiyala and Karlsson (1997), impulse responses from Bayesian models estimated subject to informative priors have typically not been presented with posterior error bands. Kadiyala and Karlsson's work in turn was extended and generalized to identified VAR models subject to informative priors in Sims and Zha (1998) and Baumeister and Hamilton (2015a). The latter methods specify priors directly on the parameters of the structural VAR representation.

Another limitation of standard Bayesian methods of inference is that the Gaussian model underlying the analysis does not allow for conditional heteroskedasticity. While it is possible to relax this assumption, the standard closed-form solutions for the posterior distribution of the reduced-form VAR model parameters no longer apply in that case. Finding the posterior distribution requires simulation methods (see Uhlig 1997). In practice, most applied users of VAR models are content to rely on methods derived for the Gaussian model.

The Role of Unit Roots and Cointegration in Bayesian Inference. Notwithstanding the obvious differences in motivation and interpretation, exact

finite-sample Bayesian error bands or regions of highest posterior density numerically coincide with frequentist confidence intervals in the asymptotic limit, as long as the model is exactly identified based on short-run or long-run restrictions. This result allows even frequentists to appeal to these intervals with suitable changes in interpretation. This result holds only for stationary models, however. In the unit root case, frequentist and Bayesian intervals differ, as illustrated by Sims and Uhlig (1991).

Formally, none of the Bayesian results we discussed thus far hinges on the presence or absence of a unit root, as the unit root model is measure zero from a Bayesian point of view. This does not mean that Bayesians do not have to worry about the possible presence of unit roots or cointegration. The mirror image of short samples being uninformative about long-run behavior is that priors regarding long-run behavior including the presence of unit roots become important. For example, it has been shown that evidence for trends and for autoregressive unit roots can be quite sensitive to the specification of the prior. The implications of VAR model priors for the long-run behavior of the model can be subtle, inadvertent, and not readily apparent.

Strong beliefs in unit root and cointegration restrictions among applied researchers have motivated the development of alternative Bayesian methods of inference that allow the user to shrink the VAR model parameters toward unit root models (see also Chapter 5). The central insight exploited in that literature is that prior information about a regression model can be introduced in the form of extra dummy observations in the data matrix. One way of allowing for the presence of unit roots is known as the sum-of-coefficients prior (see Doan, Litterman, and Sims 1984). This prior may be implemented with suitably chosen dummy observations (see Chapter 5). In the limit, as the prior restrictions are tightened, this prior implies the presence of a unit root with drift in each VAR equation, which rules out cointegration. An alternative is a prior involving a dummy initial observation proposed in Sims (1993), which allows in the limit for an unspecified number of unit roots without drift, possibly smaller than K , accommodating cointegration. For further discussion and references, see Sims and Zha (1998), who show how to construct the posterior of structural impulse responses within this alternative Bayesian framework.

The main point to take away from this discussion is that Bayesian analysis is not immune from the typical concerns that affect frequentist inference in structural VAR models. The difference is that rather than choosing between alternative asymptotic approximations and model specifications, Bayesians need to keep in mind the implications of their choice of prior.

12.11 Joint Inference on Structural Impulse Responses

As in frequentist inference about structural impulse responses, the focus among Bayesian researchers traditionally has been on pointwise inference. Concerns about this practice were first voiced in Sims and Zha (1999).

Confidence bands obtained by connecting pointwise confidence intervals tend to be too narrow and lack coverage accuracy, resulting in spurious findings of statistical significance.

To the extent that the problem of joint impulse response confidence sets has been discussed in the frequentist VAR literature, it has often been reduced to a problem of conducting joint inference across a range of horizons for a given impulse response function. For example, Jordà (2009) proposes one solution to this problem and Staszewska (2007) and Lütkepohl, Staszewska-Bystrova, and Winker (2015a) propose several alternatives. Simulation evidence on the finite-sample accuracy of these confidence bands is discussed in Kilian and Kim (2009) and Lütkepohl, Staszewska-Bystrova, and Winker (2015a).

It is important to stress that these approaches, while representing an important step forward, are too restrictive for applied work. Most users of structural VAR models are interested in conducting inference about multiple impulse response functions at the same time. Sometimes the economic question of interest by construction involves multiple impulse response functions. For example, the question of whether an oil price shock creates stagflation in the domestic economy by necessity involves studying the responses of inflation as well as real output. Likewise, researchers interested in the effects of a U.S. monetary policy shock care about the responses of both real output and inflation because the loss function of the Federal Reserve depends on both real output and inflation.

It is also common for researchers to be interested in assessing the implications of economic theory for a range of different impulse response functions simultaneously. For example, Blanchard (1989) uses a macroeconomic VAR model to evaluate the implication of standard Keynesian models that (1) positive demand innovations increase output and decrease unemployment persistently, and (2) a favorable supply shock triggers an increase in unemployment without a decrease in output. This example involves inference about four impulse response functions simultaneously. There are even cases in which users of structural VAR models are interested in studying the responses of all model variables to all structural shocks simultaneously. A good example is recent structural VAR models of the global market for industrial commodities such as crude oil (e.g., Kilian 2009). A proper solution to this problem requires taking account of the dependence of all structural impulse responses of interest, not just of the responses in a given impulse response function. Next, we review several methods designed for this task.

12.11.1 Joint Confidence Sets for Structural Impulse Responses

The Bonferroni Method. The simplest method of constructing a joint confidence set for structural impulse responses involves constructing Bonferroni bounds. The construction of these bounds is straightforward because it does

not utilize measures of the dependence across the structural impulse response coefficients (see Lütkepohl, Staszewska-Bystrova, and Winker 2015b). Let θ denote the $M \times 1$ vector of structural impulse responses of interest. Recall that $\theta = g(\alpha, \sigma)$, where $g(\cdot)$ denotes a nonlinear transformation of $\alpha = \text{vec}([A_1, \dots, A_p])$ and $\sigma = \text{vech}(\Sigma_u)$. Let $\hat{\eta}$ denote any estimator of $\eta \equiv (\alpha, \sigma)$ such that

$$\sqrt{T}(\hat{\eta} - \eta_0) \xrightarrow{d} \mathcal{N}(0, \Sigma_\eta),$$

where Σ_η is an unknown nonsingular covariance matrix.

The $(1 - \gamma)100\%$ Bonferroni band is

$$B_{1-\gamma} = [l_1, u_1] \times \dots \times [l_M, u_M],$$

where for $m = 1, \dots, M$, the lower bound l_m and the upper bound u_m are the $\gamma/(2M)$ and $(1 - \gamma/(2M))$ quantiles of the marginal distribution of the m^{th} component of $\hat{\theta}$. In practice, this distribution may be approximated by bootstrapping the reduced-form VAR model.

A well-known result is that this joint confidence band is conservative in that

$$\mathbb{P}[\theta \in B_{1-\gamma}] \geq 1 - \gamma.$$

Even for VAR models of moderate dimension, the dimension of θ tends to be large. In a recursively identified model the dimension of $\theta' = [\text{vech}(\Theta_0)', \text{vec}([\Theta_1, \dots, \Theta_H]')']$ is $M = K^2H + K(K + 1)/2$. For example, for $K = 3$ and a maximum horizon of $H = 12$, $M = 114$. For a nominal 68% joint confidence band, this would require choosing individual confidence levels of $(1 - \gamma/M) = 1 - 0.32/114 = 0.9972$. Estimating the tail quantiles therefore may require a very large number of bootstrap draws. As Lütkepohl, Staszewska-Bystrova, and Winker (2015b) observe, in some cases, only the responses to specific shocks are of interest, which helps ameliorate this problem somewhat.

The Wald Approach of Lütkepohl, Staszewska-Bystrova, and Winker (2015b). An alternative approach to constructing joint confidence sets for structural impulse responses was suggested by Lütkepohl, Staszewska-Bystrova, and Winker (2015b). Consider the Wald test statistic

$$W(\eta_0) = T(\hat{\eta} - \eta_0)' \hat{\Sigma}_\eta^{-1} (\hat{\eta} - \eta_0),$$

where $\hat{\Sigma}_\eta$ is an estimator of Σ_η . An approximation of the small-sample distribution of this Wald test statistic may be obtained by bootstrapping the reduced-form vector autoregression. Let $\hat{\eta}^{*r}$ and $\hat{\Sigma}_\eta^{*r}$ denote the r^{th} bootstrap realization, where $r = 1, \dots, R$. Define the bootstrap Wald test statistic

$$W^{*r} = T(\hat{\eta}^{*r} - \hat{\eta})' (\hat{\Sigma}_\eta^{*r})^{-1} (\hat{\eta}^{*r} - \hat{\eta}).$$

Upon ordering the bootstrap estimates of the test statistic such that $W^{*1} \leq \dots \leq W^{*R}$, choose the Wald confidence set for η such that $100\gamma\%$ of the bootstrap realizations W^{*r} exceed the Wald test statistic obtained from the original data:

$$W_{1-\gamma}^\eta = \{\eta | W(\eta) \leq W^{*(1-\gamma)R}\}.$$

As T and R go to infinity, under the usual conditions underlying applications of the bootstrap,

$$\mathbb{P}[W(\eta) \leq W^{*(1-\gamma)R}] \rightarrow 1 - \gamma,$$

and hence asymptotically, for the true parameters η_0 ,

$$\mathbb{P}[\eta_0 \in W_{1-\gamma}^\eta] \rightarrow 1 - \gamma.$$

Given the set of η values in $W_{1-\gamma}^\eta$, the implied joint confidence region for the impulse response estimator $\hat{\theta}$ may be inferred from the mapping $\theta = g(\eta)$. If this function is one-to-one, as would be the case in standard recursively identified models when $H \leq p$, the set

$$W_{1-\gamma}^{\theta=g(\eta)} = \{g(\eta) | \eta \in W_{1-\gamma}^\eta\}$$

is an asymptotically exact $1 - \gamma$ joint confidence set. If $\theta = g(\eta)$ is not one-to-one, this region has at least $1 - \gamma$ confidence level.

Plotting this joint confidence set is impractical in realistic applications. Lütkepohl, Staszewska-Bystrova, and Winker (2015b) therefore propose to represent this joint confidence set in the form of a confidence band. Let $\hat{\theta}_m^{*r} = g_m(\hat{\eta}^{*r})$ denote the r^{th} bootstrap realization of the m^{th} element of θ obtained from $\hat{\eta}^{*r}$. Then a joint confidence band for θ may be constructed by determining the lower bounds

$$l_m = \min \left\{ \hat{\theta}_m^{*r} | \hat{\eta}^{*r} \in W_{1-\gamma}^\eta \right\}$$

and upper bounds

$$u_m = \max \left\{ \hat{\theta}_m^{*r} | \hat{\eta}^{*r} \in W_{1-\gamma}^\eta \right\},$$

for $m = 1, \dots, M$, and defining

$$W_{1-\gamma}^{\theta=g(\eta), \text{band}} = [l_1, u_1] \times \dots \times [l_M, u_M].$$

Values of the structural impulse responses outside of this confidence band can be rejected with $100(1 - \gamma)\%$ confidence. Lütkepohl, Staszewska-Bystrova, and Winker (2015b) provide evidence based on stationary bivariate VAR(1) models with varying degrees of persistence that these joint confidence bands have reasonably accurate coverage rates. They tend to be more conservative

and wider than the Bonferroni bands. For larger models, one would expect the Bonferroni bands to be conservative.

The Wald Approach of Inoue and Kilian (2016). An alternative Wald approach was proposed by Inoue and Kilian (2016). Their proposal is to invert the Wald test statistic for the structural impulse response parameters of interest, denoted by

$$W = (\hat{\theta} - \theta_0)' \hat{\Sigma}_{\hat{\theta}}^{-1} (\hat{\theta} - \theta_0),$$

where $\hat{\Sigma}_{\hat{\theta}}$ is an estimate of the variance-covariance matrix of $\hat{\theta}$. One difficulty in this context is that the asymptotic distribution of $\sqrt{T}(\hat{\theta} - \theta_0)$ is degenerate when the number of structural responses exceeds the number of model parameters. Thus, even in stationary vector autoregressions, the joint asymptotic distribution of $\hat{\theta}$ may be degenerate and hence the distribution of W nonstandard (see, e.g., Lütkepohl and Poskitt 1991). This degeneracy may be overcome by transforming the estimator appropriately. Inoue and Kilian show that the joint Wald test statistic is invariant to this transformation and can be approximated by bootstrap methods, providing a theoretical justification for constructing joint confidence sets for $\hat{\theta}$ from the bootstrap realizations W^{*r} , $r = 1, \dots, R$.

To test a given null hypothesis, $\mathbb{H}_0 : \theta = \theta_0$, the value of the original Wald statistic

$$W = (\hat{\theta} - \theta_0)' \hat{\Sigma}_{\hat{\theta}}^{-1} (\hat{\theta} - \theta_0), \quad (12.11.1)$$

where

$$\hat{\Sigma}_{\hat{\theta}^*} = \frac{1}{R} \sum_{r=1}^R (\hat{\theta}^* - \hat{\theta})(\hat{\theta}^* - \hat{\theta})', \quad (12.11.2)$$

must be compared with the $100(1 - \gamma)$ percentile of the empirical distribution of the bootstrap Wald test statistics

$$W^{*r} = (\hat{\theta}^{*r} - \hat{\theta})' \hat{\Sigma}_{\hat{\theta}^{*r}}^{-1} (\hat{\theta}^{*r} - \hat{\theta}), \quad (12.11.3)$$

for $r = 1, 2, \dots, R$. Generating the bootstrap critical values requires a nested bootstrap loop, because for each bootstrap realization of the Wald statistic the term $\hat{\Sigma}_{\hat{\theta}^{*r}}$ in turn must be evaluated by bootstrap simulation.

In the absence of a specific null value, θ_0 , a joint confidence set $W_{1-\gamma}^\theta$ may be constructed by comparing the value of each of the bootstrap Wald test statistics

$$\tilde{W}^{*r} = (\hat{\theta} - \hat{\theta}^{*r})' \hat{\Sigma}_{\hat{\theta}^*}^{-1} (\hat{\theta} - \hat{\theta}^{*r}), \quad (12.11.4)$$

against the $100(1 - \gamma)$ percentile of the distribution of the bootstrap statistic (12.11.3). If the value of the statistic (12.11.4) is less than this critical value, the

bootstrap draw $\widehat{\theta}^{*r}$ is retained and becomes a member of the joint confidence set $W_{1-\gamma}^\theta$. As T and R go to infinity,

$$\mathbb{P}[\theta_\diamond \in W_{1-\gamma}^\theta] \rightarrow 1 - \gamma,$$

where θ_\diamond denotes the true value of θ . Each realization of θ contained in $W_{1-\gamma}^\theta$ may be viewed as one structural model.

Inoue and Kilian (2016) provide simulation evidence that this bootstrap Wald joint confidence set is reasonably accurate even in large-dimensional and highly persistent VAR models, whereas the Bonferroni method is conservative. The latter result is not unexpected when the number of structural impulse responses is large. Although $W_{1-\gamma}^{\theta=g(n)}$ has theoretical advantages compared with $W_{1-\gamma}^\theta$ because it restricts estimation to the parameters in η , in practice the differences in coverage accuracy tend to be small. Simulation evidence suggests that in the rare cases, in which there is a larger difference in accuracy, the joint confidence region $W_{1-\gamma}^\theta$ is more accurate.

Plotting the joint confidence region $W_{1-\gamma}^\theta$ is impractical in realistic applications. One way of representing the joint confidence set is to form a joint confidence band. Let $\widehat{\theta}_m^{*r}$ denote the r^{th} bootstrap realization of the m^{th} element of θ . Then

$$W_{1-\gamma}^{\theta,\text{band}} = [l_1, u_1] \times \cdots \times [l_M, u_M],$$

where

$$l_m = \min_r \{\widehat{\theta}_m^{*r} | \widehat{\theta}_m^{*r} \in W_{1-\gamma}^\theta\}$$

and

$$u_m = \max_r \{\widehat{\theta}_m^{*r} | \widehat{\theta}_m^{*r} \in W_{1-\gamma}^\theta\}$$

for $m = 1, \dots, M$.

This approach is sufficient to evaluate whether zero responses can be rejected at conventional significance levels, but results in a loss of information because one is no longer able to discern the evolution over time of the impulse response functions associated with any one structural model estimate in the joint confidence set. It is precisely this evolution of the response function that users of structural VAR models typically are interested in (see, e.g., Cochrane 1994). For example, many macroeconomists have abandoned frictionless neoclassical models and adopted models with nominal or real rigidities based on VAR evidence of sluggish or delayed responses of inflation and output (see, e.g., Woodford 2003). This is also true in other applications. Whereas macroeconomists may be interested in whether a response function for real output is hump-shaped or not, users of structural VAR models in international economics may be interested in whether there is delayed overshooting in the response of the exchange rate to monetary policy shocks. It is difficult

to answer such questions about the shape of a given impulse response function based on two-dimensional joint confidence bands, because such bands typically are consistent with a multitude of different response patterns. These difficulties are compounded when considering the analysis of more than one impulse response function at a time, as is common in applied work.

Inoue and Kilian (2016) therefore propose a complementary approach to evaluating joint confidence sets of the form $W_{1-\gamma}^\theta$ (or $W_{1-\gamma}^{\theta=g(\eta)}$) based on plotting all sets of impulse responses associated with the structural models contained in the joint Wald confidence set. As a result, each impulse response function viewed in isolation looks like a shotgun trajectory chart. Unlike conventional representations of confidence sets, this “shotgun plot” may be frayed around the edges. The joint confidence bands discussed earlier can be viewed as an envelope fit around this shotgun trajectory plot. This envelope may be interpreted as a conservative two-dimensional joint confidence band, but that band is only a side product of this approach. The main objective instead is to characterize the shapes and patterns of the set of impulse responses. Although the shotgun trajectory plot does not lend itself to learning about the responses associated with individual structural models in the joint confidence set, it provides important additional information not conveyed by joint confidence bands, as the following two examples from Inoue and Kilian (2016) illustrate.

A VAR Model of Monetary Policy. Consider a quarterly VAR model of monetary policy for inflation, growth in real output, and the short-term interest rate. The model is identified recursively with the interest rate ordered last. One question of economic interest is whether the implied response of real output to an unexpected monetary policy expansion is hump-shaped, as suggested by standard business cycle models. The conventional joint confidence band cannot answer this question, because this band is wide enough to accommodate any number of shapes of this impulse response function, some consistent with economic theory and some not (see Figure 12.2).

Having formally defined a hump-shaped response function (for example, as the response function reaching its maximum between horizon 1 and $H - 1$, possibly strengthened by additional monotonicity restrictions), Inoue and Kilian’s proposal is to run an iterative search on the response functions contained in the 68% joint confidence set to determine whether the response functions in the set are consistent or inconsistent with this hump shape. Evidence that all response functions in the 68% confidence set are hump shaped would lead us to reject the hypothesis that there is no hump in the response function. Evidence that none of the response functions in the set is hump shaped, in contrast, would imply a rejection of the hypothesis of a hump-shaped response. Evidence that some response functions in the set are hump shaped and some are not means that the data are not dispositive about the presence or absence

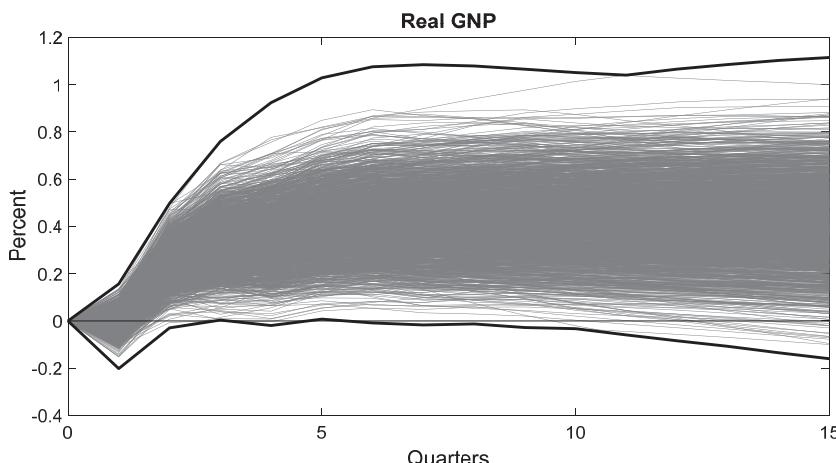


Figure 12.2. Responses of U.S. real GNP to an unexpected loosening of monetary policy: Shotgun plot implied by joint 68% Wald confidence set.

Notes: Obtained by cumulating the responses of the growth rate of real GNP and flipping the sign of the monetary policy shock in the estimated VAR model. The joint confidence band obtained by constructing an envelope around the shotgun plot is shown as bold lines.

Source: Inoue and Kilian (2016).

of a hump shape at the 68% confidence level. This exercise provides additional information about the properties of the estimated model that could not have been inferred from conventional joint confidence bands. As Table 12.1 shows, in this example, the hump shape cannot be ruled out at conventional confidence levels, nor can the hypothesis be rejected that there is no hump shape in the response of real GNP.

A VAR Model of the Stagflationary Effects of Oil Price Shocks. Often the hypotheses of interest to economists involve more than one response function in the same structural model. This situation can be illustrated using the example of a VAR model for the U.S. economy including the real price of oil, inflation (π), and growth in real output (Δgdp). The question of interest is whether oil price shocks have stagflationary effects on the U.S. economy. Answering this question requires us to consider simultaneously two response functions for the same structural model. Specifically, we need to verify whether the responses to an oil price shock of Δgdp_{t+h} and $\Delta \pi_{t+h}$ (the latter having been obtained by differencing the responses of π_{t+h}) are of opposite sign for all horizons of interest for the models in the joint confidence set $W_{1-\gamma}^\theta$ (see, e.g., Kilian 2008c). It is immediately obvious that this type of information cannot be inferred from joint confidence bands, but may be computed based on the shotgun plot. Moreover,

Table 12.1. *Percentage of Models in Joint Wald Confidence Set Consistent with a Hump-Shaped Response Function of Real GNP to an Unexpected Loosening of Monetary Policy*

Definition of hump shape	Joint Confidence Level	
	68%	95%
No piecewise monotonicity imposed	69.0%	70.1%
Piecewise monotonicity imposed	15.6%	15.3%

Notes: Based on a minimal definition of a hump-shaped response as a response reaching its maximum between horizon 1 and 15. The optional piecewise monotonicity constraint restricts the slope of the impulse response function to be positive to the left of the peak response and negative to the right of the peak. Source: Inoue and Kilian (2016).

to the extent that there is evidence of some models in $W_{1-\gamma}^\theta$ exhibiting stagflationary responses, the shotgun plot allows us to assess graphically the extent to which the subset of these models is associated with cumulative reductions in real GDP growth, which is the ultimate question of interest to users of this structural model.

It would be tempting to try to answer this question instead by observing that the data must be consistent with stagflationary responses as long as both joint confidence bands include positive as well as negative values, so hypothetically a stagflationary response fits within these bands. Inoue and Kilian (2016) provide an empirical example, in which nevertheless none of the structural models in question exhibit stagflationary behavior, reinforcing the point that joint confidence bands may obscure important information about the shapes of impulse response functions. Figure 12.3 shows that even when applying the definition of stagflation only to horizons up to one year, there are few stagflationary responses in the 68% confidence set and none that imply a large recessionary effect four years later.

12.11.2 Joint Credible Sets

The same problems of joint inference arise when using Bayesian methods of estimation and inference. A solution to this problem within the conventional Bayesian framework discussed earlier has been proposed by Inoue and Kilian (2013).

Highest-Posterior Density Joint Credible Sets. Consider a structural VAR model that is fully identified by imposing a recursive structure on B_0^{-1} . An example would be the global oil market model of Kilian (2009). When the results apply specifically to recursively identified models, we denote this specific B_0^{-1} by P , where P is the lower-triangular Cholesky decomposition of

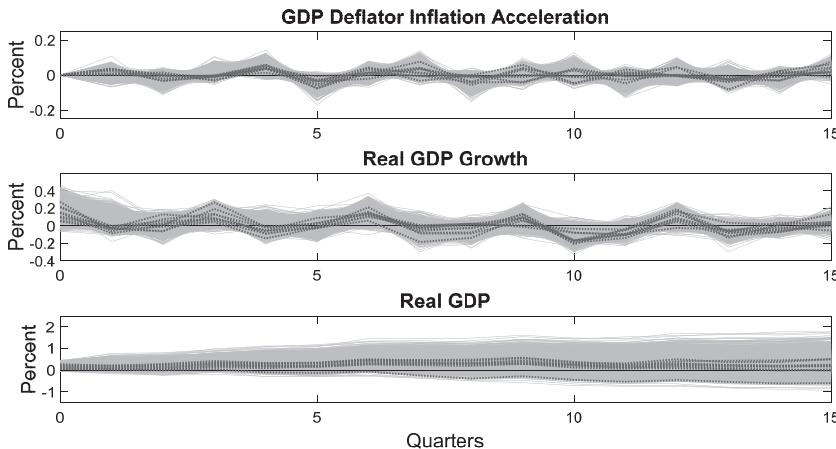


Figure 12.3. Responses to an unexpected increase in the real price of oil: Shotgun plot implied by joint 68% Wald confidence set with subset of stagflationary responses highlighted.

Notes: The pairs of response functions associated with the models in the 68% joint confidence set that exhibit stagflationary responses at horizons 1, …, 4 and a positive inflation acceleration response at horizon 1 are highlighted as dark dotted lines. There are no models with stagflationary responses at all horizons.

Source: Inoue and Kilian (2016).

Σ_u such that $PP' = \Sigma_u$. Any such structural VAR model can be defined in one of two ways: (a) as the set of the first $p + 1$ structural impulse responses $\theta = [\text{vech}(P)', \text{vec}(\Phi_1 P)', \text{vec}(\Phi_2 P)', \dots, \text{vec}(\Phi_p P)']'$, or (b) as the vector $(\text{vec}(A)', \text{vech}(P)')'$, where $A \equiv [A_1, \dots, A_p]$. In fact, there is a one-to-one mapping between these representations, given that the nonlinear function $\theta = h(\text{vec}(A), \text{vech}(P))$ is known and Σ_u is nonsingular. This fact allows the derivation of the joint distribution of θ by the change-of-variable method, if the joint distribution of $(\text{vec}(A)', \text{vech}(P)')$ is Gaussian-inverse Wishart, as postulated earlier. The posterior density takes the form

$$f(\theta) \propto \left(\left| \frac{\partial \theta}{\partial [\text{vec}(A)', \text{vech}(P)']} \right| \right)^{-1} \left| \frac{\partial \text{vech}(\Sigma_u)}{\partial [\text{vech}(P)']} \right| f(A|\Sigma_u) f(\Sigma_u),$$

where the determinant of the Jacobian has a closed form. This result allows us to rank each structural model θ by its posterior density value. Given this ranking, a $100(1 - \gamma)\%$ highest posterior density (HPD) credible set can be defined as

$$S = \{\theta | f(\theta) \geq c_\gamma\}, \quad (12.11.5)$$

where $f(\theta)$ is the posterior density of θ and c_γ is the largest constant such that

$$\mathbb{P}(S) \geq 1 - \gamma.$$

In practice, we rank each structural model θ^{*r} , $r = 1, \dots, R$, obtained by drawing from the Gaussian-inverse Wishart posterior by its posterior density value $f(\theta^{*r})$, which can be readily computed given $(\alpha^{*r}, \sigma^{*r})$, where $\alpha^{*r} = \text{vec}(A^{*r})$ and $\sigma^{*r} = \text{vech}(\Sigma_u^{*r})$. Then a 68% joint credible set, for example, would be constructed by simply retaining the 68% of the posterior draws θ^{*r} with the highest $f(\theta^{*r})$ values. These credible sets differ from conventional error bands for structural impulse responses in that the elements of the credible set consist of all K^2 structural impulse response functions up to some prespecified horizon, generated by a given structural model draw. When plotting the K^2 structural impulse response functions associated with each member of the credible set, the resulting plot of the credible set will exhibit a shotgun trajectory pattern, as discussed earlier. The evaluation of this credible set may proceed analogously to that of the joint confidence set in Inoue and Kilian (2016).

This baseline result applies to models that are fully identified, as opposed to partially identified. Partial identification here relates to a situation in which only a subset of the structural shocks is identified. For example, in many VAR models of monetary policy, we are only interested in identifying responses to monetary policy shocks with the other shocks remaining unidentified. Ranking structural models that are only partially identified requires the marginalization of the joint posterior $f\theta$ with respect to the model parameters that do not enter the posterior distribution of the structural impulse responses of interest, before joint credible sets may be constructed (see Inoue and Kilian 2013).

Although our discussion focused on recursively identified structural VAR models, with suitable changes in notation this approach can be generalized to non-recursive identification schemes. The latter results, of course, must be worked out on a case-by-case basis, depending on the nature of the identifying restrictions.

The Sims-Zha Method. An alternative Bayesian solution to the problem of joint inference was proposed by Sims and Zha (1999). Their baseline method involves draws from the first and second moment of the (unspecified) joint distribution of the posterior impulse responses, which will be a good approximation only if the joint distribution is close to Gaussian. One concern is that it is well known that the finite-sample joint posterior distribution of impulse responses is far from Gaussian, so this method will be at best an approximation, whereas the method in Inoue and Kilian (2013) is exact by construction. The other concern is that the marginalized posterior is Gaussian only if the joint posterior is Gaussian, making it problematic to conduct such joint inference in partially identified models without doing Monte Carlo integration by numerical methods.

In recognition of the first concern, Sims and Zha suggest an alternative method designed to account for asymmetry in the marginal distribution of the

impulse responses. This method is based on quantiles of the marginal distribution constructed from the same draws used for the baseline method. This proposal indeed accounts for asymmetry in the approximate marginal distribution, but it does not address the second concern regarding the non-Gaussianity of the underlying joint distribution.

Optimal Pointwise versus Optimal Joint Inference. Baumeister and Hamilton (2015b) recently argued that pointwise inference on the structural impulse responses is preferred in applied work because it is optimal from a Bayesian point of view. They made the case for reporting the vector of pointwise medians or means of the posterior of the structural responses. This approach, however, is optimal only under loss functions that imply that the user is not concerned with the dependence of the structural responses across horizons and across response functions. In other words, the user does not care about the shapes of the impulse response functions or the comovement across response functions.

Specifically, Baumeister and Hamilton (2015b) stress that under the loss function

$$\omega_0|\theta_{jk,0} - \hat{\theta}_{jk,0}| + \cdots + \omega_h|\theta_{jk,h} - \hat{\theta}_{jk,h}|,$$

the optimal estimate, for any set of positive weights $\{\omega_i\}_{i=0}^h$, is the vector of pointwise posterior medians. Although this solution is optimal conditional on this specific choice of loss function, of course, so are the solutions obtained under alternative loss functions. What is at issue therefore is not the optimality of the solution to a given loss function, but the appropriate choice of the loss function.

The solution to the loss function underlying the construction of the modal model in Inoue and Kilian (2013), for example, is as optimal as the solution proposed in Baumeister and Hamilton (2015b), but it is derived under a different loss function that arguably is more in line with the interests of applied researchers because it takes account of the dependence across structural responses. Thus, it avoids the problems of interpreting median response functions documented in Fry and Pagan (2011), Kilian and Murphy (2012), and Inoue and Kilian (2013).

Summary. It is rare for researchers to be interested in pointwise inference on structural impulse responses or related statistical objects. The conventional approach of constructing confidence bands for structural impulse responses from pointwise confidence intervals is known to be potentially misleading. With the recent introduction of methods of constructing truly joint confidence regions one would expect pointwise methods to be gradually replaced

in applied work. If we are interested in constructing joint confidence sets, both the Bonferroni method and the approach based on the Wald test statistic may be used. Simulation evidence in Inoue and Kilian (2016) suggests that both types of joint confidence bands have reasonably accurate coverage rates in realistic settings, even when the DGP is highly persistent and the maximum horizon considered is high. The coverage of Bonferroni-based methods of joint inference, however, tends to be conservative when the number of structural impulse responses of interest is large.

For a given confidence level, joint confidence sets tend to be wider than pointwise intervals if the same confidence level is used, making it more difficult for applied users to produce statistically significant results. It is, however, important to account for the true uncertainty in these estimates. In addition, joint inference may be used to investigate the shapes and patterns of impulse response functions, as emphasized in the work of Inoue and Kilian (2016), allowing researchers to address more directly the implications of economic theory.

12.12 Other Bootstrap Applications

It should be noted that the bootstrap methods discussed in this chapter are intended for approximating the unconditional distribution of statistics such as structural impulse responses or structural forecast error variance decompositions. These methods are not suitable, for example, for generating bootstrap prediction intervals because in prediction we must condition on the same last p observations as in the actual data (see Chapter 2).

12.12.1 Bootstrap Prediction

There are at present no bootstrap methods for approximating the distribution of the data conditional on the p most recent observations that do not effectively require Gaussian model innovations (see, e.g., Berkowitz and Kilian 2000). One practical solution to this problem is to bootstrap the unconditional distribution of the slope parameters, as discussed earlier, exploiting the fact that their conditional distribution under stationarity will converge to the unconditional distribution asymptotically. Consider, for example, the problem of constructing the predictive density of the reduced-form $\text{VAR}(p)$ model

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where $u_t \stackrel{iid}{\sim} F$. In that case, we can approximate the DGP by

$$y_t^* = \hat{v} + \hat{A}_1 y_{t-1}^* + \cdots + \hat{A}_p y_{t-p}^* + u_t^*,$$

where u_t^* is generated by drawing with replacement from the estimated residuals. Having generated a draw for $[\hat{v}^*, \hat{A}_1^*, \dots, \hat{A}_p^*]$ from this bootstrap DGP, for each such draw we construct predictions by iterating forward the model conditional on the p most recent observations:

$$\begin{aligned} y_{t+1}^* &= \hat{v}^* + \hat{A}_1^* y_t + \hat{A}_2^* y_{t-1} + \dots + \hat{A}_{11}^* y_{t-p+1} + u_{t+1}^* \\ y_{t+2}^* &= \hat{v}^* + \hat{A}_1^* y_{t+1}^* + \hat{A}_2^* y_t + \dots + \hat{A}_{11}^* y_{t-p+2} + u_{t+2}^* \\ &\vdots \end{aligned}$$

Once the predictive distribution has been simulated by bootstrap, the construction of prediction intervals based on the percentiles of the predictive distribution is straightforward. It is useful to keep in mind that prediction intervals differ from confidence intervals in that they relate to a random variable (rather than a model parameter) and hence are intended to include future realizations with a prespecified probability. Pascual, Romo, and Ruiz (2004) provide a proof of the asymptotic validity of this approach in the univariate AR(p) model. Kim (1999, 2004) provides evidence that the accuracy of bootstrap prediction intervals may be improved by relying on the bias adjustments proposed by Kilian (1998c) for structural impulse response analysis. As in the case of impulse response analysis, the case can be made that inference should be joint across the horizons of the forecasts rather than pointwise. Wolf and Wunderli (2015) discuss the construction of joint bootstrap prediction regions for VAR models and VECMs and prove their asymptotic validity.

12.12.2 Bootstrapping the Critical Values of Test Statistics

The bootstrap algorithms discussed so far must also be modified if we are interested in constructing critical values for test statistics. In that case, the bootstrap DGP is obtained from the best-fitting model under the null hypothesis rather than from the best-fitting unrestricted model. Consider the example of testing for Granger causality in a bivariate stationary VAR(p) model of the form

$$\begin{pmatrix} y_{1t} \\ y_{2t} \end{pmatrix} = \begin{pmatrix} v_1 \\ v_2 \end{pmatrix} + \sum_{i=1}^p \begin{bmatrix} a_{11,i} & a_{12,i} \\ a_{21,i} & a_{22,i} \end{bmatrix} \begin{pmatrix} y_{1,t-i} \\ y_{2,t-i} \end{pmatrix} + \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix}.$$

Under the null hypothesis that y_{2t} is not Granger-causal for y_{1t} , we have $a_{12,i} = 0$, $i = 1, 2, \dots, p$. Thus, the bootstrap DGP under the null hypothesis reduces to

$$\begin{pmatrix} y_{1t}^* \\ y_{2t}^* \end{pmatrix} = \begin{pmatrix} \hat{v}_1 \\ \hat{v}_2 \end{pmatrix} + \sum_{i=1}^p \begin{bmatrix} \hat{a}_{11,i} & 0 \\ \hat{a}_{21,i} & \hat{a}_{22,i} \end{bmatrix} \begin{pmatrix} y_{1,t-i}^* \\ y_{2,t-i}^* \end{pmatrix} + \begin{pmatrix} u_{1t}^* \\ u_{2t}^* \end{pmatrix},$$

where the coefficient estimates are obtained by restricted ML or restricted GLS methods (see Chapter 2). The bootstrap innovations are generated by drawing

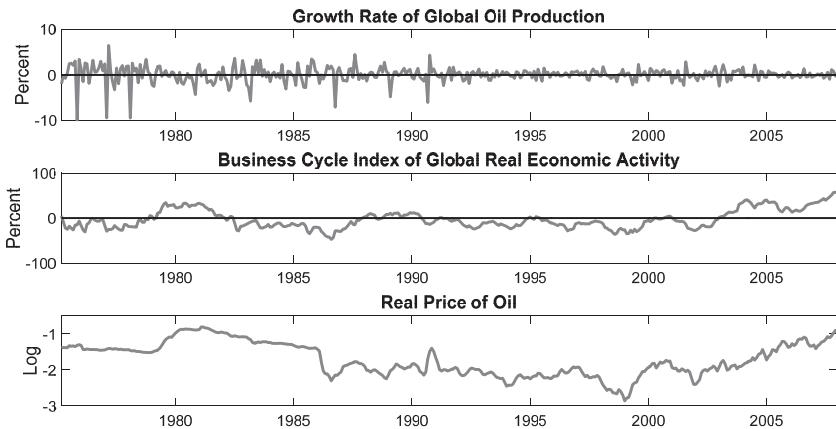


Figure 12.4. Global oil market data.

Notes: The sample period is 1973m2–2007m12. For a description of the data sources, see Kilian (2009).

with replacement from the estimated residuals. The power of the test may be improved by utilizing residuals from the unrestricted model rather than the restricted model, because the latter are consistent estimates under both the null and under the alternative hypothesis, whereas the former are consistent estimates only under the null hypothesis.

12.13 Examples of Impulse Response Confidence Intervals

12.13.1 An Exactly Identified Model

For expository purposes we consider the recursively identified monthly structural VAR(24) model of the global market for crude oil of Kilian (2009), as already discussed in Chapter 8. Recall that $y_t = (\Delta prod_t, rea_t, rpoil_t)$, where $\Delta prod_t$ denotes the percent change in world crude oil production, rea_t is a business cycle index measuring global real economic activity, and $rpoil_t$ is the log of the real price of oil. The model is estimated by LS with an intercept included. The estimation period is 1973m2–2007m12. Figure 12.4 shows the $T - 24$ net sample observations of the transformed time series. The dominant root of the estimated process is 0.989. The identifying assumptions imply that

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} \end{bmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix},$$

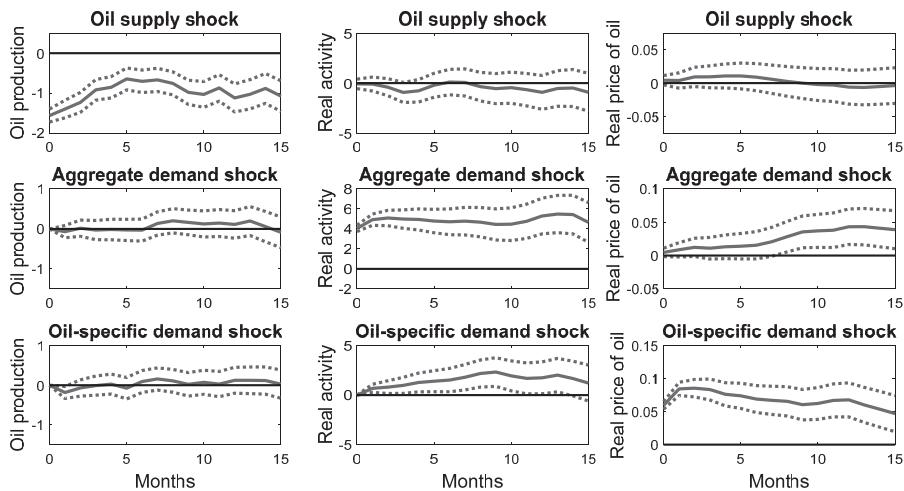


Figure 12.5. 95% delta method confidence intervals based on bootstrap standard error estimates.

Notes: All estimates were generated using the data in Figure 12.4 and the recursively identified structural VAR model in Kilian (2009).

allowing us to estimate B_0^{-1} as the lower-triangular Cholesky decomposition of $\widehat{\Sigma}_u$.

Pointwise confidence intervals for the structural impulse responses can be generated by a wide range of alternative methods. Figure 12.5 plots the point estimates of the structural impulse responses together with 95% pointwise confidence intervals constructed by the delta method. Rather than relying on asymptotic closed-form solutions for the standard errors of the structural impulse responses, we generated estimates of the standard errors from 500 bootstrap replications of the estimated VAR(24) process under the assumption that the innovations are iid with unknown distribution.

The oil supply shocks in Figure 12.5 have been normalized to represent responses to unexpected oil supply disruptions. Because the lower-triangular Cholesky decomposition generates responses to positive one-standard-deviation shocks, this normalization involves multiplying all responses to oil supply shocks by -1 . The responses of global oil production are obtained by cumulating the responses of the growth rate of global oil production.

Figure 12.5 shows that negative oil supply shocks are associated with a statistically significant decline in oil production at all horizons, but the positive responses of the real price of oil and the negative responses of global real activity are not individually statistically significant. In contrast, positive aggregate demand shocks (defined as demand shocks that affect both the market for crude oil and the market for other globally traded industrial commodities) have a statistically significant positive effect on global real activity at all horizons

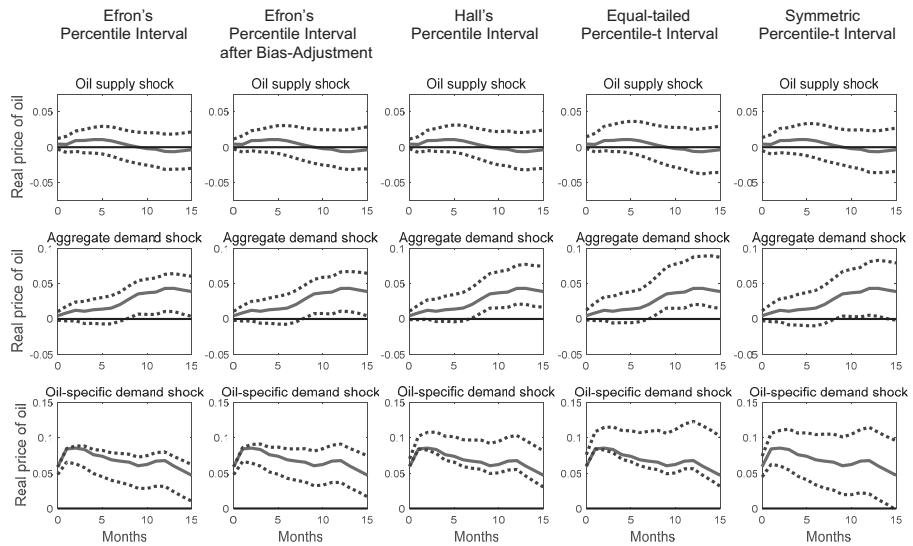


Figure 12.6. Alternative 95% bootstrap confidence intervals.

Notes: All estimates were generated using the data in Figure 12.4 and the recursively identified structural VAR model in Kilian (2009).

and cause a statistically significant increase in the real price of oil at horizons of eight or more months. They do not have a statistically significant effect on global oil production at any horizon, however. Finally, positive oil-market-specific demand shocks trigger a highly statistically significant increase in the real price of oil at all horizons. They are also associated with statistically significant increases in global real activity at horizons up to 13 months and a statistically significant decline in global oil production at horizon 1.

For comparison, Figure 12.6 presents the corresponding pointwise 95% bootstrap confidence intervals obtained by applying a recursive-design bootstrap algorithm under the premise that the VAR model innovations are iid. Estimates of the bootstrap interval endpoints are generated from 2,000 bootstrap replications. To conserve space, only the responses of the real price of oil are shown.

The first column of Figure 12.6 shows confidence intervals constructed from Efron's percentile interval. These interval endpoints tend to be downward biased relative to the point estimate, as is evident especially in the response of the real price of oil to the oil-market specific demand shock. The second column shows results for Efron's percentile interval obtained after applying the first-order mean bias correction to the slope parameters, as proposed in Kilian (1999b). The bias estimates were constructed using the closed-form solution of Pope (1990), as discussed in Section 2.3.3. As a result of the bias adjustment,

the interval endpoints tend to be shifted upward, as is especially evident for the response of the real price of oil to the oil-specific demand shock.¹³

Hall's percentile interval applied to the original LS estimate of the slope parameters embodies an implicit bias adjustment for the structural impulse response estimator, which causes an even more dramatic upward shift in the bootstrap confidence interval endpoint, as shown in the third column. These adjustments are further compounded when using the equal-tailed percentile- t interval in the fourth column. In contrast, the lower interval endpoints of the symmetric percentile- t interval in the last column are closer to those in the second column, while the upper interval endpoints tend to be considerably higher. The standard error estimates for the symmetric percentile- t intervals were obtained by bootstrapping each bootstrap estimate of the VAR model.

Notwithstanding these differences, from an economic point of view, all methods agree that a positive aggregate demand shock causes a statistically significant increase in the real price of oil at horizons beyond about half a year, that negative oil supply shocks do not have a statistically significant effect on the real price of oil, and that positive oil-market-specific demand shocks cause a highly statistically significant increase in the real price of oil at all horizons.

12.13.2 Guarding against Conditional Heteroskedasticity

The case can be made that at least some of the innovations in the Kilian (2009) oil market model are conditionally heteroskedastic. This fact suggests that allowing for conditional heteroskedasticity may improve the accuracy of inference. Figure 12.7 compares three alternative versions of Hall's percentile interval constructed using a recursive design bootstrap under the assumption of iid errors (as in Figure 12.6), constructed using the multivariate generalization of the recursive-design residual wild bootstrap of Gonçalves and Kilian (2004), and constructed using the recursive design residual block bootstrap of Brüggemann, Jentsch, and Trenkler (2016). We use a block length $l = 36$. The results are not very sensitive to the choice of l . Figure 12.7 illustrates that often the wild bootstrap implies intervals that are wider than the iid bootstrap, as expected. The residual-based block bootstrap, which is the only method designed to fully protect against conditional heteroskedasticity at all horizons, tends to imply even wider intervals. Overall, nevertheless, the results are quite similar.

¹³ It should be noted that it is possible for the least-squares point estimate to be above the upper interval endpoint. Likewise, it is possible for the LS estimator to lie below the lower interval endpoint. Only the delta method interval and the symmetric percentile- t interval impose symmetry on the interval, preventing such situations from arising. This does not mean that there is anything wrong with confidence intervals not imposing symmetry. Rather, evidence of the point estimate lying outside the confidence band could be an indication of the LS point estimate being biased.

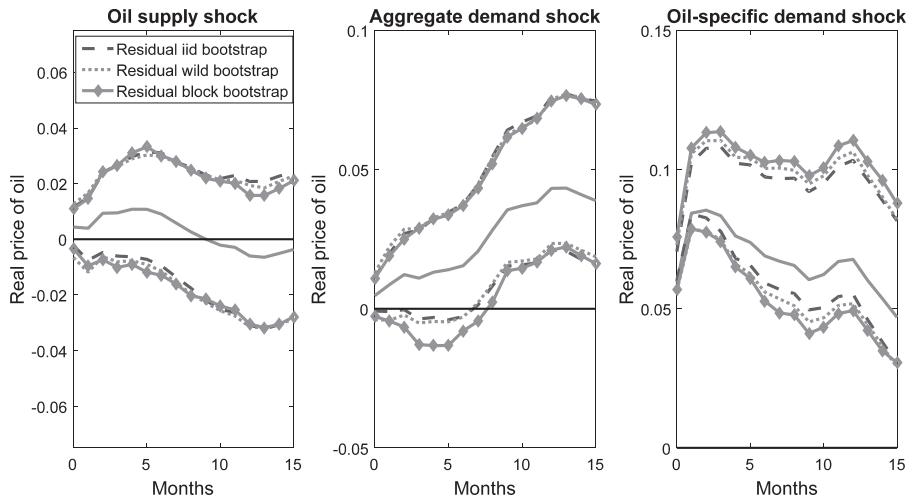


Figure 12.7. Responses of the real price of oil to oil demand and supply shocks with alternative 95% confidence intervals.

Notes: All results are based on Hall's percentile interval. The residual block bootstrap relies on block length $l = 36$. Other block sizes produced similar results.

12.13.3 Extensions to Overidentified Models

Sometimes, a researcher is interested in imposing one or more overidentifying restrictions on an already exactly identified structural VAR model. A case in point is the Kilian (2009) model. Recall that

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} \end{bmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix}.$$

Given that we already imposed the identifying restriction that an oil-specific demand shock does not lower global real economic activity on impact, it would seem natural to impose a similar restriction on the impact effect of oil supply shocks on global real economic activity. This reasoning suggests the presence of an overidentifying restriction such that $b_0^{21} = 0$:

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 \\ 0 & b_0^{22} & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} \end{bmatrix} \begin{pmatrix} w_t^{\text{oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix}.$$

After imposing this extra restriction, B_0^{-1} is no longer recursive and standard method-of-moments estimators no longer apply because there are more identifying restrictions than unknowns.

Estimation Subject to Overidentifying Restrictions. As discussed in Chapter 9, constructing the GMM estimator of overidentified models requires choosing the unknown elements of B_0^{-1} to minimize

$$J = T(\text{vech}(\widehat{\Sigma}_u) - \text{vech}(B_0^{-1}B_0^{-1}'))' \widehat{W} (\text{vech}(\widehat{\Sigma}_u) - \text{vech}(B_0^{-1}B_0^{-1}')),$$

where all identifying and overidentifying restrictions have been imposed on B_0^{-1} and \widehat{W} is the inverse of the estimate of the variance-covariance matrix of the sample moment restrictions.

Recall that we wish to impose four exclusion restrictions such that

$$B_0^{-1} = \begin{bmatrix} b_0^{11} & 0 & 0 \\ 0 & b_0^{22} & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} \end{bmatrix}.$$

There are five unknown elements in B_0^{-1} . In the absence of the overidentifying restriction

$$\widehat{P} = \text{chol}(\widehat{\Sigma}_u) = \begin{bmatrix} 1.5617 & 0 & 0 \\ 0.0735 & 4.0588 & 0 \\ -0.0044 & 0.0047 & 0.0594 \end{bmatrix}.$$

A reasonable initial guess for minimizing J therefore is

$$\begin{bmatrix} 1.5617 & 0 & 0 \\ 0 & 4.0588 & 0 \\ -0.0044 & 0.0047 & 0.0594 \end{bmatrix},$$

where the overidentifying restriction has simply been imposed on the (1, 2) element of \widehat{P} .

Next we need to construct the weighting matrix. Let $u_t = (u_{1,t}, u_{2,t}, u_{3,t})'$. Then $\widehat{W} = (\overline{W}'\overline{W}/T)^{-1}$, where \overline{W} is obtained by demeaning the elements of the matrix

$$\begin{bmatrix} \widehat{u}_{1,1}^2 & \widehat{u}_{1,1}\widehat{u}_{2,1} & \widehat{u}_{1,1}\widehat{u}_{3,1} & \widehat{u}_{2,1}^2 & \widehat{u}_{2,1}\widehat{u}_{3,1} & \widehat{u}_{3,1}^2 \\ \widehat{u}_{1,2}^2 & \widehat{u}_{1,2}\widehat{u}_{2,2} & \widehat{u}_{1,2}\widehat{u}_{3,2} & \widehat{u}_{2,2}^2 & \widehat{u}_{2,2}\widehat{u}_{3,2} & \widehat{u}_{3,2}^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \widehat{u}_{1,T}^2 & \widehat{u}_{1,T}\widehat{u}_{2,T} & \widehat{u}_{1,T}\widehat{u}_{3,T} & \widehat{u}_{2,T}^2 & \widehat{u}_{2,T}\widehat{u}_{3,T} & \widehat{u}_{3,T}^2 \end{bmatrix}$$

which is constructed by stacking vectors of observations for the $K(K + 1)/2$ elements of $\text{vech}(\hat{u}_t \hat{u}'_t)$. \bar{W} in this example is of dimension $T \times 6$. Thus

$$\hat{W} = \begin{bmatrix} 0.0000 & 0.0000 & 0.0001 & -0.0000 & 0.0000 & -0.0003 \\ 0.0000 & 0.0000 & 0.0000 & -0.0000 & 0.0000 & 0.0001 \\ 0.0001 & 0.0000 & 0.0167 & -0.0000 & -0.0002 & 0.1078 \\ -0.0000 & -0.0000 & -0.0000 & 0.0000 & -0.0000 & -0.0002 \\ 0.0000 & 0.0000 & -0.0002 & -0.0000 & 0.0027 & -0.0291 \\ -0.0003 & 0.0001 & 0.1078 & -0.0002 & -0.0291 & 3.5177 \end{bmatrix} \times 10^4.$$

Given \hat{W} and the initial guess for B_0^{-1} , the expression for J may be iterated to convergence using a suitable numerical optimizing routine.¹⁴ We obtain the GMM estimator

$$\hat{B}_0^{-1} = \begin{bmatrix} 1.5624 & 0 & 0 \\ 0 & 4.0604 & 0 \\ -0.0044 & 0.0047 & 0.0594 \end{bmatrix}.$$

The J statistic is minimized with respect to the unknown elements of B_0^{-1} at $\hat{J} = 0.1311$.

The point estimates in Figure 12.8 show that not only the estimate of the structural impact multiplier matrix B_0^{-1} but also the implied structural impulse response functions are quite similar to those in the recursively identified model.

Testing for Overidentifying Restrictions. Before imposing an overidentifying restriction, it is useful to test its validity. It has to be stressed that this test is conditional on the remaining identifying assumptions being valid. It is not a test of the validity of the set of identifying restrictions more generally. As discussed in Chapter 9, under the null hypothesis that the overidentifying restrictions are correct,

$$\hat{J} \xrightarrow{d} \chi^2(n),$$

where n is the number of overidentifying restrictions. In our empirical example, $n = 1$ and we are unable to reject the overidentifying restriction, given the asymptotic p -value of 0.7173.

Inference Subject to Overidentifying Restrictions. Bootstrap inference for the GMM estimator of the structural impulse responses in overidentified structural VAR models may be conducted based on the results in Hall and Horowitz (1996). The key difference compared with the bootstrap methods already

¹⁴ For example, the function *fminsearch* in the MATLAB optimization toolbox finds a vector x such that a given function $f(x)$ is minimized.

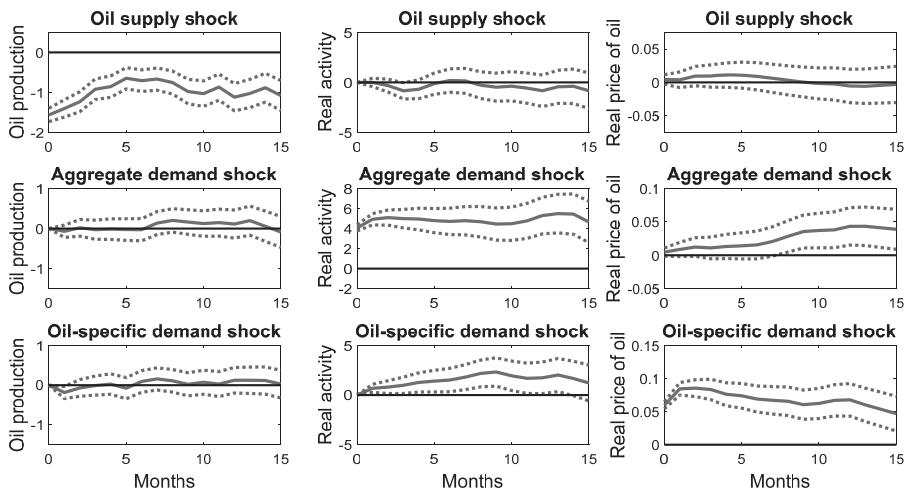


Figure 12.8. 95% delta method confidence intervals based on bootstrap standard error estimates in the overidentified model.

Notes: All estimates were generated using the data in Figure 12.4 and the recursively identified structural VAR model in Kilian (2009) with the overidentifying restriction $b_0^{21} = 0$ imposed.

discussed for exactly identified structural VAR models is that the bootstrap moment conditions must be recentered. This recentering is necessary because in overidentified models the population moment conditions hold with equality, whereas the sample moment conditions evaluated at the GMM estimator by construction do not. Thus, we need to recenter the bootstrap moment conditions and choose the unknown elements of B_0^{*-1} to minimize

$$J^* = T(\text{vech}(\widehat{\Sigma}_u^*) - \text{vech}(B_0^{*-1} B_0^{*-1'}) - c)' \widehat{W}^* \\ \times (\text{vech}(\widehat{\Sigma}_u^*) - \text{vech}(B_0^{*-1} B_0^{*-1'}) - c),$$

where $c = \text{vech}(\widehat{\Sigma}_u) - \text{vech}(\widehat{B}_0^{-1} \widehat{B}_0^{-1'})$ is a vector of constants and \widehat{B}_0^{-1} denotes the GMM estimate of B_0^{-1} . Otherwise, the construction of bootstrap confidence intervals proceeds exactly as discussed earlier. Figure 12.8 shows the GMM point estimates of the structural impulse responses together with pointwise 95% delta method confidence intervals. The standard errors of the impulse response estimates were estimated by bootstrap methods with the bootstrap moment condition suitably recentered. The results are virtually indistinguishable from those in Figure 12.5, which did not impose the overidentifying restriction.

The most common use of overidentifying restrictions in the literature has been to discriminate between alternative models of the transmission of monetary policy shocks by considering a structural VAR model that encompasses

multiple channels of transmission, the existence of which implies additional overidentifying restrictions (see, e.g., Bernanke 1986; Sims 1986; Bernanke and Mihov 1998b; Gordon and Leeper 1994; Leeper, Sims, and Zha 1996). Other applications of overidentified structural VAR models include Galí (1992), Fisher (2006), and Inoue, Kilian, and Kiraz (2009).

Alternative Estimation Methods for Overidentified Models. As discussed in Chapter 9, the model of interest in this empirical example cannot be estimated by IV methods because the identifying restrictions are imposed on B_0^{-1} rather than B_0 . An alternative estimation approach would be to employ the FIML estimator for overidentified structural VAR models.

Another alternative to the GMM estimator for overidentified structural VAR models is the use of Bayesian estimation methods. As in the case of bootstrap methods, care must be exercised in evaluating the posterior of the structural impulse responses in the overidentified case. The question of how to conduct inference in overidentified structural VAR models from a Bayesian point of view was first discussed in Sims and Zha (1998, 1999). Sims and Zha showed that standard methods of computing regions of high posterior density for structural impulse responses are not valid when the structural VAR model is overidentified, because overidentifying restrictions imply restrictions on the error covariance matrix. As discussed in Chapter 9, in this case it is necessary to derive the posterior from the structural VAR representation rather than the reduced-form VAR representation. The posterior of the structural impulse responses is no longer Gaussian-inverse Wishart and has to be evaluated using an importance sampler or Metropolis-Gibbs sampler.¹⁵

These Bayesian methods, however, are designed for structural VAR models with linear identifying restrictions on B_0 and cannot be used in the current context. It is not clear whether the generalized Bayesian estimation method by Canova and Pérez Forero (2015), which allows particular nonlinear restrictions on B_0 , could be adapted to the estimation of overidentified structural VAR models with restrictions on B_0^{-1} .

¹⁵ In related work, Waggoner and Zha (2003) propose a Gibbs sampler. The latter approach is applicable to a wider class of overidentified models including models with restrictions on lagged slope coefficients. Such restrictions may arise from exogeneity restrictions on the VAR model, for example, and differ from the class of overidentifying restrictions on the error covariance matrix considered here.

13 Identification by Sign Restrictions

The approach of using sign restrictions to identify structural VAR models was pioneered by Faust (1998), Canova and De Nicolo (2002), and Uhlig (2005). This approach has become increasingly popular in applied work as an alternative to traditional approaches to identification based on exclusion restrictions.

13.1 A Model of Demand and Supply

To understand why this approach is attractive from an economic point of view, consider a simple bivariate model of a goods market with a demand shock (w_t^{demand}) and a supply shock (w_t^{supply}). The observables are price (p_t) and quantity (q_t). Let us write the relation between the reduced-form residuals of a VAR model and the structural shocks as $u_t = B_0^{-1} w_t$, as in Chapter 8, where in our present example $u_t = (u_t^q, u_t^p)'$ and $w_t = (w_t^{\text{supply}}, w_t^{\text{demand}})'$. The implication of an exogenous shift in the demand or supply curve, respectively, for u_t^p and u_t^q depend on the slopes of the demand and supply curves, which may range from flat to vertical (see Figure 13.1).

A common approach of identifying the effects of demand and supply shocks by short-run exclusion restrictions relies on the short-run supply curve being vertical. In this case, demand shocks have no contemporaneous effect on quantity, which implies one exclusion restriction,

$$\begin{pmatrix} u_t^q \\ u_t^p \end{pmatrix} = \begin{pmatrix} * & 0 \\ * & * \end{pmatrix} \begin{pmatrix} w_t^{\text{supply}} \\ w_t^{\text{demand}} \end{pmatrix}, \quad (13.1.1)$$

where asterisks denote unrestricted elements. This amounts to imposing that production does not respond within the impact period to a price increase triggered by a demand shock (see Figure 13.2). While this restriction may be reasonable in some contexts, as discussed in Chapter 8, in typical situations such recursive models are difficult to justify from an economic point of view.

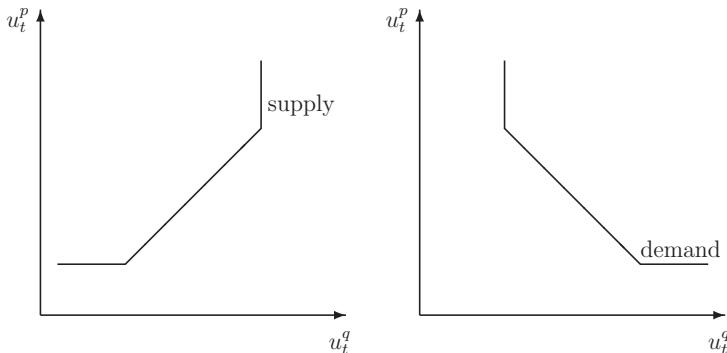


Figure 13.1. Demand and supply curves may have different slopes.

In general, all we know is that, under reasonable assumptions, a supply shock (represented by an exogenous shift to the right of the supply curve along the demand curve) will increase quantity and reduce the price, whereas a demand shock (represented as an exogenous shift to the right of the demand curve along the supply curve) will increase both price and quantity. These implications of economic theory relate to the sign of the responses of price and quantity to demand and supply shocks, respectively, and may be used for identification. Specifically, we may postulate that

$$\begin{pmatrix} u_t^q \\ u_t^P \end{pmatrix} = \begin{bmatrix} + & + \\ - & + \end{bmatrix} \begin{pmatrix} w_t^{\text{supply}} \\ w_t^{\text{demand}} \end{pmatrix}, \quad (13.1.2)$$

where + and – indicates a strictly positive and a strictly negative sign of the parameters in the structural impact multiplier matrix. Intuitively, knowing that positive supply shocks and positive demand shocks move the price in opposite directions, but quantity in the same direction helps us differentiate shifts of the demand curve from shifts of the supply curve (see Figure 13.3). If two shocks shared the same sign pattern, in contrast, one would be unable to identify them separately.

One key difference from the recursive model is that the parameters of the impact multiplier matrix in the sign-identified model are no longer point identified, but set identified. This means that even with an infinite amount of data we will only be able to bound the parameters of interest. As discussed later in this chapter, this fact greatly complicates estimation and inference in sign-identified VAR models.

In some studies, sign restrictions are represented as weak inequalities. The reason why weak inequalities are not permitted here is that identification would be lost if both weak inequalities were binding. For example, a solution for B_0^{-1}

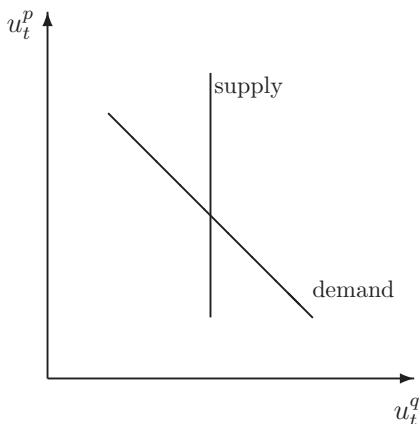


Figure 13.2. The case of a vertical supply curve.

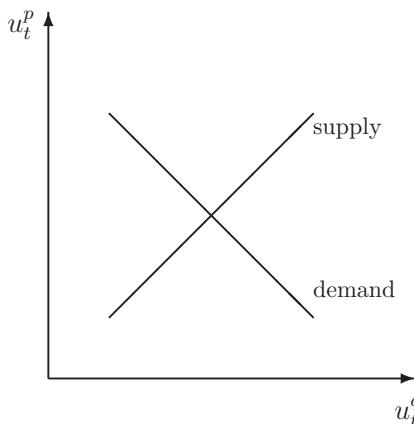


Figure 13.3. Standard demand and supply model.

of the form

$$\begin{bmatrix} + & + \\ 0 & 0 \end{bmatrix},$$

would be permitted under weak inequalities, but clearly is inadmissible because the two shocks would be observationally equivalent and B_0^{-1} would be rank deficient.

In contrast, it is possible in principle to achieve identification by judiciously combining weak and strict inequalities or by combining zero restrictions and strict inequalities. For example, the matrices

$$\begin{bmatrix} + & + \\ 0 & + \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} + & + \\ - & 0 \end{bmatrix},$$

would be admissible solutions for B_0^{-1} under alternative assumptions about the slopes of the demand and supply curves. The first of these matrices would be consistent with the pattern of demand and supply curves shown in Figure 13.4. The second matrix would be consistent with the pattern shown in Figure 13.5. Structural models that combine zero and sign restrictions will be discussed in more detail in Section 13.9.

It is sometimes argued that sign-identified models are more general than recursively identified models. Note, however, that the sign-identified model (13.1.2) does not nest the recursively identified model (13.1.1). While the sign-identified model relaxes the exclusion restriction in the recursively identified model, it does so at the cost of imposing sign restrictions on other parameters that were previously unrestricted. Thus, what we gain in generality in one dimension, we lose in the other dimensions. This means that the recursive

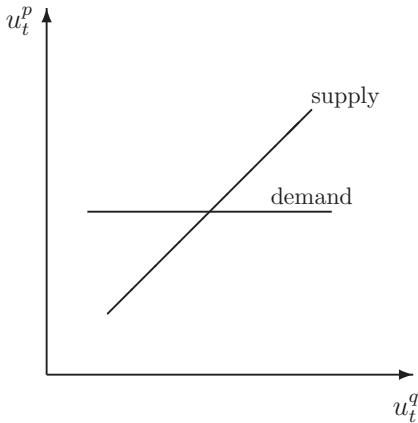


Figure 13.4. The case of a horizontal demand curve.

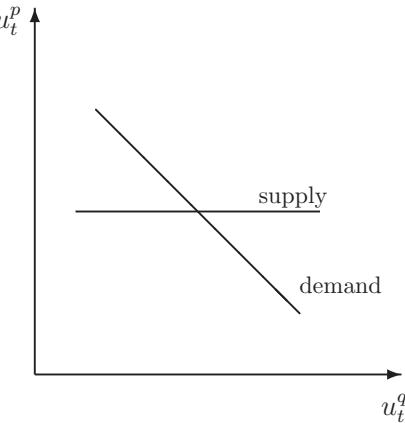


Figure 13.5. The case of a horizontal supply curve.

model is not nested in the sign-identified model. Hence, it is not possible to validate or invalidate the implications of recursively identified VAR models based on sign-identified models. Rather, these approaches are alternatives.

For now we focus on models identified by strict sign restrictions only. Although identification may be achieved by imposing a combination of exclusion and strict inequality restrictions, as discussed in Section 13.9, it is not possible to estimate sign-identified structural VAR models subject to weak inequalities. As shown later in this chapter, standard solution algorithms for sign-identified models are designed for strict inequalities only. All these algorithms have the property that if one were to impose a weak inequality on the response of the price to one shock, while imposing a strict inequality on the price response to the other shock, for example, the limiting case of zero is an event with probability measure zero.

One of the central questions is how to obtain numerical estimates of the structural model when only qualitative information is available about the model structure. It is useful to begin with the case of VAR models in which sign restrictions are imposed only on the impact responses of the observables to structural shocks. These responses correspond to the elements of the structural impact multiplier matrix, B_0^{-1} . Sign restrictions on the elements of B_0^{-1} are referred to as static sign restrictions.

13.2 How to Impose Static Sign Restrictions

Consider a structural vector autoregressive model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t.$$

The variance-covariance matrix of the structural error term w_t is normalized such that

$$\mathbb{E}(w_t w_t') \equiv \Sigma_w = I_K.$$

Let $u_t = P\eta_t$, where u_t is the reduced-form VAR innovation and P is the lower-triangular Cholesky decomposition of Σ_u .¹ By construction, the shocks η_t are mutually uncorrelated and have unit variance. There is, of course, no reason for these shocks to correspond to economically interpretable structural shocks such as the demand and supply shocks in the bivariate model (13.1.2) above, but one can search for candidate solutions w_t^* for the unknown structural shocks w_t by constructing a large number of combinations of the shocks η_t of the form

$$w_t^* = Q'\eta_t,$$

where Q' is a square orthogonal matrix such that $Q'Q = QQ' = I_K$ and $u_t = PQQ'\eta_t = PQw_t^*$. Hence, each candidate solution w_t^* consists of uncorrelated shocks with unit variance. Whether any of these candidate solutions w_t^* is an admissible solution for the unknown structural shock w_t , given the vector of reduced-form parameters, depends on whether the implied structural impact multiplier matrix, PQ , satisfies the maintained sign restrictions on B_0^{-1} . We retain solutions that satisfy these sign restrictions and discard the remaining solutions. Repeating this procedure allows us to characterize the set of all structural models that are consistent with the maintained sign restrictions and the reduced-form parameters. More generally, knowledge of PQ allows the construction of all implied structural impulse response coefficients of interest from the estimates of the reduced-form slope parameters.

The ability to generate large numbers of candidate matrices Q from the set of all orthogonal matrices \mathcal{O} thus is essential for the construction of sign-identified VAR models. In the following we denote the set of $K \times K$ orthogonal matrices by $\mathcal{O}(K)$, i.e.,

$$\mathcal{O}(K) \equiv \{Q \mid QQ' = I_K\}.$$

If the dimension of the matrices is not important or evident, we sometimes simply use \mathcal{O} instead of $\mathcal{O}(K)$. There are two common approaches to constructing orthogonal matrices Q . One is based on Givens rotation matrices; the other is the Householder transformation approach.

¹ It should be noted that nothing hinges on P being the lower-triangular Cholesky decomposition. Any solution for P that satisfies $PP' = \Sigma_u$ will do as well.

13.2.1 Givens Rotation Matrices

Givens rotation matrices can be used to construct orthogonal matrices. In the bivariate model, Givens matrices have the form

$$Q(\phi) = \begin{bmatrix} \cos \phi & -\sin \phi \\ \sin \phi & \cos \phi \end{bmatrix},$$

where ϕ lies between 0 and 2π . Each choice of $\phi \in [0, 2\pi]$ implies an orthogonal matrix $Q(\phi)$ and

$$\mathcal{O}(2) = \{Q(\phi) \mid \phi \in [0, 2\pi]\}.$$

In practice, one defines a finite-dimensional grid over the possible values of ϕ (or, alternatively, draws ϕ from a uniform distribution on $[0, 2\pi]$), computes the implied $Q(\phi)$ and the corresponding structural impact multiplier matrices $PQ(\phi)$, and retains only those solutions that agree with the maintained sign restrictions. This allows one to characterize the space of admissible structural VAR models, with each structural model defined as a set of structural impulse responses, conditional on the reduced-form estimate.

In the trivariate model, one way of forming an orthogonal matrix Q is to generate the product

$$Q(\phi_1, \phi_2, \phi_3) = Q_{12}(\phi_1) \times Q_{13}(\phi_2) \times Q_{23}(\phi_3),$$

of the Givens rotation matrices

$$Q_{12} = \begin{bmatrix} \cos \phi_1 & -\sin \phi_1 & 0 \\ \sin \phi_1 & \cos \phi_1 & 0 \\ 0 & 0 & 1 \end{bmatrix}, \quad Q_{13} = \begin{bmatrix} \cos \phi_2 & 0 & -\sin \phi_2 \\ 0 & 1 & 0 \\ \sin \phi_2 & 0 & \cos \phi_2 \end{bmatrix},$$

$$Q_{23} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & \cos \phi_3 & -\sin \phi_3 \\ 0 & \sin \phi_3 & \cos \phi_3 \end{bmatrix},$$

where each ϕ_i , $i = 1, 2, 3$, lies between 0 and 2π . Given $\cos^2 \phi_i + \sin^2 \phi_i = 1$, it can be shown that $Q'_{12}Q_{12} = I_3$, $Q'_{13}Q_{13} = I_3$, and $Q'_{23}Q_{23} = I_3$. By construction, $Q(\phi_1, \phi_2, \phi_3)$ is an orthogonal matrix.

Canova and De Nicolo (2002) suggest defining a finite-dimensional grid of values for each ϕ_i between 0 and 2π , computing all implied $Q(\phi_1, \phi_2, \phi_3)$, and retaining only those solutions that yield a structural impact multiplier matrix $PQ(\phi_1, \phi_2, \phi_3)$ that agrees with the maintained sign restrictions.

Generalizations to $K > 3$ are possible, but it is clear that this algorithm becomes computationally burdensome for large-dimensional VAR models and may be computationally infeasible for large K , which explains why it is rarely used in practice.

13.2.2 The Householder Transformation

An alternative and more common approach for computing suitable impact multiplier matrices, proposed by Rubio-Ramírez, Waggoner, and Zha (2010), is an algorithm involving the QR decomposition (or QR factorization). Recall that any real square matrix W can be decomposed as $W = QR$, where Q is an orthogonal matrix (i.e., its columns are orthogonal unit vectors such that $QQ' = I$) and R is an upper-triangular matrix. If W is invertible, then the factorization will be unique, provided the diagonal elements of R are restricted to be positive. The QR factorization of a matrix may be computed using the Householder transformation discussed in Stewart (1980).

The algorithm suggested by Rubio-Ramírez, Waggoner, and Zha (2010) covers the space $\mathcal{O}(K)$ of $K \times K$ orthogonal matrices Q by drawing each column of a $K \times K$ matrix W at random from a $\mathcal{N}(0, I_K)$ distribution and applying the QR factorization to each draw W . They show that if we choose Q by a QR decomposition of W with the diagonal of the upper triangular matrix R normalized to be positive, then this amounts to drawing Q from a uniform distribution over the space of orthogonal matrices $\mathcal{O}(K)$. Drawing the elements of the matrix W independently from a $\mathcal{N}(0, 1)$ distribution will result in a nonsingular matrix with probability 1. This fact ensures that invertibility holds when applying this algorithm. The mathematical underpinnings of this approach follow from Stewart (1980).

The algorithm can be used for generating a large number of candidate solutions for B_0^{-1} as PQ , where P denotes the lower-triangular Cholesky decomposition of Σ_u and Q is obtained from a random draw for W . The matrix Q in the literature is often referred to as the rotation matrix. Fry and Pagan (2011) observe that this method is equivalent to using Givens matrices. Its main advantage is that it is more computationally efficient for large K .

This leaves the question of how to compute the QR decomposition in practice. Commonly used software provides built-in functions for this purpose such as the *qr* function in MATLAB. Note, however, that this function does not ensure positive diagonal elements of R . Hence, that normalization has to be performed on the output of the *qr* function by reversing all signs of each row of R corresponding to a negative diagonal element and adjusting Q accordingly to ensure that $W = QR$ holds. Since we are not interested in R , we may equivalently just reverse all signs in the i^{th} column of Q if the i^{th} diagonal element of R is negative. Note also that the sign normalization should not be performed on Q because that may violate the premise of uniform sampling from $\mathcal{O}(K)$. For example, if one were to standardize the main diagonal elements of Q instead of R to be positive, one would still obtain a unique QR decomposition. However, one would miss all elements of $\mathcal{O}(K)$ with negative diagonal elements.

Having obtained many candidate solutions for B_0^{-1} , one retains only those solutions that yield a structural impact multiplier matrix that agrees with the

maintained sign restrictions. More generally, knowledge of admissible B_0^{-1} matrices also allows the construction of the implied set of structural impulse responses from the reduced-form slope parameters. Any draw for B_0^{-1} that does not satisfy all sign restrictions on the impulse responses must be rejected. The simplest approach is to discard this draw and to generate another draw. This approach is computationally inefficient if all responses to a given shock have the wrong sign. Rubio-Ramírez, Waggoner, and Zha (2010) observe that in that case, instead of generating a new draw, one can simply change the sign of the column of Q in question, resulting in a new orthogonal matrix that satisfies the sign restrictions. Notice that multiplying a column of an orthogonal matrix Q by -1 results in another orthogonal matrix.

A substantial further computational gain may be achieved by checking each column of PQ for the sign pattern associated with a given shock, exploiting the fact that in sign-identified VAR models the ordering of the variables does not determine which shock is contained in which column of the structural impact multiplier matrix (see, e.g., Baumeister and Hamilton 2015a). Suppose we are interested in a bivariate model with sign restriction matrix

$$\begin{bmatrix} + & + \\ - & + \end{bmatrix},$$

where the first column refers to the impact effects of a supply shock and the second column refers to a demand shock. If we obtain a draw

$$\begin{bmatrix} 0.41 & 0.01 \\ 0.30 & -0.88 \end{bmatrix}$$

for B_0^{-1} , we can interpret the first column as representing the impact effects of the demand shock and the second column as representing the impact effects of the supply shock rather than discarding this draw simply because the signs of the first column do not match the sign restriction matrix. This approach is equivalent to taking an alternative draw for Q that flips the columns of this matrix.

13.2.3 The Ouliaris-Pagan Approach

Yet another approach for computing suitable structural impact multiplier matrices was proposed by Ouliaris and Pagan (2016) who focus on the problem of imposing static sign restrictions on the elements of B_0 rather than on the elements of the structural impact multiplier matrix. Let B_0 have a unit diagonal and Σ_w be a diagonal matrix with unrestricted main diagonal. Then, if Σ_u and the $K(K - 1)/2$ elements of B_0 above the main diagonal are known, one can solve the system

$$\Sigma_u = B_0^{-1} \Sigma_w B_0^{-1'}$$
 (13.2.1)

for the elements of B_0 below the main diagonal and for the diagonal elements of Σ_w . Building on this observation, Ouliaris and Pagan propose to draw the above-diagonal elements of B_0 at random such that possible sign restrictions on B_0 are preserved. They then solve for the remaining elements of B_0 and the diagonal elements of Σ_w . They retain the resulting candidate solution for B_0 if all sign restrictions on B_0 are satisfied. Otherwise it is discarded. The procedure is repeated many times to ensure that the entire admissible parameter space is represented.

The procedure for drawing the above-diagonal elements of B_0 is as follows. For the ij^{th} element of B_0 , denoted $b_{ij,0}$, with $i < j$, a random variable ϕ is drawn from the uniform distribution

$$\begin{aligned} \mathcal{U}(-1, 1) &\quad \text{if } b_{ij,0} \text{ is unrestricted,} \\ \mathcal{U}(0, 1) &\quad \text{if } b_{ij,0} > 0, \\ \mathcal{U}(-1, 0) &\quad \text{if } b_{ij,0} < 0. \end{aligned}$$

Then $b_{ij,0}$ is set to $\phi/(1 - |\phi|)$. Independent draws are used for all elements $b_{ij,0}$ with $i < j$. Note that this algorithm, rather than sampling from the space of all possible orthogonal matrices Q , samples from the set of possible values for the elements of B_0 , treating these elements as independent. By construction, $b_{ij,0}$ has infinite support, but the algorithm assigns more probability mass to values of $b_{ij,0}$ close to zero.

Given the above-diagonal elements of B_0 and an estimate of Σ_u , the below-diagonal elements of B_0 and the diagonal elements of Σ_w can be estimated by solving the system (13.2.1) with a nonlinear equation solver as discussed in Chapter 10. An alternative possibility is to use IV estimation for estimating the values of the below-diagonal elements. Given the recursive structure of the estimation problem, we can use LS estimation of the first equation,

$$y_{1t} + b_{12,0}y_{2t} + \cdots + b_{1K,0}y_{Kt} = \sum_{k=1}^K \sum_{i=1}^p b_{1k,i}y_{k,t-i} + w_{1t},$$

where the left-hand-side parameters are fixed. The estimation error \widehat{w}_{1t} can then be used as an instrument for y_{1t} in estimating the second equation,

$$y_{2t} + b_{23,0}y_{3t} + \cdots + b_{2K,0}y_{Kt} = -b_{21,0}y_{1t} + \sum_{k=1}^K \sum_{i=1}^p b_{1k,i}y_{k,t-i} + w_{2t},$$

where again the left-hand-side parameters are given. More generally, the k^{th} equation,

$$\begin{aligned} y_{kt} + b_{k,k+1,0}y_{k+1,t} + \cdots + b_{kK,0}y_{Kt} \\ = -b_{k1,0}y_{1t} - \cdots - b_{k,k-1,0}y_{k-1,t} + \sum_{k=1}^K \sum_{i=1}^p b_{1k,i}y_{k,t-i} + w_{kt}, \end{aligned}$$

can be estimated by using the residuals from the previous equations, \hat{w}_{jt} , as the instrument for y_{jt} , $j = 1, \dots, k - 1$. If the estimates obtained in this way satisfy the sign restrictions, the implied estimate of B_0 is retained. Otherwise it is discarded.

As an example, consider a 3-dimensional VAR(1) model where no inequality restrictions are imposed on the above-diagonal elements of

$$B_0 = \begin{bmatrix} 1 & b_{12,0} & b_{13,0} \\ b_{21,0} & 1 & b_{23,0} \\ b_{31,0} & b_{32,0} & 1 \end{bmatrix}.$$

The first step of the procedure is to estimate the reduced form, $y_t = A_1 y_{t-1} + u_t$, by LS. We obtain an estimate $\hat{\Sigma}_u$ of the reduced-form error covariance matrix. We then draw three independent numbers ϕ_1, ϕ_2, ϕ_3 from a uniform distribution $\mathcal{U}(-1, 1)$ and fix the above-diagonal elements of B_0 at $b_{12,0}^* = \phi_1/(1 - |\phi_1|)$, $b_{13,0}^* = \phi_2/(1 - |\phi_2|)$, and $b_{23,0}^* = \phi_3/(1 - |\phi_3|)$, respectively. Finally, we estimate $b_{21,0}$, $b_{31,0}$, and $b_{32,0}$ and the diagonal elements of Σ_w by solving (13.2.1) for the unknown parameters with Σ_u replaced by $\hat{\Sigma}_u$.

Alternatively, estimates of the unknown parameters may be obtained by LS estimation of

$$y_{1t} + b_{12,0}^* y_{2t} + b_{13,0}^* y_{3t} = b_{11,1} y_{1,t-1} + b_{12,1} y_{2,t-1} + b_{13,1} y_{3,t-1} + w_{1t}.$$

The residuals of this regression then are used as the instrument for y_{1t} in the IV estimation of

$$y_{2t} + b_{23,0}^* y_{3t} = -b_{21,0} y_{1t} + b_{21,1} y_{1,t-1} + b_{22,1} y_{2,t-1} + b_{23,1} y_{3,t-1} + w_{2t}.$$

Finally,

$$y_{3t} = -b_{31,0} y_{1t} - b_{32,0} y_{2t} + b_{31,1} y_{1,t-1} + b_{32,1} y_{2,t-1} + b_{33,1} y_{3,t-1} + w_{3t}$$

is estimated by IV using the residuals from the previous two equations as instruments for y_{1t} and y_{2t} .

This method may also be adapted to allow the B_0 matrix to be restricted in other ways, provided that enough model parameters are fixed, so the remaining parameters can be solved for or estimated. For example, if the main diagonal of B_0 is unrestricted, Σ_w is the identity matrix, and Σ_u is known, the system $\Sigma_u = B_0^{-1} B_0^{-1'}$ can be solved for the elements on and below the main diagonal, as long as the above-diagonal elements of B_0 are fixed by assigning random draws to them.

13.3 Partially Identified VAR Models

Implicitly, the discussion thus far assumed that the model is fully identified in that all structural shocks are individually identified. This is not always the case.

A common situation in applied work is that the researcher only has knowledge of the signs of the responses to some shocks. A situation in which the number of identified shocks is less than K is known as a partially identified VAR model (see, e.g., Rubio-Ramírez, Waggoner, and Zha 2010; Inoue and Kilian 2013). Most common are situations in which only a single structural shock is of interest.²

For example, consider the following partially identified bivariate model. Suppose that structural shock w_{1t} is known to raise both observables on impact, whereas we have no a priori knowledge of the sign of the responses to structural shock w_{2t} :

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{bmatrix} + & ? \\ + & ? \end{bmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix}.$$

It may seem that in this case it would be enough to check the signs in the first column of the structural impact multiplier matrix. This is not the case if we insist that the sign pattern of the first shock must be distinct from that of the other shock. In that case, we still need to verify the signs of the responses to w_{2t} because a realization of PQ may give rise to a candidate solution

$$\begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{bmatrix} + & + \\ + & + \end{bmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix}.$$

Given that we can interchange the two candidate structural shocks in this case, it is not clear which of the two shocks we should focus on. Since the two shocks are orthogonal, they are distinct in the statistical sense, but they are not distinct in the economic sense. Simply choosing the first structural shock to be the demand shock would be arbitrary. Moreover, we know that this particular candidate solution is inconsistent with any reasonable economic model of this market because it implies the existence of two demand shocks but no supply shock. Hence, we need to verify that the signs in the second column are the complement of the signs in the first column, namely,

$$\begin{bmatrix} + & + \\ + & - \end{bmatrix} \quad \text{or} \quad \begin{bmatrix} + & - \\ + & + \end{bmatrix}.$$

If they are not, this draw for PQ must be discarded.

Although in this simple example the requirement that the identified shock be distinguished by a unique sign pattern effectively implies a fully identified

² Set identification arises whenever inequality constraints are imposed. In microeconomics it is common to refer to set-identified models interchangeably as partially identified models. In the present context, partial identification has a different meaning, and it is important to keep in mind that a model may be set identified without being partially identified and partially identified without being set identified. Nevertheless, there are some VAR studies that ignore this distinction and refer to set identification as partial identification.

model, this is not the case in higher-dimensional partially identified models. In general, all we are imposing is that the sign pattern of each of the unidentified shocks is different from that of the identified shocks. We are not imposing that any of the unidentified shocks must have a specific sign pattern.

It should be noted that there is no consensus in the literature on whether this additional requirement should be imposed in estimating partially identified sign-identified models. As a result, some studies focus only on the responses to the shock of interest without inspecting the responses to other shocks in the same model. In other words, they consider only one column of the impact multiplier matrix at a time, looking for solutions that satisfy the prespecified sign pattern while ignoring the other columns. All columns satisfying the prespecified sign pattern are considered admissible solutions, and the possibility that other shocks in the same structural model may have the same sign pattern is ignored. Further discussion of this issue can be found in Uhlig (2005), Fry and Pagan (2011), and Canova and Paustian (2011).

13.4 Beyond Static Sign Restrictions

The basic algorithm for characterizing the set of admissible models can be extended to allow for additional restrictions on the structural impulse responses. Such additional restrictions are often required for the estimates of sign-identified models to be economically meaningful.

13.4.1 Dynamic Sign Restrictions

It is straightforward to extend the algorithms above to allow for additional sign restrictions on the structural impulse responses beyond the impact period. One drawback of this approach is that there is less consensus in economic theory about the signs of structural impulse responses at longer horizons (see, e.g., Canova and Paustian 2011). It has also been noted that imposing dynamic sign restrictions may be redundant in some cases and hence unhelpful (see Fry and Pagan 2011). Nevertheless, such restrictions can be useful in restricting further the space of admissible models. For example, we may be willing to agree that a monetary policy tightening is bound to reduce real GDP after half a year, even if the sign of the short-run response and the sign of the longer-run response of real GDP is debatable (see Inoue and Kilian 2013).

13.4.2 Elasticity Bounds

A common approach among applied researchers has been to favor models that are agnostic in the sense of only imposing minimal identifying assumptions.

Being agnostic in the context of sign-identified models involves the risk of allowing problematic structural models to be retained in the admissible set. This point was first illustrated in Kilian and Murphy's (2012) analysis of the global oil market. Their model may be viewed as the analogue of the recursively identified oil market model of Kilian (2009) in Chapter 8.

The model includes monthly data for the growth of global crude oil production ($\Delta prod_t$), a measure of global real activity expressed as a business cycle index (rea_t), and the log of the real price of oil ($rpoil_t$). The three structural shocks are a shock to the flow of crude oil coming out of the ground ($w_t^{\text{flow supply}}$), a shock to the flow demand for oil associated with unexpected fluctuations in the global business cycle ($w_t^{\text{flow demand}}$), and a shock to the demand for oil not associated with the global business cycle ($w_t^{\text{other demand}}$). The latter demand shock is designed to capture, for example, shifts in the precautionary demand for crude oil that shift the demand for stocks of oil. A conventional analysis of this market would start with restrictions on the signs of the structural impact multiplier matrix:

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} - & + & + \\ - & + & - \\ + & + & + \end{bmatrix} \begin{pmatrix} w_t^{\text{flow supply}} \\ w_t^{\text{flow demand}} \\ w_t^{\text{other demand}} \end{pmatrix}.$$

All shocks have been normalized to imply an increase in the real price of oil on impact. For example, a negative flow supply shock can be thought of as a shift of the short-run oil supply curve to the left, causing a reduction in oil production and an increase in the real price of oil. Such a shock would also lower global real economic activity. In contrast, a positive flow demand shock would shift the demand curve for oil to the right, causing both oil production and the real price of oil to increase while raising global real activity. Finally, we can think of the other demand shock as a precautionary demand shock that raises the price of oil by shifting the demand for oil stocks. To accommodate the accumulation of stocks, oil production has to increase and oil consumption has to fall, which implies a fall in global real activity.

One of the central questions in the literature on the global oil market is how large the effects of flow supply shocks are on the real price of oil. It can be shown that the set of admissible models consistent with the minimal assumptions outlined thus far includes many models in which oil supply shocks cause much larger oil price responses than in the recursively identified model of Kilian (2009).

What is not immediately obvious, however, is that these models all imply a large impact price elasticity of oil supply. The latter elasticity is defined as the ratio of the impact response of global oil production triggered by an exogenous demand shock relative to the impact response of the real price of oil triggered

by the same shock.³ Large elasticity estimates not only violate the conventional wisdom among oil experts but also are inconsistent with formal and informal empirical evidence from other sources showing that this elasticity is close to zero. They are also inconsistent with economic theory (see, e.g., Anderson, Kellogg, and Salant 2016).

Kilian and Murphy (2012) show that restricting the impact price elasticity of oil supply not to exceed a reasonable magnitude suffices to eliminate the structural models in the admissible set that imply large responses of the real price of oil to flow supply shocks and produces results much more in line with those in Kilian (2009). Although this additional restriction does not directly restrict the response of the real price of oil to a supply shock, it suffices to estimate fairly precisely the set of responses of the real price of oil to oil supply shocks. In contrast, there is considerable uncertainty about the magnitude and pattern of the responses of the real price of oil to the two demand shocks in this model.

This example shows how a failure to impose relevant identifying restrictions may give credence to models that upon reflection are unrealistic. Kilian and Murphy (2012) conclude that the burden of proof is on the researcher to show that all available identifying information has been imposed on sign-identified models. There is no comfort in having remained agnostic and no excuse for having failed to examine other possible sources of identifying information.

In practice, elasticity bounds may be imposed by disregarding the candidate solutions for all structural models with higher elasticity values than permitted. The approach of bounding price elasticities was subsequently refined in Kilian and Murphy (2014) who showed that the impact price elasticity of oil demand as well may be bounded by independent estimates of the long-run price elasticity of demand from cross-sectional studies. The identifying assumption is that the long-run price elasticity of oil demand at least weakly exceeds the short-run price elasticity. Although this example focuses on the oil market, similar restrictions may also be used in many other market models. Even if a specific bound may be difficult to justify, sensitivity analysis based on alternative elasticity bounds may reveal how fragile the empirical results are to imposing extra information.

Although the example of market models is the most natural context for the use of elasticity bounds, such bounds have also been discussed in other contexts. For example, Caldara and Kamps (2017) investigate the implications of bounds on output elasticities of fiscal variables for the fiscal multiplier in sign-identified VAR models.

³ This definition corresponds to the textbook definition of the price elasticity of oil supply. In applied regression analysis it is common to regress the log of one variable on the log of another variable and to refer to the regression coefficient as an elasticity. The latter elasticity concept in general is not related to the microeconomic concept of the price elasticity of demand or the price elasticity of supply.

13.4.3 Shape Restrictions

A natural extension of the idea of sign restrictions is the use of shape restrictions on structural impulse response functions. Such restrictions may involve monotonicity constraints on the evolution of the response function or may involve imposing a hump shape on the response function. A hump shape may be natural in modeling the effects of demand shocks on real output, for example (see, e.g., Blanchard and Quah 1989; Inoue and Kilian 2016). Shape restrictions may also arise in modeling the delayed overshooting in asset prices in response to monetary policy shocks, as discussed in Eichenbaum and Evans (2005) and Scholl and Uhlig (2008). These restrictions can be written as sign restrictions on the change in the impulse response coefficients across the horizon and hence can be accommodated by standard solution algorithms.

13.5 Can Sign Restrictions Be Verified?

It is uncontroversial that imposing incorrect identifying assumptions could result in the set of admissible models being empty. Perhaps motivated by this observation, it is sometimes suggested that only a small fraction of the random draws of PQ being admissible (in the sense of satisfying the maintained sign restrictions) means that the identifying assumptions are suspect. For example, Fry and Pagan (2011) suggest that in this case the data are incompatible with the maintained sign restrictions.

This view is misleading for several reasons. First, suppose for expository purposes that the reduced-form parameters are known. Then the fraction of admissible models only depends on Q , but Q does not depend on the data, so the fraction of admissible models does not constitute empirical evidence. In fact, all rotations by construction imply models that fit the data equally well. Second, reporting the fraction of admissible models is meaningless, if we do not know the computational efficiency of the algorithm used. Merely by changing algorithms, this fraction may change by a factor of 10 or 20. Third, it is evident upon reflection that, for a given algorithm, the fraction of admissible models is simply an indication of how informative the identifying restrictions are. This last point is best made by example.

To build intuition, first consider an extreme example, in which the data are generated from a recursive model. In that case, no one would expect random draws from the QR algorithm conditional on the VAR parameters to generate realizations of PQ that preserve this recursive structure. Although the probability of encountering a recursive model estimate is zero, the correct model is nevertheless recursive. The same intuition still applies if we relax the assumptions somewhat. Consider instead a set of data generating processes with some structural impulse responses not equal to zero, but ranging in value between 0 and some small $\epsilon > 0$ in population. If we impose this constraint in our search for

admissible models, the number of draws generated from the QR decomposition that lie within this region will be small relative to the total number of draws by construction. Again, this does not mean that the identifying assumption is suspect. Exactly the same phenomenon arises if — rather than restricting some structural impulse response to lie within a small region — one imposes multiple inequality restrictions on the impulse responses that effectively restrict the admissible range of the structural impulse responses. Indeed, a low fraction of admissible models in this case could actually be an indication of having imposed all relevant economic structure on the model. Conversely, a high fraction of admissible models is bound to arise in insufficiently identified models. Thus, the fraction of admissible structural models is not a meaningful measure of the quality of the identifying restrictions of a sign-identified VAR model.

It is, of course, correct that by imposing a certain sign on the structural impulse responses, we rule out all models that imply a different sign. As in all structural VAR models, the estimates of sign-identified models are conditional on the chosen identifying assumptions. These assumptions are not testable within this VAR framework. The only purpose of estimating the sign-identified model therefore is to quantify the magnitude of the response of interest conditional on the assumed signs.

It has been suggested that one way of evaluating the plausibility of sign restrictions would be to establish that the sign-identified model can recover the correct sign of the response of interest, when other response functions in the VAR model are constrained, but not the response function of interest. Indeed, this was the explicit motivation for the agnostic identification procedure proposed by Uhlig (2005). Paustian (2007) addresses this question in the context of sign-identified VAR models applied to data generated from DSGE models. He provides simulation evidence that for this approach to work, the variance of the structural shock that triggers the response of interest must be much higher in population than seems empirically plausible.

This result is not surprising because sign restrictions are comparatively weak, and hence models that do not impose all available identifying information tend to be uninformative. This point has been illustrated by Inoue and Kilian (2013) in the context of Uhlig's agnostic identification procedure. Their analysis confirms that there is no information about the response of real GDP to monetary policy shocks in this model without further identifying restrictions on the response function of interest. This result, of course, does not mean that sign-identified models with richer identification structures are not a useful tool in quantifying dynamic economic relationships. A similar point has been made in Canova and Paustian (2011) who conclude that — leaving aside pathological examples — even when the variance of the shock of interest is low, the sign-identified model usually does not make systematic mistakes, provided the estimated model has a sufficiently rich shock structure and employs enough identifying restrictions.

13.6 Estimation and Inference in Sign-Identified VAR Models

A fundamental problem in interpreting VAR models identified based on sign restrictions is that there is not a unique point estimate of the structural impulse response functions. Unlike conventional structural VAR models based on short-run restrictions, sign-identified VAR models are only set identified. This problem arises because sign restrictions represent inequality restrictions. The cost of remaining agnostic about the precise values of the structural model parameters is that the data are potentially consistent with a wide range of structural models that are all admissible in that they satisfy the identifying restrictions. Without further assumptions all these structural models are equally likely.

A typical outcome in practice is that the structural impulse responses implied by the admissible structural models will disagree on the substantive economic questions of interest. One early approach to this problem, exemplified by Faust (1998), has been to focus on the admissible model that is most favorable to the hypothesis of interest. This allows us to establish the extent to which this hypothesis could potentially explain the data. It may also help us rule out a hypothesized explanation if none of the admissible models supports this hypothesis. The problem is that this approach is not informative about whether any one of the admissible models is a more likely explanation of the data than some other model. There are examples in which the admissible structural models are sufficiently similar to allow unambiguous answers to the question of economic interest (see, e.g., Kilian and Murphy 2012). Typically, however, the set of admissible models will be equally consistent with competing economic hypotheses. This problem is compounded once we allow for estimation uncertainty in the reduced-form parameters. Moreover, in the latter case, it has to be kept in mind that confidence sets or credible sets in sign-identified VAR models by construction reflect both uncertainty about the identification of the model and estimation uncertainty.

There have been both frequentist and Bayesian approaches to summarizing estimates of the admissible set of sign-identified structural VAR models. These approaches differ from the methods discussed in Chapter 12. Standard asymptotic and bootstrap methods of inference on structural impulse responses, in particular, are invalid when working with inequality restrictions. This result follows from the absence of a consistent point estimator of the structural impulse response. Section 13.6.1 reviews the construction of alternative frequentist confidence sets for structural impulse responses in sign-identified VAR models. Section 13.6.2 discusses how to generate draws from the posterior distribution of the structural impulse responses. Section 13.6.3 critically examines the proposal of summarizing the posterior with the help of so-called median response functions and discusses the alternative approach of reporting the responses of the most likely structural model in the identified set. It also

discusses the construction of credible sets. Yet another approach is the use of penalty functions, as reviewed in Section 13.6.4. Finally, Section 13.6.5 outlines how historical information may be used to narrow the set of admissible models both in a Bayesian and in a frequentist setting.

13.6.1 Frequentist Approaches

The procedures for characterizing the set of admissible models outlined in Sections 13.2 and 13.4 condition on a given reduced-form VAR model and do not account for estimation uncertainty. In practice, we need to account for both identification uncertainty and estimation uncertainty in sign-identified models. One method of constructing classical confidence intervals for structural impulse responses in stationary sign-identified VAR models with iid innovations has recently been developed by Moon, Schorfheide, and Granziera (2013) under the simplifying assumption that the reduced-form impulse responses have an asymptotic normal distribution. They view the estimation of sign-identified VAR models as an estimation problem subject to moment-inequality restrictions. The presence of inequality constraints means that the structural impulse responses are only set identified and that the large-sample numerical equivalence between Bayesian credible sets and frequentist confidence sets breaks down.

Let ϕ denote the vector of orthogonalized reduced-form impulse responses, obtained by post-multiplying the reduced-form VAR responses by the lower-triangular Cholesky decomposition of Σ_u , θ the vector of structural impulse responses, and q the vector of nuisance parameters corresponding to a suitable parameterization of Q . Sign restrictions generate an identified set for θ and q , denoted by $F^{\theta,q}(\phi)$. Conditional on q and ϕ , the vector θ is point identified. Let $\sqrt{T}(\widehat{\phi} - \phi) \xrightarrow{d} \mathcal{N}(0, \Lambda)$ for all $\phi \in \mathcal{P}$, where \mathcal{P} is the space of the reduced-form parameters ϕ and the limiting covariance matrix Λ is of full rank.

Moon et al. propose two approaches to obtaining a $(1 - \gamma)100\%$ confidence set CS^θ

$$\inf_{\phi \in \mathcal{P}} \inf_{\theta \in F^\theta(\phi)} \mathbb{P}_\phi \{ \theta | \theta \in \text{CS}^\theta(\widehat{\phi}) \} \geq 1 - \gamma,$$

where \mathbb{P}_ϕ denotes the probability measure for a given reduced-form draw of ϕ and $F^\theta(\phi)$ is the set of structural impulse responses obtained for a given ϕ . Both approaches involve the marginalization of the joint confidence set for (θ, q) . One approach is to project the joint confidence set for $(\theta, q) \in F^{\theta,q}(\phi)$ onto the θ ordinate. The other approach involves first constructing confidence intervals for the set-identified nuisance parameters in q and then taking the union of standard confidence sets for θ that are generated conditional on all q in the first-stage confidence set. Moon et al. use the Bonferroni inequality to control the coverage probability of the resulting confidence set for θ .

In practice, the distribution of the reduced-form parameters is approximated by standard bootstrap methods. A detailed description of the algorithms is provided in Moon, Schorfheide, and Granziera (2013). The construction of these nonstandard confidence intervals is computationally costly and tends to be infeasible in fully identified models. Moreover, their coverage probabilities tend to be conservative.

More recently, Gafarov, Meier, and Montiel Olea (2015a) proposed a computationally less demanding alternative delta method approach for partially identified set-identified structural VAR models that tends to produce tighter intervals than the algorithm of Moon et al. A more general method that also accommodates fully identified models is discussed in Gafarov, Meier, and Montiel Olea (2015b). The latter method is conservative in that its coverage tends to exceed the nominal confidence level, but the coverage accuracy may be calibrated, resulting in tighter confidence bounds. Unlike the method in Moon, Schorfheide, and Granziera (2013), Gafarov, Meier, and Montiel Olea (2015a, 2015b) only require the asymptotic normality of the reduced-form parameters and are able to achieve uniformly accurate coverage.

Even with these adjustments, frequentist confidence sets for sign-identified models tend to be wide and not informative about the shape of the impulse response functions. A given confidence band tends to be consistent with a wide range of different impulse response function patterns. This fact tends to make it difficult to interpret the results from an economic point of view. An open question is whether this problem may be overcome in larger models with more identifying restrictions.

Finally, Kitagawa, Montiel Olea, and Payne (2015) extend the analysis to the related problem of constructing confidence sets for the maximum of the h -period ahead forecast error variance decomposition of a variable with respect to a given structural shock (see Chapter 4). The key innovation of this study is that it focuses on possibly nondifferentiable functions of a parameter vector θ such as $|\theta|$ or $\max(0, \theta)$, for which standard delta method or bootstrap inference, as employed in earlier studies, fails. The maximal contribution of monetary policy shocks to the variability of output growth is one such example. Kitagawa et al. provide precise conditions under which a confidence set for such functions may nevertheless be obtained by calibrating the highest-posterior density Bayesian credible set to achieve the desired frequentist coverage accuracy.

Notwithstanding these theoretical advances, there are few empirical applications to date of frequentist confidence sets for sign-identified structural impulse responses. Moreover, it is important to keep in mind that these confidence sets not only tend to be wide, but that they tend to be uninformative about the shape of the impulse response functions in that a given confidence band is typically consistent with a wide range of different impulse response function patterns. This fact tends to make it difficult to interpret the results from an economic point of view.

13.6.2 Bayesian Approaches

The most common approach in the literature on sign-identified VAR models has been to rely on Bayesian methods of inference (see Chapter 5). For example, under the assumption of a conventional Gaussian-inverse Wishart prior on the reduced-form parameters and an independent uniform prior on the rotation matrices, one can construct the posterior distribution of the structural impulse responses by simulating posterior draws from the reduced-form posterior, applying the QR algorithm to each reduced-form posterior draw, and discarding solutions that do not satisfy the sign restrictions.

Generating the Posterior in Fully Identified Models. Consider the K -variate reduced-form VAR(p) model:

$$y_t = v + A_1 y_{t-1} + A_2 y_{t-2} + \cdots + A_p y_{t-p} + u_t, \quad (13.6.1)$$

for $t = 1, \dots, T$, where $u_t \stackrel{iid}{\sim} \mathcal{N}(0_{K \times 1}, \Sigma_u)$ and Σ_u is positive definite. Define $A = [v, A_1, \dots, A_p]$.

Specify a Gaussian-inverse Wishart prior distribution for the reduced-form VAR parameters of the form

$$\text{vec}(A) | \Sigma_u \sim \mathcal{N}(\text{vec}(A^*), V_{\text{vec}(A)}) = V \otimes \Sigma_u$$

and

$$\Sigma_u \sim \mathcal{IW}_K(S_*, n),$$

where $\text{vec}(A^*)$, V , S_* , and n are prior parameters specified by the analyst (see Section 5.2.4). Then the posterior is given by

$$\text{vec}(A) | \Sigma_u, \mathbf{y} \sim \mathcal{N}(\text{vec}(\bar{A}), \bar{\Sigma}_{\text{vec}(A)}), \quad \Sigma_u | \mathbf{y} \sim \mathcal{IW}_K(S, \tau), \quad (13.6.2)$$

where $\mathbf{y} \equiv \text{vec}(Y)$, $Y \equiv [y_1, \dots, y_T]$ denotes the data,

$$\begin{aligned} \bar{A} &= (A^*V^{-1} + YZ')(V^{-1} + ZZ')^{-1} \\ \bar{\Sigma}_{\text{vec}(A)} &= (V^{-1} + ZZ')^{-1} \otimes \Sigma_u, \\ S &= T\tilde{\Sigma}_u + S_* + \hat{A}ZZ'\hat{A}' + A^*V^{-1}A^{**} - \bar{A}(V^{-1} + ZZ')\bar{A}', \end{aligned}$$

and $\tau = T + n$. Here A^* and \bar{A} are $K \times (Kp + 1)$ matrices, $\hat{A} = YZ'(ZZ')^{-1}$, and $\tilde{\Sigma}_u = (Y - \hat{A}Z)(Y - \hat{A}Z)'/T$. Moreover, $Z \equiv [Z_0, \dots, Z_{T-1}]$ with $Z_{t-1} \equiv (1, y'_{t-1}, \dots, y'_{t-p})'$.

Simulating the posterior of the structural impulse responses requires draws for $\text{vec}(A)$ and for B_0^{-1} . Let A^{*r} denote the r^{th} posterior draw of A and Σ_u^{*r} the r^{th} posterior draw for Σ_u . Then $\tilde{B}_0^{-1} = P^{*r}Q$, where P^{*r} is the lower-triangular Cholesky decomposition of Σ_u^{*r} such that $P^{*r}P^{*r'} = \Sigma_u^{*r}$, and Q is an orthogonal matrix. \tilde{B}_0^{-1} is a potential solution for the unknown structural impact multiplier matrix B_0^{-1} that satisfies $\tilde{B}_0^{-1}\tilde{B}_0^{-1'} = P^{*r}QQ'P^{*r'} = P^{*r}P^{*r'} = \Sigma_u^{*r}$.

Following Uhlig (2005), the prior distribution for the matrix Q is postulated to be uniform on the space of orthogonal matrices $\mathcal{O}(K)$, which can be drawn via the Householder transformation as described in Section 13.2.2, allowing us to simulate the set of potential solutions for \tilde{B}_0^{-1} , given Σ_u^{*r} and A^{*r} .

As discussed in Chapter 4, there is a known nonlinear function that allows us to construct the structural impulse responses associated with every potential solution $(A^{*r}, \tilde{B}_0^{-1})$ for the structural model. Candidate solutions that imply structural impulse responses that do not satisfy all sign restrictions are discarded. The other posterior draws for the structural impulse responses are retained and used to approximate the posterior distribution of the structural impulse responses. Unlike in conventional VAR models, this distribution reflects both estimation uncertainty and identification uncertainty.

In practice we proceed as follows:

- Step 1.** Take a random draw, (A^{*r}, Σ_u^{*r}) , from the posterior of the reduced-form VAR parameters and compute the lower-triangular Cholesky decomposition $P^{*r} = \text{chol}(\Sigma_u^{*r})$.
- Step 2.** For (A^{*r}, P^{*r}) , consider N random draws of the rotation matrix Q , and for each combination (A^{*r}, P^{*r}, Q) compute the set of implied structural impulse responses Θ^{*r} .
- Step 3.** If Θ^{*r} satisfies the sign restrictions, store the value of Θ^{*r} . Otherwise discard Θ^{*r} .
- Step 4.** Repeat steps 1, 2, and 3 M times.

In simulating the posterior distribution of the structural responses, care must be taken that the posterior is approximated using a sufficiently large number of reduced-form draws as well as a sufficiently large number of rotation draws for each posterior draw from the reduced-form. The tighter the identifying restrictions, the more draws will be required to approximate the posterior of the structural impulse responses, because fewer realizations will satisfy the sign restrictions. The precise number of draws required for an accurate approximation also depends on which algorithm is employed. In practice, it is recommended to verify that alternative seeds of the Gaussian random number generator generate similar sets of admissible models. If not, the number of draws must be increased.

Generating the Posterior in Partially Identified Models. A simplified algorithm for partially identified models was proposed by Uhlig (2005). Uhlig's objective was to identify the responses to a monetary policy shock without taking a stand on the identification of the remaining structural shocks. Uhlig does not solve his model by constructing draws for Q as discussed earlier. Rather, he observes that constructing the responses of the model variables to the monetary policy shock does not require knowledge of the entire structural

impact multiplier matrix B_0^{-1} , but only of the vector b representing the impact responses of his model variables to the monetary policy shock. The vector b is the first column of B_0^{-1} in this example.

Focusing on one structural shock only simplifies the analysis. By analogy to $B_0^{-1} = PQ$, let q denote a K -dimensional vector of unit length such that $b = Pq$. Then the sign-identified structural impulse responses can be expressed as $\theta_i = \Phi_i b$, where θ_i denotes the vector of responses at horizon i to the monetary policy shock. In practice, Uhlig draws a vector \tilde{q} from the $\mathcal{N}(0, I_K)$ distribution and normalizes the length of this vector to unity such that $q = \tilde{q}/\|\tilde{q}\|$, where $\|\tilde{q}\| = \sqrt{\sum_{k=1}^K \tilde{q}_k^2}$ denotes the length of the vector \tilde{q} . Given P and a candidate solution for q , one can construct the implied candidate solution for b . Only candidates for b that imply solutions for θ_i that satisfy the sign restrictions are retained.

This approach does not rule out the possibility that there are other orthogonal shocks with exactly the same sign pattern as the monetary policy shock, which may raise the question of whether the monetary policy shock is properly identified.

13.6.3 Evaluating the Posterior of the Structural Impulse Responses

Having obtained enough draws for the structural responses to approximate their posterior distribution, we can make probability statements about the structural impulse responses. Note that the posterior distribution incorporates both estimation uncertainty and identification uncertainty. The standard approach in the literature for many years has been to report the vector of pointwise posterior medians of the structural impulse responses (often referred to as the median response function) as a measure of the central tendency of the impulse response functions. In fact, many applied users treat these median response functions as though they were traditional point estimates from exactly identified models.

This approach suffers from two distinct shortcomings. One shortcoming is that the vector of pointwise posterior median responses will not correspond to the response function of any of the admissible models, unless the pointwise posterior medians of all impulse response coefficients in the VAR system correspond to the same structural model, which is highly unlikely a priori. The problem is not only that for different horizons h the pointwise posterior median responses coincide with responses from different admissible structural models. Similar problems may also arise when the order of the models differs for two response functions at the same horizon h . Thus, the median response function lacks a structural economic interpretation (see, e.g., Fry and Pagan 2011; Kilian and Murphy 2012; Inoue and Kilian 2013).

This point is illustrated in Figure 13.6. The figure focuses on the response of real GDP to an unanticipated monetary policy shock for a horizon of up

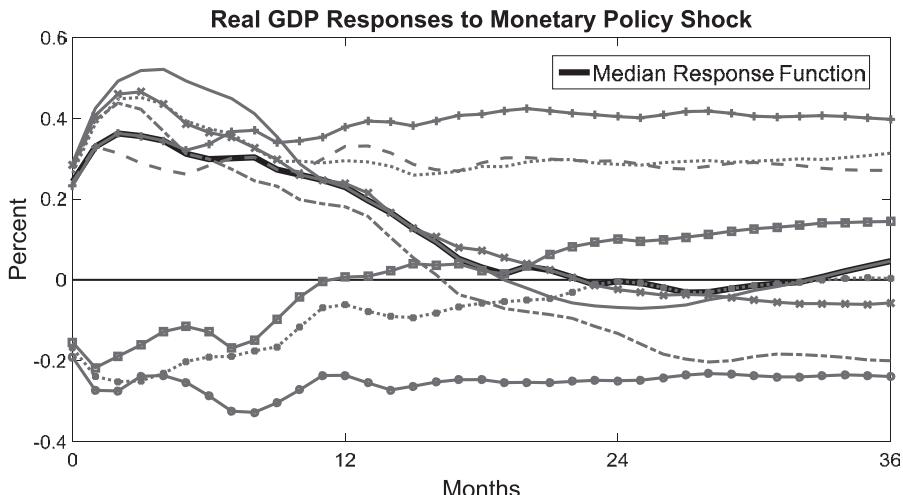


Figure 13.6. Randomly selected response functions from the sign-identified VAR model in Uhlig (2005).

to 36 months. This example was constructed by plotting a randomly chosen subset of nine admissible response functions for the sign-identified VAR(12) model of monetary policy proposed by Uhlig (2005). The model consists of monthly U.S. data for the log of interpolated real GDP, the log of the interpolated GDP deflator, the log of a commodity price index, total reserves, nonborrowed reserves, and the federal funds rate. The sample period is 1965m1–2003m12. The identifying restrictions are that an unexpected monetary policy contraction is associated with an increase in the federal funds rate, with price decreases, and with declines in nonborrowed reserves for some time following the policy shock.

Figure 13.6 demonstrates that, for different horizons, the pointwise posterior median responses of real GDP to a monetary policy shock coincide with responses of different admissible structural models. Specifically, the median response function for real GDP coincides with the response function of model 1 at horizons 23–32, with the response function of model 3 at horizons 19 and 20, with that of model 5 at horizons 12–18 and 33–36, with that of model 6 at horizons 0, 1, and 7–9, with that of model 7 at horizons 10, 11, 21, and 22, with that of model 8 at horizons 5 and 6, and with that of model 9 at horizons 2, 3, and 4. There is, in fact, no structural model in the admissible set that could replicate the response pattern implied by the posterior median response function, rendering this statistic economically meaningless.

The second shortcoming is that median response functions are not a useful statistical summary of the set of admissible impulse response functions. It is well known that the vector of medians is not the median of a vector valued random variable. In fact, the median of a vector valued random variable does not

exist, rendering the vector of pointwise medians inappropriate as a statistical measure of the central tendency of the impulse response functions (see, e.g., Chaudhuri 1996; Koltchinskii 1997; Liu, Parelis, and Singh 1999).⁴

This means that even if there were an admissible structural model for which all impulse response functions coincide with the corresponding median response functions, there would be no compelling reason to focus on this model in interpreting the evidence. In fact, it has been shown that posterior median response functions may be quite misleading about the most likely response dynamics in sign-identified models (see, e.g., Kilian and Murphy 2012; Inoue and Kilian 2013).

This criticism also extends to the proposal in Fry and Pagan (2011) designed to overcome the lack of structural interpretation of median response functions. Their proposal is to search for the admissible structural model with impulse response functions closest to the median response function. This proposal deals with the first shortcoming highlighted above, but not with the second. Because the median response function is not a well-defined statistical measure of central tendency, there is no compelling reason to focus on the structural model with responses closest to the pointwise posterior medians.

The same conceptual problem arises with the upper and lower quantiles constructed from the pointwise posterior distribution of the structural impulse responses. It is common to connect the upper quantiles for a given impulse response function to form an upper band, and similarly to connect the lower quantiles to form a lower band. These error bands, of course, fail to account for the dependence of structural impulse responses across horizons and across response functions, and hence are misleading (see, e.g., Sims and Zha 1999).

A solution to these two problems within the Bayesian framework of Uhlig (2005) has been proposed in Inoue and Kilian (2013). This study shows how to characterize the most likely admissible model within the set of structural VAR models that satisfy the sign restrictions. The most likely structural model can be computed from the posterior mode of the joint distribution of admissible models both in the fully identified and in the partially identified case. The resulting set of structural response functions is well defined from an economic and a statistical point of view. Inoue and Kilian also propose a highest posterior density credible set that characterizes the joint uncertainty about the set of admissible models. Unlike conventional posterior error bands for sign-identified VAR models, the implied credible sets for the structural response functions characterize the full uncertainty about the structural response functions.

⁴ It is worth noting that this problem persists even if we restrict attention to the structural responses at horizon 0 because $\text{vec}(B_0^{-1})$ is a multidimensional object. Put differently, the posterior median for one element of $\text{vec}(B_0^{-1})$ may come from a different structural model than the posterior median of another element of $\text{vec}(B_0^{-1})$.

The procedure proposed by Inoue and Kilian (2013, 2017) can be summarized as follows. We first consider the case of a fully identified model, before extending the discussion to partially identified models.

Fully Identified Models. The objective is to rank the structural models in the admissible set. Ignoring the intercept, which does not enter the definition of the structural impulse responses, let $A \equiv [A_1, \dots, A_p]$, let P be the lower-triangular Cholesky decomposition of Σ_u and let $\text{vech}(P)$ denote the $K(K+1)/2 \times 1$ vector that consists of the on-diagonal elements and the below-diagonal elements of P . The $K \times K$ real orthogonal matrix $Q \in \mathcal{O}(K)$ satisfies $Q'Q = I_K$ and its determinant is $|Q| = 1$. Then

$$\mathbf{S} = I_K - 2(I_K + Q)^{-1}$$

is a $K \times K$ skew-symmetric matrix. Let \mathbf{s} denote the $K(K-1)/2 \times 1$ vector that consists of the above-diagonal elements of \mathbf{S} . Then the matrix Q and the vector \mathbf{s} have a one-to-one relationship. If Q is uniformly distributed over the space of real orthogonal $K \times K$ matrices $\mathcal{O}(K)$, the density of \mathbf{s} is given by

$$f(\mathbf{s}) = \left(\prod_{i=2}^K \frac{\Gamma(i/2)}{\pi^{i/2}} \right) \frac{2^{(K-1)(K-2)/2}}{|I_K + \mathbf{S}|^{K-1}},$$

where $\Gamma(x) \equiv \int_0^\infty z^{x-1} e^{-z} dz$ and $|\cdot|$ denotes the determinant (see equation (4) of León, Massé, and Rivest 2006, p. 415).

Let $B_0^{-1} = PQ$. Then there is a one-to-one mapping between the first $p+1$ structural impulse responses

$$\Theta = [B_0^{-1}, \Phi_1 B_0^{-1}, \Phi_2 B_0^{-1}, \dots, \Phi_p B_0^{-1}],$$

on the one hand, and the tuple

$$(\text{vec}(A), \text{vech}(P), \mathbf{s})$$

on the other hand. This one-to-one mapping allows us to derive the posterior density $f(\Theta)$ from the joint posterior density of $(\text{vec}(A), \text{vech}(P), \mathbf{s})$, where Θ is defined by the known nonlinear function $\Theta = h(\text{vec}(A), \text{vech}(P), \mathbf{s})$. Using the change-of-variables method, the posterior density of Θ can be written in closed form as

$$f(\Theta) \propto \left| \frac{\partial \text{vec}(\Theta)}{\partial [\text{vec}(A)', \text{vech}(P)', \mathbf{s}']} \right|^{-1} \left| \frac{\partial \text{vech}(\Sigma_u)}{\partial \text{vech}(P)'} \right| f(A | \Sigma_u) f(\Sigma_u) f(\mathbf{s}).$$

This result allows one to compute the numerical value of the posterior density for every possible draw Θ of the structural model. Let Θ denote the set of all admissible structural models that satisfy the sign restrictions. If the objective is to choose one admissible structural model from all possible realizations in Θ , then a natural approach is to select the modal (or most likely) model, defined as the structural model that maximizes the value of $f(\Theta)$ among all

models that satisfy the sign restrictions. By construction this model is immune from both shortcomings of the posterior median response functions discussed earlier. It must be emphasized that Inoue and Kilian (2013) do not propose to focus on the mode of the marginal distribution of the structural impulse responses, but rather on the mode of the joint distribution of the structural models.

The corresponding $(1 - \gamma)100\%$ highest posterior density (HPD) credible set may be defined by

$$\mathcal{S} = \{\Theta \in \Theta | f(\Theta) \geq c_\gamma\}, \quad (13.6.3)$$

where $f(\Theta)$ is the posterior density of Θ and c_γ is the largest constant such that

$$\mathbb{P}(\mathcal{S}) \geq 1 - \gamma.$$

This credible set is a joint credible set accounting for the dependence of the elements of Θ across time as well as across variables.

In practice we proceed as follows:

Step 1. Take a random draw, (A^{*r}, P^{*r}) , from the posterior of the reduced-form VAR parameters.

Step 2. For (A^{*r}, P^{*r}) , consider N random draws of the rotation matrix Q and for each combination (A^{*r}, P^{*r}, Q) compute the set of implied structural impulse responses Θ^{*r} .

Step 3. If Θ^{*r} satisfies the sign restrictions, store the value of Θ^{*r} and the value of $f(\Theta^{*r})$. Otherwise discard Θ^{*r} .

Step 4. Repeat steps 1, 2, and 3 M times. Sort in descending order by the value of $f(\Theta^{*r})$ the pairs $\{(\Theta^{*r}, f(\Theta^{*r}))\}$ that satisfy the sign restrictions.

Then the Θ^{*r} in the first sorted pair is the most likely model and the $(1 - \gamma)100\%$ HPD credible set is obtained by selecting the Θ^{*r} of the first $(1 - \gamma)100\%$ sorted pairs.

Inoue and Kilian (2013) make the case for focusing on the mode of the posterior distribution of the admissible structural models (with each model characterized by an entire set of structural impulse responses) as opposed to the pointwise median of the structural impulse response vector. It may seem that instead of focusing on the posterior mode in the admissible set of models one could have reported the posterior mean of the responses of the admissible structural models (or the responses of the admissible structural model whose impulse responses are closest to that posterior mean), given that the mean is a well-defined measure of central tendency for vector-valued random variables. There are two reasons for not proceeding along these lines. One is that it is not clear how to conduct inference about the posterior mean given that $f(\Theta)$ is highly non-Gaussian. The other is that the distribution in question need not have finite moments, as discussed in Baumeister and Hamilton (2015a).

The case has been made that the median response function may be motivated as the solution to a loss function involving the sum of the absolute loss of the vector of structural responses (see, e.g., Baumeister and Hamilton 2015a). While the median response function is indeed optimal conditional on this particular loss function, there are many possible loss functions one could maintain. The real question therefore is not whether there is a loss function under which the median response function is optimal, but rather what the right loss function is from the point of view of an economist evaluating the VAR model. It is not enough for a summary statistic to be statistically coherent. It also has to be economically meaningful. The loss function proposed by Baumeister and Hamilton (2015a), under which the median response function is optimal, effectively postulates that the economist using this structural VAR model is not interested in the shapes of the impulse response functions or in their comovement. Reporting pointwise posterior quantiles thus does not account for the dependence of structural impulse responses over time and across variables. Hence, the case can be made that Inoue and Kilian's loss function is more in line with the objectives of applied users. Moreover, the questions answered by their methodology, namely what the most likely structural models are within the admissible set, cannot be answered by computing the median response function.

Partially Identified Models. If we are concerned with a subset of structural impulse response functions only, as in the Uhlig (2005) example, what matters for constructing the posterior mode is not the joint impulse response distribution, but the marginalized distribution obtained by integrating out responses to shocks that are not economically identified. To simplify the exposition, we focus on the case in which only impulse responses to one structural shock are identified. Let b denote the column vector of B_0^{-1} relating to this shock and denote by q a random vector drawn from $\mathcal{N}(0, I_K)$ and normalized to have unit length. In other words, $\sum_{i=1}^K q_i^2 = 1$, where q_i refers to element i of the vector q . The K -dimensional vector b represents the impact responses of the model variables to the shock in question. It is computed as $b = Pq$ by analogy to $B_0^{-1} = PQ$. The implied structural impulse responses at higher horizons are $\theta_1 = \Phi_1 b, \dots, \theta_p = \Phi_p b$. Given $\theta_0 = b$ and Kp restrictions of the form

$$\theta_i = \Phi_i b, \quad i = 1, \dots, p,$$

the derivation of the marginal posterior density $f(\theta_0, \theta_1, \dots, \theta_p)$ requires numerical integration (see Inoue and Kilian 2013). On the basis of this density one may proceed exactly as in the fully identified case. In practice, this density must be evaluated by Monte Carlo integration, which greatly increases the computational cost compared with the algorithm for fully identified models.

Of course, as noted before, this approach does not ensure that the remaining structural shocks have a sign pattern that is distinct from the shock of interest.

13.6.4 The Penalty Function Approach

An alternative approach to selecting one model from the set of admissible structural models is to minimize a criterion function (or penalty function). This approach was first discussed in Uhlig (2005). Uhlig proposed to select from the set of candidate models the structural model that minimizes a penalty function. Rather than directly discarding models that violate sign restrictions, as in the conventional sign restriction approach, Uhlig attaches a large numerical penalty to such models. His penalty function not only punishes for violations of sign restrictions more than it rewards models with correct signs, but it involves an additional reward for large responses which may be tailored to specific responses and horizons. For example, a model in which monetary policy shocks generate a large initial increase in the federal funds rate, all else equal, may receive a larger reward than other models.

Uhlig's identifying restrictions are that an unexpected monetary policy contraction is associated with an increase in the federal funds rate, with price decreases, and with declines in nonborrowed reserves for some time following the policy shock. Uhlig's model is partially identified in that he is only interested in identifying responses to the monetary policy shocks. Thus, rather than having to identify B_0^{-1} , it suffices to identify the vector b , which represents the structural responses of the VAR model variables to a monetary policy shock. Given $(A, P = \text{chol}(\Sigma_u))$, Uhlig proposes to use a simplex algorithm to solve the problem

$$b = \arg \min_{b=Pq} \Psi(b),$$

where the penalty function is defined as

$$\Psi(b) = \sum_k \sum_{h=0}^H f\left(\iota_k \frac{\theta_{k,h}(b)}{\sigma_k}\right),$$

with the first sum being over $k \in \{\text{GDP deflator, commodity price index, non-borrowed reserves, federal funds rate}\}$. The parameter σ_k is the standard deviation of the growth rate of variable k , $\theta_{k,h}$ is the structural response of variable k at horizon h to a monetary policy shock, $\iota_k = -1$ for the federal funds rate and $\iota_k = 1$ otherwise, and

$$f(x) = \begin{cases} x & \text{if } x < 0 \\ 100 \times x & \text{if } x \geq 0 \end{cases}.$$

Note that the sign of the penalty is flipped for the federal funds rate, reflecting the nature of the sign restrictions in Uhlig's model. The rescaling by σ_k serves

to make the deviations across structural impulse responses more comparable to one another.

It is worth pointing out that Uhlig's penalty function never directly imposes the sign restrictions, although it is very unlikely that these restrictions would not hold in the model that minimizes the penalty function. Another difference from the conventional approach is that this algorithm, by construction, will produce solutions even when the conventional approach yields the empty set, because the identifying sign restrictions do not have to be satisfied at the minimum of the penalty function. In the latter case, the penalty function given (A, P) will find the structural model that comes closest to satisfying the sign restrictions.

The particular form of the penalty function originally proposed by Uhlig (2005) as an alternative to the pure sign restriction approach has evolved to include many other loss functions. Examples include Mountford and Uhlig (2009) and Beaudry, Nam, and Wang (2012). The latter two studies postulate additional short-run exclusion restrictions, however, and hence do not belong into the current section. Penalty function approaches also play an important role in studies of forward-looking behavior (see, e.g., Barsky and Sims 2011).

Critiques of Penalty Function Estimators. It is important to differentiate this approach from earlier work by Faust (1998) that also involved penalty functions. Faust applies a penalty function to select among many candidate monetary policy VAR models identified by sign restrictions the model that explains most of the variation in real output growth at a horizon of nine years. Both Faust and Uhlig are interested in measuring the effects of monetary policy shocks on the variability of real output growth, given a set of admissible structural models. Moreover, both use a penalty function. It may seem that these approaches are closely related, therefore, but there are important differences.

Faust uses a penalty function to assess the best-case scenario relative to the economic hypothesis of interest, conditional on the sign restrictions having been imposed. The result is best thought of as providing an answer to the question of whether some outcome is possible rather than whether it is true or whether it is the most likely outcome. In contrast, Uhlig uses the penalty function to select the best model from the point of view of the economic hypothesis of interest and treats it like a point estimate. Moreover, his penalty function in practice involves an additional layer of identifying restrictions that all admissible structural models must satisfy. In Uhlig's words, "one is, in effect, imposing somewhat more than just sign restrictions" (see Uhlig 2005, p. 414). Finally, whereas Faust looks at the set of admissible structural models satisfying only the sign restrictions, Uhlig also considers models that come close to satisfying the restrictions implied by the penalty function. As a result, one would expect the individual structural impulse response estimates selected by this procedure to be different from the original estimates in the admissible

set. Clearly, this approach can only be recommended if we are comfortable with all the identifying restrictions implied by the use of a penalty function.

The problem is that the additional restrictions usually are not very transparent. Uhlig (2005) is aware of this concern and cautions that the pure sign restriction approach is cleaner and more appealing because it only imposes weak prior beliefs about the signs of impulse responses. Many applied users of penalty functions, however, seem unaware of the fact that they are imposing additional implicit identifying assumptions or, at any rate, have no idea what additional restrictions may be implied by the penalty function. Indeed, making these assumptions explicit can be prohibitively difficult.

In some cases the penalty function approach has been shown to imply additional unintended or at least unanticipated sign restrictions. Intuitively, the problem is that by choosing a particular orthogonal matrix that minimizes the response of one of the model variables to a given shock, we end up biasing the responses of other variables to the same shock. As a result, the penalty function involves additional restrictions on variables that are seemingly unrestricted (see Arias, Rubio-Ramírez, and Waggoner 2013). For example, Caldara and Kamps (2017) demonstrate that the penalty function used in Mountford and Uhlig (2009) amounts to imposing an additional sign restriction on the output response to a tax increase. A more detailed analysis of this example and the related work of Beaudry, Nam, and Wang (2012) can be found in Arias, Rubio-Ramírez, and Waggoner (2013), who explicitly trace the additional sign restrictions to the fact that the penalty functions in question reward structural models with large responses of some variables to some structural shocks.

These examples cast doubt on the use of penalty function estimators more generally. On the one hand, the use of a penalty function clearly cannot be recommended unless we understand its implications for identification. On the other hand, if we do, and if the implied additional restrictions involve sign restrictions as in the examples above, these sign restrictions should have been imposed directly, with the penalty function only serving to select a unique structural model.

Inference on Penalty Function Estimators. There are also questions about how to conduct valid inference for penalty function estimators. Uhlig (2005) proposes to apply the penalty function to the subset of sign-identified structural models obtained by exploring many solutions for Q , conditional on a given draw of the posterior for (A, P) . This means that for each draw from the reduced-form posterior, the procedure will return just one estimate of the structural model. By repeating this procedure for many draws from the reduced-form posterior, one can approximate the posterior distribution of the penalty function estimator. Uhlig proposes to summarize this posterior by constructing posterior median response functions and pointwise posterior quantile bands.

There are three concerns with this approach. One concern is that, as noted earlier, the use of median response functions and pointwise error bands can be misleading. A second concern is that the penalty function approach tends to underestimate the posterior uncertainty (relative to the correct measure of uncertainty based on the explicit sign restrictions only) to the extent that it implicitly imposes additional identifying structure. The third concern is that measures of the posterior uncertainty of estimates obtained by the penalty function approach underestimate the uncertainty because they condition on one possible choice of the rotation matrix ignoring the existence of other solutions that satisfy the identifying restrictions (see Arias, Rubio-Ramírez, and Waggoner 2013).

Alternative Uses of Penalty Functions. None of these concerns arises if a penalty function is applied to choose among the models contained in the admissible set, as in Faust (1998). In that case we first generate many draws from the reduced-form posterior for (A, P) and, conditional on each of these draws, generate many draws of rotation matrices Q . Only the structural model solutions that satisfy the sign restrictions are retained. The resulting admissible set simultaneously accounts for estimation uncertainty and identification uncertainty. Applying a penalty function to choose among the models in this set will generate a unique structural model.

For example, Faust (1998) chooses the structural VAR model that maximizes the share of the variability of real output growth explained by the monetary policy shock. Similarly, Francis, Owyang, Roush, and DiCecio (2014) identify a technology shock as the shock that satisfies suitable sign restrictions and maximizes the forecast error variance share in labor productivity at some finite horizon. Unlike Uhlig (2005), they first impose the sign restrictions and then search for the structural model in the identified set that maximizes the forecast error variance.

Another example is Kilian and Lee (2014) who use an external estimate of the price elasticity of oil demand to select the structural oil market VAR model in the identified set that comes closest to implying a price elasticity of oil demand with that value. Rather than just establishing a best-case or worst-case scenario, as in Faust (1998), their analysis uses an external estimate of a model parameter to determine the best-fitting structural model within the admissible set.

13.6.5 Using Historical Information to Narrow the Set of Admissible Models

A very different approach to narrowing down the set of admissible models is to discriminate among these models based on their historical decompositions.

A case in point is Kilian and Murphy (2014) who utilize extraneous information about the role of oil supply shocks and speculative oil demand shocks during specific historical episodes to judge the economic plausibility of alternative admissible models. For example, there is extraneous information from oil industry sources about a surge in speculative demand in the second half of 1979, allowing us to discard models that do not replicate this feature. Likewise we know that the spike in the price of oil in 1990, following the invasion of Kuwait, must be related at least in part to an oil supply disruption and was not caused by a booming economy. We also know that there was an important shift in oil price expectations that must have been reflected in increased speculative demand and higher oil prices in 1990.

This approach amounts to imposing additional identifying restrictions that can only be imposed on the historical decompositions rather than on the structural impulse responses. Unlike the penalty function approach, this approach typically does not produce a single admissible model, but it may be used to narrow down the admissible set considerably. Of course, its application is limited to settings in which the researcher has accurate information about the quantitative importance of a given structural shock during a specific historical episode.

This approach has recently been formalized by Antolín-Díaz and Rubio-Ramírez (2016). Building on the framework of Rubio-Ramírez, Waggoner, and Zha (2010), they show how both the sign of shocks occurring on specific dates and the sign of the cumulative contribution of structural shocks during specific episodes may be restricted when estimating the sign-identified VAR models. For example, one may impose the restriction that the contribution of one shock is larger in absolute magnitude than that of some other shock (or possibly of all other structural shocks combined) during a given episode. These additional identifying restrictions based on extraneous evidence may considerably reduce the range of model solutions consistent with the data and tend to tighten the credible sets. Antolín-Díaz and Rubio-Ramírez (2016) apply this approach to the oil market VAR model of Kilian and Murphy (2012) and the semistructural VAR model of monetary policy of Uhlig (2005).

13.7 The Role of the Prior for the Rotation Matrix

The reason why classical estimation methods are inherently uninformative about which of the admissible structural models is most likely is that the likelihood is flat with respect to the choice of the rotation matrix. An obvious question is how Bayesian methods are seemingly able to overcome this problem. From a mechanical point of view, the answer is that they rely on a prior distribution for the rotation matrix Q . This prior also is known as a Haar prior in the literature. Applying the Givens and Householder transformations to generate draws for Q is equivalent to generating draws from this prior. Although

the marginal prior distribution of Q is uniform in the Haar space, as noted earlier, the implied priors for the impact multiplier matrix used in constructing the structural impulse responses are clearly informative. This result follows from the fact that the impact responses are a weighted average of the elements of Q which themselves are not uniform. The resulting informative prior for the structural impulse responses is not based on economic information, however, and there is no way for the data to overrule this prior even asymptotically because the likelihood does not depend on Q . This fact raises obvious concerns about the conventional Bayesian approach to estimating structural VAR models subject to sign restrictions.

An important practical question is to what extent the posterior distribution of the estimated structural VAR models, Θ , depends on the prior for Q , as opposed to the data. On the one hand, one certainly can construct examples in which the posterior essentially mirrors the prior. On the other hand, given that Inoue and Kilian (2013) evaluate the posterior of Θ , which includes structural impulse responses beyond the impact period, it is possible that much of the posterior information, which depends in part on A and P , may come from the data rather than the prior on Q . This is an empirical question. If the location and the concentration about the mode of the posterior distribution differed substantially from the location and concentration of the prior distribution of Θ , this outcome would increase our confidence in the posterior of the sign-identified model. If not, the outcome would cast doubt on the results from sign-identified VAR models. Of course, this diagnostic only reveals the extent to which the prior on Q affects inference about the structural impulse responses without addressing the root cause of this problem.

The ultimate cause of this problem is that the traditional approach does not specify the prior directly on the object of interest, which are the structural impulse responses, but on the model parameters on which the structural impulse responses depend. One response to this problem has been the development of the frequentist approaches to inference in sign-identified models reviewed in Section 13.6.1. These approaches have their own limitations, however. Another response has been to modify the way Bayesian estimation and inference is conducted.

13.7.1 An Approach Based on Explicit Bayesian Priors for B_0

Baumeister and Hamilton (2015a) formally demonstrate that the marginal prior distribution for Q is informative about the structural model parameters even asymptotically. They show that the nature of this implicit prior distribution varies with the dimension K of the VAR model. For example, in a standard bivariate model of price and quantity driven by demand and supply shocks, in which sign restrictions on the impact responses are the only identifying assumptions, the price elasticities of demand and supply, which can be shown

to be linear functions of selected elements of B_0 , have a truncated Cauchy distribution under conventional prior specifications for sign-identified structural VAR models. If the reduced-form residuals are positively correlated, the model a priori allows any value of the price elasticity of demand but restricts the price elasticity of supply to fall within a certain interval. With negatively correlated errors, in contrast, the elasticity of supply could be any positive number, whereas the elasticity of demand is restricted to fall within a particular interval. Thus the choice of the Haar prior in this example is anything but innocuous. Although such prior restrictions on the range of the price elasticities need not arise in larger-dimensional models, as shown in Kilian and Murphy (2014), there is no doubt that the use of the Haar prior may affect the posterior of the structural impulse responses and the implied price elasticities in unknown and possibly undesirable ways.

Baumeister and Hamilton's proposed solution is to specify the prior directly on the elements of B_0 , generalizing the approach of Sims and Zha (1998). They impose the restriction that the marginal priors for each element of B_0 are independent of one another, facilitating the task of specifying the joint prior. The marginal priors on the elements of B_0 may be uninformative (flat or diffuse) or deliberately informative. One obvious drawback of this proposal is that any prior on the elements of B_0 implies an informative prior on B_0^{-1} . The nature of that prior is not made explicit and need not coincide with the prior views of the researcher about the structural impulse responses. Thus, unless the case can be made that the elements of B_0 are the object of ultimate interest rather than the structural impulse responses, this approach suffers from exactly the same problem as the traditional approach it seeks to replace. Put differently, Baumeister and Hamilton (2015a) do not propose a solution to the problem of specifying a prior directly on B_0^{-1} (or more generally on all structural impulse responses), but they are proposing a solution to the problem of modeling the parameters of the structural VAR model.

There are examples where their approach is appealing, as illustrated by the bivariate empirical example in Baumeister and Hamilton (2015a). For many applications of sign-identified VAR models in the literature, however, this approach is problematic. This point is illustrated by the additional empirical examples provided in Baumeister and Hamilton (2015c).

Priors on Elasticities. Baumeister and Hamilton (2015c) focus on oil market VAR models including the model of Kilian and Murphy (2012). The central premise of Baumeister and Hamilton is that the price elasticities of demand and supply can be written as linear functions of the elements of B_0 , allowing them to impose priors on these elasticities. This approach makes sense provided that at each point in time the production and consumption of the commodity in question coincide, as assumed in the standard bivariate model of prices and quantities. That assumption does not hold for storable commodities such as

crude oil, however. In the latter case, production equals consumption plus the change in inventories. As a result, the price elasticity of oil demand no longer is a linear function of the elements of B_0 (see Kilian and Murphy 2014).

This fact not only makes it impossible to impose economically motivated priors on this elasticity by means of priors on selected elements of B_0 , but it also invalidates Baumeister and Hamilton's maintained assumption that the marginal priors on B_0 are independent. The price elasticity of oil demand, properly defined to account for the role of oil inventories, depends on both the response of production and the response of inventories to a change in the real price of oil triggered by an exogenous shift in the oil supply curve. Hence, a prior on this price elasticity by construction has implications for the prior of more than one element of B_0 , violating the assumption of independence. In short, Baumeister and Hamilton's approach is not suitable for this class of problems.

In addition, this fact matters for the interpretation of the results. Baumeister and Hamilton (2015c) suggest that oil market VAR models such as Kilian and Murphy (2012) or Kilian (2009) that impose tight upper bounds on the price elasticity of oil supply are fundamentally flawed because they imply unrealistically large price elasticities of oil demand. That conclusion, however, is based on Baumeister and Hamilton misinterpreting the coefficients of B_0 in the model of Kilian and Murphy (2012) as implying an estimate of the price elasticity of oil demand. In fact, that model (like the earlier model of Kilian 2009) is not designed to infer the price elasticity of oil demand. If the structural oil market VAR model is suitably modified to allow proper estimation of this elasticity, as shown in Kilian and Murphy (2014), price elasticities of oil supply close to zero are perfectly consistent with reasonably low price elasticities of oil demand.

Informative versus Uninformative Priors. Baumeister and Hamilton suggest that we embrace the fact that priors in sign-identified VAR models are informative and make this information content explicit. An obvious concern with intentionally specifying informative priors for the elements of B_0 in sign-identified models is that it is not clear why a prior that merely reflects the personal views of the user should be of wider interest to other economists. In fact, an unsympathetic observer may view the use of subjective priors as an attempt to circumvent the lack of identification by simply imposing the answer. Thus, at best this approach may provide an opportunity for illustrating how alternative subjective priors affect the posterior.

Baumeister and Hamilton are aware of this concern and make the case that imposing informative priors makes sense when there is extraneous identifying information on the structural model parameters that has not been utilized in existing research. For example, there may be extraneous information about the value of elasticity parameters from microeconomic studies that can be imposed

in estimation. Given the uncertainty about such estimates, it makes sense to think of this information as being stochastic with the prior parameterizing our degree of confidence in these estimates. Unlike conventional subjective priors, priors based on extraneous evidence can be viewed as objective in that there should be agreement among economists on this prior information.

This approach clearly is appealing at first sight. As the oil market VAR examples in Baumeister and Hamilton (2015c) illustrate, however, it can be difficult to find such identifying information beyond restrictions such as elasticity bounds that have already been imposed in the existing literature (see Kilian and Murphy 2014; Kilian and Lee 2014). Moreover, it is not clear how exactly to map extraneous evidence about elasticities into priors without introducing subjective elements.

What Do We Learn from Extraneous Elasticity Estimates? It is important to be precise about what we learn from extraneous evidence about the price elasticities of oil demand and oil supply. These parameters play a central role in oil market VAR models. Kilian and Murphy (2012) demonstrate that the price elasticity of oil supply is crucial in determining the relative importance of oil demand and oil supply shocks in explaining oil price fluctuations. If that elasticity is close to zero, one obtains the by now standard result that oil demand shocks have been the primary driver of oil price fluctuations all the way back to the 1970s, with rare exceptions such as 1990 and to some extent 2014/15.

The rationale for a price elasticity of oil supply close to zero in Kilian and Murphy (2012) and related studies is threefold: (a) Narrative evidence on decisions by major oil producers such as Saudi Arabia shows a reluctance to respond to price fluctuations driven by oil demand (see Kilian 2009); (b) microeconometric evidence based on production decisions by U.S. oil producers suggests elasticity estimates of 0.0009 (e.g., Anderson, Kellogg, and Salant 2016); (c) economic theory shows that in equilibrium, oil producers do not respond to oil price fluctuations caused by oil demand shocks in the short run; all the adjustment works through investment (see Anderson, Kellogg, and Salant 2016).

Baumeister and Hamilton depart from this consensus and make the case for priors that put more weight on lower price elasticities of oil demand and higher price elasticities of oil supply than found in earlier oil market VAR studies. Such a change in the prior has far-reaching implications because it is expected to improve the ability of oil supply shocks to explain fluctuations in the real price of oil at the expense of oil demand shocks. In support of their prior, Baumeister and Hamilton appeal to a range of estimates of price elasticities of oil demand and oil supply in the existing literature, some of which date back to the 1970s and 1980s.

It may seem tempting simply to aggregate all existing extraneous elasticity estimates, but that approach would be misleading because not all estimates in the literature are equally credible. Although there are numerous extraneous microeconometric estimates of the short-run price elasticity of oil demand in the literature, many of these estimates are questionable. Not only do traditional estimates of this elasticity ignore the storability of crude oil, but they are based on OLS reduced-form regressions of quantity on price and hence suffer from simultaneous equation bias. Econometric theory implies that these elasticity estimates are biased toward zero and are not economically meaningful. More recent estimates of the price elasticity of oil demand from structural econometric models are invariably larger, as noted by Kilian and Murphy (2014). This conclusion is consistent with recent IV estimates of the short-run price elasticity of gasoline demand, which in turn have been externally validated (see Coglianese, Davis, Kilian, and Stock 2017). For example, the estimate of the one-month price elasticity of oil demand of -0.25 in Kilian and Murphy (2014) is much higher than the modal value of -0.1 in Baumeister and Hamilton's baseline prior specification. Likewise, for the short-run price elasticity of oil supply, there are few credible microeconometric estimates, but the available empirical evidence, anecdotal evidence, case studies, and economic theory all point to an elasticity much closer to zero than Baumeister and Hamilton's modal value of 0.1 . Thus, there is no evidence that the existing literature on oil market VAR models has ignored important identifying information, but there are questions about the empirical support for the elasticity priors proposed by Baumeister and Hamilton.

How Do We Translate Extraneous Elasticity Estimates into a Prior?

Even if one can agree on a reasonable prior mean or mode for these elasticities, the problem remains of how to translate the extraneous information into prior densities for the elasticities that reflect more than introspection. Say, for example, that some earlier study reports that the short-run price elasticity of oil supply is 0.02 . This does not tell us what the functional form of the prior density should be. Nor does it tell us what the dispersion of the prior density should be. In fact, even if one had multiple credible estimates of the price elasticity of supply, say 0.05 and 0.02 , this would not suffice to specify the prior density any more than one estimate. In other words, the elasticity priors specified by Baumeister and Hamilton in the end come from introspection rather than extraneous information. Perhaps for this reason, Baumeister and Hamilton (2015c) appear to favor priors intended to be fairly diffuse. In the end, they seek to establish that their main results are robust to alternative prior specifications for B_0 rather than relying on additional extraneous evidence.

Incompatibility with the Identifying Restrictions Used in Earlier Studies. Baumeister and Hamilton specify independent marginal priors on each element

of B_0 . These priors may take the form of a truncated uniform or a t distribution, for example. As discussed earlier, in applied work, it is determined by the economic model of interest whether a user imposes identifying restrictions on the elements of B_0 or B_0^{-1} . These representations in general are not exchangeable.

Thus, when Baumeister and Hamilton (2015c) attempt to illustrate the benefits of their approach to estimation and inference in the context of the oil market VAR model of Kilian and Murphy (2012), for example, they do not rely on the same economic structure and on the same identifying assumptions as the original study. Their analysis both ignores some of the identifying sign restrictions imposed by Kilian and Murphy (2012) and imposes additional exclusion restrictions on B_0 not found in the original specification. Specifically, Baumeister and Hamilton postulate that global real economic activity only affects oil production contemporaneously through its effect on the real price of oil and that global oil production does not enter directly in the global real economic activity equation. Hence, the model considered by Baumeister and Hamilton is at best similar to the original model, making it difficult to attribute differences in the estimates to the choice of the prior. Put differently, the point is not that Baumeister and Hamilton chose not to replicate the original identifying assumptions in Kilian and Murphy (2012) because they disagree with their economic content, but that their methodology does not allow them to incorporate these assumptions.

More generally, the alternative approach proposed by Baumeister and Hamilton does not allow for any dynamic sign restrictions or cross-equation restrictions of the type utilized by many earlier oil market VAR models. Baumeister and Hamilton (2015a) argue that such restrictions are irrelevant for identification, but that conclusion only holds under the particular loss function they adopt.

In short, the objective of the analysis in Baumeister and Hamilton is not to provide an improved way of estimating the sign-identified models proposed by other researchers; rather it is to replace the structural models and identifying restrictions used by those researchers by alternative models and assumptions. Whether these alternative models are more or less economically appealing than the original models has to be decided on a case-by-case basis.

Choice of the Loss Function. Any comparisons with the results in the existing literature are further complicated by the fact that Baumeister and Hamilton choose to rely on a specific loss function that allows them to effectively use the median response function as the point estimate of the structural impulse response function. This choice of the loss function is controversial. Median response functions suffer from the shortcomings we discussed earlier and are not informative about the questions most applied users care about. Whether the

prior for the elements of B_0^{-1} is implicit or explicit does not affect this conclusion. As illustrated in Inoue and Kilian (2013), for a given prior, evaluating the set of admissible models based on the median response function rather than the responses of the most likely structural model may affect not only the magnitude but even the sign of the structural responses.

Thus, it is difficult to know whether any differences between the estimates reported in Baumeister and Hamilton (2015c) and in the original studies are driven by changes in the model specification, in the prior, in the identifying assumptions, or in the loss function. Notwithstanding these caveats, Baumeister and Hamilton (2015c) end up confirming many of the substantive conclusions of earlier oil market studies, suggesting that the prior for Q embodied in conventional Bayesian estimates of sign-identified structural VAR models is not overly informative in these applications.

13.7.2 An Approach Based on Explicit Bayesian Priors for the Structural Impulse Responses

A closely related alternative approach that allows the user to impose priors directly on elements of B_0^{-1} and on the shapes and smoothness of the structural impulse response functions has been proposed by Plagborg-Møller (2016). Let w_t be a K -dimensional vector of structural shocks. In practice, w_t is typically modeled as iid Gaussian. The central idea is to estimate the structural moving average representation,

$$y_t = \Theta(L)w_t,$$

up to the maximum horizon of interest directly rather than estimating the structural VAR representation,

$$B(L)y_t = w_t,$$

first and then inverting that representation, as proposed by Baumeister and Hamilton (2015a). Plagborg-Møller proposes to impose a multivariate prior directly on the first $H + 1$ coefficient matrices of $\Theta(L)$, including the structural impact multiplier matrix B_0^{-1} and the structural impulse responses at horizons $h = 1, \dots, H$. As in Baumeister and Hamilton (2015a), this approach can be used to explicitly allow for identification uncertainty or it may be used to impose a degenerate prior on certain elements of B_0^{-1} . Thus, traditional short-run and long-run exclusion restrictions as well as static and dynamic sign restrictions on structural impulse responses are covered as special cases. Short-run exclusion restrictions may be imposed by dropping the elements in question from the set of structural impulse responses. Long-run exclusion restrictions may be imposed as a constraint in evaluating the likelihood. Sign restrictions may be imposed by restricting the parameter space of the

structural impulse responses to the subspace where the inequality restrictions hold.

One potential advantage of this approach compared with Baumeister and Hamilton (2015a) is that it is in line with most applied studies in seeking to restrict the structural impulse responses rather than the parameters of the structural model. It also allows the user to impose shape restrictions on the structural impulse response functions. Such restrictions can be incorporated as constraints in the evaluation of the likelihood. Moreover, the smoothness of the structural impulse response functions can be controlled by the prior covariance across structural impulse responses.

The obvious cost is that the parameters of the Wold reduced-form MA representation are no longer pinned down by the data alone. To conduct posterior inference about the structural impulse responses, Plagborg-Møller (2016) develops a simulation algorithm that exploits the Whittle (1953) likelihood approximation. He shows how this posterior may be approximated by a Markov Chain Monte Carlo method. He then derives the frequentist limit of this posterior distribution under weaker assumptions than Moon and Schorfheide (2012). Under some simplifying assumptions including covariance stationarity of the data, it can be shown that the data will asymptotically pin down the coefficients of the reduced-form Wold representation, but the structural impulse responses in this framework are only set-identified, making their posterior distribution sensitive to the choice of the prior for B_0^{-1} . Plagborg-Møller follows Baumeister and Hamilton in embracing this fact, arguing for the imposition of explicitly informative priors in identifying the structural impulse responses.

The main challenge in implementing Plagborg-Møller's approach thus is the derivation of the priors. There are two distinct concerns. One is that much of what we learn from the posterior reflects the prior imposed in estimation. This is a bigger concern than in other approaches to sign-identified models, because imposing a prior on the first $H + 1$ coefficient matrices of $\Theta(L)$ implicitly imposes an informative prior on the reduced-form representation of the model. Even if the effect of that prior on the posterior is negligible asymptotically, it will matter in finite samples.

The other concern is that coming up with universally accepted priors may be difficult. For example, it is not clear how to derive a nondogmatic prior density for the elements of B_0^{-1} in general. Nor is there empirical support for the claim that economists in the past have routinely failed to impose readily available identifying information. Plagborg-Møller shows by example that in some situations priors for the parameters of $\Theta(L)$ may be elicited from DSGE models, but in many applications that approach may not be feasible or desirable. Moreover, even if a particular economic model suggests a certain pattern of identifying restrictions, this identifying information may not be robust to changes in the specification of this economic model. Imposing a diffuse prior

on the dynamics of the structural impulse responses, in contrast, avoids this problem, but is likely to imply very wide credible sets.⁵

Another unresolved question is how to report the posterior evidence. Plagborg-Møller relies on posterior mean vectors for the structural responses and pointwise HPD error bands. He also reports selected elements of the shotgun trajectory plot. A useful generalization would be to derive the most likely structural model and the joint credible set from the posterior distribution, building on the analysis in Inoue and Kilian (2013).

In short, the approach taken by Plagborg-Møller (2016) provides a useful complement to the analysis in Baumeister and Hamilton (2015a) in that it considers identifying restrictions not covered by their analysis. At this point, we do not have much experience in making this alternative approach operational, however. It remains to be seen whether this proposal may be developed into a practically useful alternative to conventional sign-identified VAR models.

13.7.3 A Robust Bayesian Approach

A very different response to the problem of informative priors on Q in sign-identified structural VAR models is due to Giacomini and Kitagawa (2015). Whereas Baumeister and Hamilton (2015a) propose deriving prior densities that reflect the user's beliefs about the parameters of B_0 and Plagborg-Møller (2016) proposes imposing priors directly on the structural impulse responses (including the elements of B_0^{-1}), Giacomini and Kitagawa's proposal is to construct posterior bounds on B_0^{-1} (and more generally on the structural impulse responses) without taking a stand on the nature of the prior for Q . In other words, Giacomini and Kitagawa are not concerned with the economic content of the prior for Q . The only economic information used in their approach involves the imposition of sign restrictions (or other identifying restrictions) on the structural impulse responses.

Giacomini and Kitagawa show that the problem of constructing pointwise credible sets from the posterior of the structural responses derived under all possible priors for Q is equivalent to constructing pointwise posterior bounds on each of the structural impulse responses. Their procedure allows the structural VAR model to be partially identified (or underidentified) in that only a subset of the structural shocks is identified. In practice, Giacomini and Kitagawa's procedure only relies on a prior for the reduced-form parameters of the VAR model and on a set of identifying restrictions. One practical drawback of their procedure is that the construction of the bounds involves a complicated

⁵ A closely related Bayesian approach has been developed by Barnichon and Matthes (2016). Unlike Plagborg-Møller, they impose additional structure on the coefficients of the structural MA representation to reduce the dimensionality of the estimation problem. Specifically, they approximate the impulse response functions with so-called Gaussian basis functions.

nonlinear optimization problem for each posterior draw. Moreover, in practice, the resulting credible sets are likely to be so wide as to be uninformative. Finally, as in frequentist approaches to inference in sign-identified models, the estimates can be difficult to interpret from an economic point of view.

13.7.4 An Agnostic Bayesian Approach

An alternative Bayesian approach was recently proposed by Arias, Rubio-Ramírez, and Waggoner (2015). Their objective is to ensure that the prior we use in applied work is agnostic in that it does not imply additional restrictions beyond the sign restrictions that the user wishes to impose. The analysis is based on the observation that we can represent a structural VAR model in terms of the parameters of the structural VAR model representation, B_0, \dots, B_p , in terms of a set of structural impulse responses, Θ , or in terms of the representation (A, Σ_u, Q) . Given A and P such that $PP' = \Sigma_u$, each value of the orthogonal matrix Q determines a particular structural model (see also Inoue and Kilian 2013). Arias et al. define a prior over the structural model representation (or over the structural impulse response representation) to be agnostic with respect to the identification if the prior density is invariant to the choice of $Q \in \mathcal{O}(K)$ and show that a prior is agnostic if and only if it is equivalent to a prior over (A, Σ_u, Q) that is flat over $Q \in \mathcal{O}(K)$. This result immediately implies that the conventional Bayesian approach to estimating models subject to sign restrictions discussed earlier is agnostic in this sense. It is important to note, however, that a prior being agnostic in this technical sense does not mean that the prior is not informative for the structural impulse responses. In particular, the implied prior for the structural impulse responses is not flat.

As an alternative to the agnostic prior, Arias et al. also consider the construction of jointly flat priors over the structural model representation or over the structural impulse response representation, respectively. Their representation of this problem is more general than that in Baumeister and Hamilton (2015a) in that it allows for the imposition of dynamic sign restrictions in addition to static sign restrictions. Arias et al. prove that if a prior is flat over either the structural model representation or the structural impulse response representation, then the equivalent of that prior written in terms of the representation (A, Σ_u, Q) is flat over $\mathcal{O}(K)$. The fact that the conventional prior discussed in Rubio-Ramírez, Waggoner, and Zha (2010) is informative about the structural model parameters and about the structural impulse response parameters thus does not result from the assumption of a flat prior over the rotation matrices Q ; rather, it stems from their choice of the prior for the reduced-form parameters. This fact suggests that we need to change the reduced-form prior to ensure that the prior is flat in the desired dimension as well as agnostic.

Arias et al. propose to choose a conjugate prior within the class of Gaussian-inverse Wishart priors such that the prior is flat over either the structural model

representation or the structural impulse response representation, depending on the user's preferences. One can be flat in one of these dimensions but not in both. They show that a prior over the structural model representation is flat if and only if the equivalent prior over the representation (A, Σ_u, Q) equals

$$2^{\frac{-K(K+1)}{2}} |\det(\Sigma_u)|^{-\left(\frac{K_p}{2} + K + 1\right)},$$

which implies a flat prior for Q on $\mathcal{O}(K)$ and a Gaussian-inverse Wishart prior of a particular form. Similarly, a prior over the structural impulse response representation is flat if and only if the equivalent prior over the representation (A, Σ_u, Q) equals

$$2^{\frac{-K(K+1)}{2}} |\det(\Sigma_u)|^{-\left(\frac{K_p}{2} - 1\right)},$$

which implies a flat prior for Q on $\mathcal{O}(K)$ and a Gaussian-inverse Wishart prior of a different form.

13.7.5 A Non-Bayesian Approach

Finally, a very different approach has been pursued in Baumeister and Kilian (2016b) in the context of a global oil market VAR model. Their objective is to quantify the extent to which the decline in the price of oil in the second half of 2014 is explained by oil demand and oil supply shocks taking place after June 2014. For this purpose, Baumeister and Kilian develop a novel identification strategy based on recursive estimates of the reduced-form VAR representation. As in Kilian and Murphy (2014), the vector of VAR model variables includes global oil production, global real economic activity, global crude oil inventories, and the real price of crude oil. At each point in time, the VAR model generates a set of reduced-form prediction errors for each of the model variables obtained by comparing the one-step ahead model predictions with the subsequent realizations of the model variables. These prediction errors by construction reflect the oil demand and oil supply shocks in the underlying structural VAR model.

If a given set of prediction errors is driven primarily by one of the structural oil demand and oil supply shocks, we can use identifying information about the signs and relative magnitudes of the prediction errors to infer which of the structural shocks can explain the observed pattern of prediction errors. For example, a negative oil supply shock would be associated with a large positive prediction error in the real price of oil and a large negative prediction error in oil production, but a small negative prediction error for real activity and a small negative prediction error for oil inventories. Whether a prediction error for oil production is large can be judged based on the existing literature on large oil supply shocks (see, e.g., Hamilton 2003). Likewise, prediction errors for oil inventories can be expressed relative to the stock of oil inventories

to obtain a measure of their magnitude. The corresponding patterns implied by flow demand or speculative demand shocks may be derived in a similar manner.

This identifying information allows one to assess whether the pattern of prediction errors for a given month is consistent with any one of the structural shocks. It is possible, of course, for the prediction errors to reflect a multitude of structural shocks of large magnitude. In the latter case, the pattern of prediction errors will not fit the patterns implied by any one structural shock, and the shocks cannot be identified. It is also possible that the shocks in a given month are so small that the prediction error for the real price of oil is negligible, in which case the question of identification is moot.

In their particular application, Baumeister and Kilian (2016b) show that between July 2014 and December 2014 there were only two large oil price prediction errors, one in July and one in December. The first prediction error was unambiguously associated with a reduction in speculative demand driven by a shift in expectations about the future price of oil, and the second prediction error was unambiguously driven by a reduction in flow demand, reflecting an unexpected weakening of the global economy. In contrast, the alternative hypothesis of a shift in oil price expectations in December 2014, triggered by the OPEC announcement of late November, can be ruled out on the basis of the pattern of prediction errors.

The advantage of this approach to identification is that it does not require any prior for Q , and, in fact, does not require the user to employ a Bayesian approach to estimating the VAR model, yet it may allow one to quantify numerically the extent to which structural shocks shifted the real price of oil. Its most important disadvantage is that it does not generate numerical estimates of the structural shocks. The effect of a structural shock can only be judged informally by comparing the path of model predictions after the shock has occurred to the path predicted before this shock occurred.

13.8 Examples of Models Identified by Sign Restrictions

13.8.1 A Small-Scale Macroeconomic Model

In Chapter 8, we examined a stylized quarterly semistructural model of monetary policy, in which the monetary policy shock was identified by exclusion restrictions that ruled out contemporaneous feedback from the policy shock to the variables ordered above the interest rate. These identifying assumptions are neither credible in general nor consistent with the implications of most general equilibrium models. Sign restrictions provide a natural alternative.

As discussed in Fry and Pagan (2011), we may postulate instead a positive demand shock that on impact raises economic growth (Δgdp), inflation (π), and the interest rate (i); a positive cost-push (or negative supply)

shock that on impact lowers growth but raises inflation and the interest rate; and a contractionary monetary policy shock that on impact raises the interest rate but lowers inflation and growth. Formally these sign restrictions can be represented as

$$\begin{pmatrix} u_t^{\Delta gdp} \\ u_t^\pi \\ u_t^i \\ u_t^i \end{pmatrix} = \begin{bmatrix} + & - & - \\ + & + & - \\ + & + & + \end{bmatrix} \begin{pmatrix} w_t^{\text{AD}} \\ w_t^{\text{AS}} \\ w_t^{\text{monetary policy}} \end{pmatrix}.$$

13.8.2 A Slightly Larger Macroeconomic Model

Peersman (2005) postulates a model of the U.S. economy based on a covariance stationary structural VAR model for the percent change in the nominal price of oil (Δo), real output growth (Δq), consumer price inflation (Δp), and the short-term nominal interest rate (i). Cointegration among the variables in levels has been ruled out. The real price of oil is implicitly assumed to be $I(1)$ so that Δo is $I(0)$. The vector of structural shocks includes a nominal oil price shock, a domestic aggregate supply shock, a domestic aggregate demand shock, and a domestic monetary policy shock. Identification involves sign restrictions of the form

$$\begin{pmatrix} u_t^{\Delta o} \\ u_t^{\Delta q} \\ u_t^{\Delta p} \\ u_t^i \end{pmatrix} = \begin{bmatrix} + & * & + & - \\ - & + & + & - \\ + & - & + & - \\ + & - & + & + \end{bmatrix} \begin{pmatrix} w_t^{\text{oil price}} \\ w_t^{\text{AS}} \\ w_t^{\text{AD}} \\ w_t^{\text{monetary policy}} \end{pmatrix},$$

where * stands for an unrestricted element.

An unexpected monetary policy tightening is associated with higher interest rates but lower real output and lower inflation on impact. The impact effect on the nominal price of oil is negative, consistent with the response of the price level. A positive domestic aggregate demand shock is associated with positive impact responses of all model variables. A positive oil price shock is interpreted as a negative domestic aggregate supply shock. It raises the price of oil, raises inflation, lowers real output, and raises the interest rate. A positive domestic aggregate supply shock, in contrast, shows the opposite signs, except that its effect on the nominal price of oil is treated as unknown because of the conflicting signs of the real and nominal effects of such a shock. On the one hand, one would expect the nominal price to increase because of higher real demand for oil; on the other hand, one would expect it to fall because of the deflationary effects of positive domestic aggregate supply shocks on dollar-denominated prices.

On the basis of the sign restrictions alone, it cannot be ruled out that a positive oil price shock may be observationally equivalent to a negative domestic

aggregate supply shock because both shocks would have in common the same sign pattern if the negative aggregate supply shock were to lower the nominal price of oil, which is not precluded by the assumptions so far. Because the oil price shock and the aggregate supply shock are not individually identified, Peersman assumes that of these shocks the oil price shock is the shock with the larger impact effect on the nominal price of oil. This restriction may be imposed as an additional inequality restriction enforcing that one response is larger than the other. The sign restrictions on the price level and real output are imposed not only on impact but for the first four quarters.

13.8.3 A Model of Unemployment and Vacancies

Fujita (2011) proposes a quarterly structural VAR model including the job separation rate, the job finding rate, and the number of vacancies in the United States. The number of vacancies is measured by the number of help-wanted advertisements in newspapers. All data are detrended to remove low-frequency variation not explained by economic theory. The model is partially identified in that only the responses to a shock in the profitability of the employment relationship is identified. A positive shock to profitability is assumed to increase vacancies on impact, resulting in a sign restriction on B_0^{-1} . A positive shock to profitability also is assumed to cause declines in unemployment for the first two quarters. Although changes in unemployment are not included in the reduced-form model, the unemployment response may be inferred from the responses of the transition rates via the implied gross job flows (for details see Fujita 2011). Thus, the second identifying assumption imposes nonlinear restrictions on the responses of the two transition rates. This very simple model is consistent with a wide range of Mortensen-Pissarides style search and matching models with and without endogenous job separation decisions.

13.8.4 An Extended Model of Unemployment and Vacancies

Fujita (2011) also considers an extended model designed to differentiate between two different sources of shocks to profitability. The model includes inflation and productivity growth in addition to job finding and job separation rates and vacancies. The model is again partially identified, but rather than identifying one shock to profitability, Fujita identifies separately a demand shock and a productivity shock. It is assumed that a positive demand shock raises the price level for the first four quarters, causes vacancies to rise on impact and the change in unemployment to be negative for the first two quarters. A positive technology shock is assumed to increase labor productivity for the first 20 quarters, to lower the price level for the first four quarters, and to raise unemployment. The change in unemployment again is defined on the basis of the responses of the rates of transition.

13.8.5 A Model of Technology Shocks

Dedola and Neri (2007) propose a quarterly structural VAR of the U.S. macroeconomy designed to study the responses of the economy to a positive technology shock. The variables in the model are the log of labor productivity, real wages, per capita hours worked, per capita real investment, per capita real consumption, the GDP deflator rate of inflation, and the short-term interest rate.

The sign restrictions are explicitly derived from a representative DSGE model. A positive technology shock increases labor productivity for the first 20 quarters, investment and output for the first 10 quarters, real wages for all quarters between the third and the twentieth quarter, and consumption for the first 5 quarters. The responses of hours, inflation, and the short-term interest rate are left unrestricted.

13.8.6 A Model of Exchange Rate Responses to Monetary Policy Shocks

Scholl and Uhlig (2008) propose a model of the response of the monthly nominal exchange rate to monetary policy shocks that is identified by static and dynamic sign restrictions as well as shape restrictions on selected impulse responses. They study the relationship between the United States and other economies, one economy at a time. The reduced-form is identical to the recursively identified model of Eichenbaum and Evans (1995). The model includes seven variables: U.S. and foreign industrial production, U.S. and foreign short-term interest rates, the U.S. price level, the ratio of nonborrowed reserves to total reserves in the United States, and the bilateral nominal dollar exchange rate in dollars per foreign currency. All variables but the interest rates are in logs.

Scholl and Uhlig (2008) consider two alternative sets of identifying assumptions. The first identification scheme postulates that an unanticipated monetary tightening in the U.S. lowers the price level and the ratio of nonborrowed to total reserves for the first year, while raising the interest rate for the first year.

The second identification scheme postulates that an unanticipated monetary tightening in the U.S. lowers the price level and the ratio of nonborrowed to total reserves for the first half year, while raising the interest rate for the first half year. It also adds the restrictions that the response of the U.S. interest rate exceeds the response of the foreign interest rate for the first half year. Finally, it imposes that the impact response of the nominal exchange rate is negative, and it imposes the absence of delayed overshooting in the exchange rate. The latter identifying assumption implies a shape restriction on the response function of the exchange rate that can be written as

$$s_0 < 0, |s_j| < |s_{j-1}|, |s_l| < |s_0| \text{ for } j=0, 1, 2 \text{ and } l=j+1, \dots, 23,$$

where s_j denotes the response of the exchange rate after j periods to an unexpected monetary policy tightening at date 0. Delayed overshooting refers to a situation in which a contractionary monetary policy shock causes a gradual appreciation of the exchange rate, followed by a gradual depreciation. In other words, the response reaches its maximum only with a delay. This pattern is inconsistent with traditional rational expectations open-economy sticky price models such as Dornbusch (1976) and is ruled out by the shape restriction above.

13.8.7 A Medium-Scale Macroeconomic Model

Canova and Paustian (2011) is an example of a medium-scale structural VAR model of the U.S. economy that simultaneously identifies many shocks. In Chapter 6 we already discussed in some detail the relationship between VAR models and DSGE models. Canova and Paustian first examine impulse responses for a wide range of medium-scale DSGE models. They document robust sign patterns for a range of structural shocks in these theoretical models and then proceed to impose these sign patterns as identifying assumptions on structural VAR models.

One implication of their analysis is that dynamic sign restrictions tend not to be robust across alternative theoretical models and hence should be avoided when fitting macroeconomic models. Another implication is that the choice of model variables matters. In VAR models with few variables it may not be possible to discriminate between the responses to alternative structural shocks because different structural shocks imply the same sign pattern in the responses of variables that are included in the VAR model. Usually, this problem may be overcome by judiciously adding variables, the responses of which help discriminate between alternative shocks.

The baseline model in Canova and Paustian (2011) includes five variables (nominal interest rate, real wage, inflation, real output, and hours worked) and identifies four structural shocks (markup shock, monetary shock, taste shock, technology shock) by restricting the impact multiplier matrix. If we postulate a flexible price, sticky wage model with measurement error added to the real wage, for example, the static identifying restrictions are

$$\begin{pmatrix} u_t^i \\ u_t^{\text{real wage}} \\ u_t^{\text{inflation}} \\ u_t^{\text{output}} \\ u_t^{\text{hours}} \end{pmatrix} = \begin{bmatrix} + & + & + & - & * \\ - & + & - & + & * \\ + & - & + & - & * \\ - & - & + & + & * \\ - & - & + & - & * \end{bmatrix} \begin{pmatrix} w_t^{\text{markup}} \\ w_t^{\text{monetary}} \\ w_t^{\text{taste}} \\ w_t^{\text{technology}} \\ w_t^{\text{measurement error}} \end{pmatrix}.$$

Using simplified versions of this model, Canova and Paustian (2011) demonstrate that the ability of sign-identified VAR models to recover the

structural responses implied by DSGE models improves when more shocks are identified (even if those shocks are not the shocks of interest) and when more variables are restricted for a given number of shocks. Especially the estimation of responses to monetary policy shocks requires many restrictions. Canova and Paustian also provide evidence that sign-identified VAR models of lower dimension may still be capable of recovering the DSGE population responses, as long as the omitted shocks do not exhibit the same sign patterns as the shocks to be identified.

This result addresses to some extent concerns in Fry and Pagan (2011) that the approach of deriving sign restrictions from DSGE models relies on the DSGE model having the same reduced-form VAR representation that is imposed in applied work. They observe that in practice omitting some of the variables contained in the underlying DSGE model may undermine the ability of sign-identified VAR models to recover the population responses. Fry and Pagan's conclusion is that caution is called for in imposing sign restrictions derived from DSGE models, unless there is a one-to-one mapping between the VAR model and the DSGE model. The results in Canova and Paustian (2011) suggest that sign-identified VAR models are more robust than exactly identified VAR models to model misspecification such as omitted variables.

13.8.8 A Model of Speculation in the Global Oil Market

Kilian and Murphy (2014) and Kilian and Lee (2014) propose a monthly model of the global oil market that allows for an explicit role of speculators in the physical oil market. In addition to the percent change in global oil production ($\Delta prod$), a business cycle measure of global real economic activity (rea), and the log of the real price of oil ($rpoil$), this model includes the change in above-ground global crude oil inventories ($\Delta Inventories$). Given the seasonality of the inventory data, the model is estimated using seasonal dummies (see Chapter 19). With the inclusion of the inventory data, it is no longer possible to defend a recursively identified model. Instead, the structural shocks are identified based on a combination of sign restrictions and bounds on the short-run price elasticities of oil demand and oil supply.

The key identifying assumptions are restrictions on the signs of the impact responses of the four observables to the structural shocks. First, an unanticipated disruption in the flow supply of oil ($w_t^{\text{flow supply}}$) causes oil production to fall, the real price of oil to increase, and global real activity to fall on impact. Second, an unanticipated increase in the flow demand for oil ($w_t^{\text{flow demand}}$), defined as an increase in oil demand for current consumption, causes global oil production, global real activity, and the real price of oil to increase on impact. Third, a positive speculative demand shock ($w_t^{\text{speculative demand}}$) is defined as an increase in inventory demand driven by, for example, a higher expected

real price of oil that is not already captured by flow demand or flow supply shocks. Such a shock in equilibrium causes an accumulation of oil inventories and raises the real price of oil. The accumulation of inventories requires oil production to increase and oil consumption to fall (associated with a fall in global real activity). Finally, the model also includes a residual shock (w_t^{residual}) designed to capture idiosyncratic shocks driven by a myriad of reasons that cannot be classified as one of the first three structural shocks. This shock is defined implicitly as the complement to the remaining shocks. These assumptions imply that

$$\begin{pmatrix} u_t^{\Delta \text{prod}} \\ u_t^{\text{rea}} \\ u_t^{\text{rpoil}} \\ u_t^{\Delta \text{Inventories}} \end{pmatrix} = \begin{bmatrix} - & + & + & * \\ - & + & - & * \\ + & + & + & * \\ * & * & + & * \end{bmatrix} \begin{pmatrix} w_t^{\text{flow supply}} \\ w_t^{\text{flow demand}} \\ w_t^{\text{speculative demand}} \\ w_t^{\text{residual}} \end{pmatrix}.$$

Note that the inventory responses to the flow supply and flow demand shocks are left unrestricted, but can be shown to be negative in the data, consistent with stabilizing inventory responses.

In addition to these static sign restrictions, the model imposes the additional restriction that the response of the real price of oil to a negative flow supply shock must be positive for at least twelve months, starting in the impact period. This restriction is necessary to rule out structural models in which unanticipated flow supply disruptions cause a decline in the real price of oil below its starting level. Such a decline would be at odds with conventional views of the effects of unanticipated oil supply disruptions. Because the positive response of the real price of oil tends to be accompanied by a persistently negative response of oil production, once we impose this additional dynamic sign restriction, it furthermore must be the case that global real activity responds negatively to oil supply shocks. This is the only way for the oil market to experience higher prices and lower quantities in practice, because in the data the decline of inventories triggered by an oil supply disruption is much smaller than the shortfall of oil production. This implies a joint set of dynamic sign restrictions such that the responses of oil production and global real activity to an unanticipated flow supply disruption are negative for the first twelve months, while the response of the real price of oil is positive.

Finally, the model imposes the restrictions that the impact price elasticity of oil supply is close to zero and that the impact price elasticity of oil demand cannot exceed the long-run price elasticity of oil demand, consistent with conventional views in the literature. A benchmark for that long-run elasticity is provided by studies of nonparametric gasoline demand functions based on U.S. household survey data such as Hausman and Newey (1995) which have consistently produced long-run price elasticity estimates near -0.8 . Their estimate suggests a bound of -0.8 on the impact price elasticity of demand. The

construction of this price elasticity of oil demand is complicated by the presence of oil inventories. For details the reader is referred to Kilian and Murphy (2014).

The model's focus on above-ground crude oil inventories is consistent with conventional accounts of speculation involving the accumulation of oil inventories in oil-importing economies. An alternative view is that speculation may also be conducted by oil producers who have the option of leaving oil below the ground in anticipation of rising prices (see Hamilton 2009). An accumulation of below-ground inventories by oil producers in anticipation of rising prices is equivalent to a reduction in flow supply. In short, flow supply shocks and speculative supply shocks are observationally equivalent in this model.

13.9 Mixing Sign and Exclusion Restrictions

It is possible to combine sign restrictions with other identifying restrictions such as short-run or long-run exclusion restrictions. Of course, mixing sign and exclusion restrictions further complicates inference. Little is known at this point about how to provide valid summary statistics for such models. Empirical studies have tended to rely on the posterior median response functions and posterior quantile bands.

One situation encountered in applied work is a mixture of sign restrictions and long-run exclusion restrictions. As in the case of models mixing short-run and long-run exclusion restrictions, special care must be exercised in specifying the reduced-form model and in imposing the long-run restrictions when constructing the initial recursive model, for which alternative rotations are considered. Fisher, Huh, and Pagan (2016) provide an illustrative example, building on the analysis in Peersman (2005). A more common situation in applied work is the combination of sign restrictions with short-run exclusion restrictions.

13.9.1 Examples of Models Mixing Sign and Short-Run Zero Restrictions

A Model of Fiscal Policy Shocks. Mountford and Uhlig (2009) propose a quarterly structural VAR model for the U.S. economy including ten variables designed to identify a generic business cycle shock, a monetary policy shock, a government spending shock, and a government revenue shock. The model includes real GDP, real consumption, total government expenditures, total government revenue, real wages, private nonresidential investment, the short-term interest rate, adjusted reserves, the producer price index for crude materials, and the GDP deflator. All components of national income are in per capita terms. All variables but the interest rate are in logs.

A positive business cycle shock is defined as a shock that increases output, consumption, investment, and government revenue for the first year following the shock. A positive monetary policy shock increases interest rates. It also lowers adjusted reserves and all prices for the first year. The monetary policy shock is defined to be orthogonal to the business cycle shock. A positive government spending shock increases government spending for one year after the shock. A positive unanticipated government revenue shock, in contrast, increases government revenue for one year after a positive shock. The two fiscal policy shocks are defined to be orthogonal to the business cycle shock and the monetary policy shock, but not necessarily orthogonal with respect to each other. Mountford and Uhlig (2009) also consider an alternative identification allowing for anticipated government revenue shocks. In that case, government revenue is restricted to rise only one year after the shock, which implies exclusion restrictions on the earlier responses, followed by a positive response in the second year.

In this model, no explicit sign restrictions are imposed on the responses of output, investment, and consumption. Because the model is estimated using a penalty function along the lines discussed earlier, however, as shown in Arias, Rubio-Ramírez, and Waggoner (2013), there are implicit sign restrictions on the responses of these variables to fiscal policy shocks. If the same model is estimated imposing only the sign restrictions discussed by Mountford and Uhlig and not using the penalty function approach, there is no support for Mountford and Uhlig's main conclusion that deficit-financed tax cuts are better for stimulating economic activity than deficit-financed increases in government spending. In contrast, when the VAR model is estimated imposing additional sign restrictions on the responses of GDP, consumption, and investment, the credible set narrows and qualitatively the same results as in the penalty function approach emerge.

A Model of Shocks to Optimism about the Economy. In related work, Beaudry, Nam, and Wang (2012) consider a structural VAR model including a suitably adjusted measure of total factor productivity, the stock price, real consumption, the real federal funds rate, and hours worked. The baseline structural model only imposes the restriction that positive shocks to optimism about the state of the U.S. economy have no contemporaneous effect on adjusted total factor productivity, but raise stock prices. Beaudry et al. estimate this model using a penalty function as in Mountford and Uhlig (2009). They show that a positive shock to optimism triggers an increase in hours worked and consumption. If correct, this result would seem to endorse the view that business cycles are driven by spells of optimism and pessimism.

As Arias, Rubio-Ramírez, and Waggoner (2013) show, however, the use of a penalty function in solving this model implies the additional restriction that

positive shocks to optimism generate an increase in consumption and hours worked. Thus, it is not the case that Beaudry et al.'s conclusion only relies on the identifying assumption that positive shocks to optimism have no contemporaneous effect on productivity, but raise stock prices. Upon estimating the model without the penalty function, these results disappear and the estimates become uninformative. They reappear only if the additional sign restrictions are explicitly imposed in estimation.

A Small Open Economy Model. A third example, motivated by the analysis in Mumtaz and Surico (2009), is a small open economy model. All global variables are marked with an asterisk. The model includes global real GDP growth (Δgdp^*), global inflation (Δp^*), global money growth (Δm^*), and the global interest rate (i^*). It also includes real GDP growth (Δgdp), inflation (Δp), and the short-term interest rate (i) in the U.K. domestic economy.

The model has an international block and a domestic block. The international block is ordered first and consists of a goods market and a money market. A positive global aggregate demand shock raises all global macroeconomic aggregates on impact. A positive global supply shock raises global real output, but lowers global inflation on impact. The responses of global money growth and the global interest rate remain unrestricted. A positive shock to global money demand causes a decline in inflation and real output on impact as well as an increase in the global interest rate and money growth. A positive shock to global money supply causes a decline in the global interest rate on impact as well as an increase in real output, global inflation, and global monetary aggregates.

The domestic block consists of the last three equations. The domestic monetary policy shock is identified recursively as in standard semistructural models of monetary policy. The other two domestic shocks are not separately identified. A key identifying assumption is that there is no immediate feedback from the three domestic shocks to any of the global macroeconomic aggregates, resulting in a block of exclusion restrictions in the upper right corner of B_0^{-1} :

$$\begin{pmatrix} u_t^{gdp,*} \\ u_t^{\Delta p,*} \\ u_t^{\Delta m,*} \\ u_t^{i,*} \\ u_t^{\Delta p} \\ u_t^{\Delta gdp} \\ u_t^i \end{pmatrix} = \begin{bmatrix} + & + & - & + & 0 & 0 & 0 \\ + & - & - & + & 0 & 0 & 0 \\ + & * & + & + & 0 & 0 & 0 \\ + & * & + & - & 0 & 0 & 0 \\ * & * & * & * & * & 0 & 0 \\ * & * & * & * & * & * & 0 \\ * & * & * & * & * & * & * \end{bmatrix} \begin{pmatrix} w_t^{\text{AD},*} \\ w_t^{\text{AS},*} \\ w_t^{\text{money demand,*}} \\ w_t^{\text{money supply,*}} \\ w_{5t} \\ w_{6t} \\ w_t^{\text{monetary policy, UK}} \end{pmatrix}. \quad (13.9.1)$$

13.9.2 How to Combine Sign Restrictions and Exclusion Restrictions

A number of algorithms have been proposed for combining sign restrictions with selected short-run or long-run exclusion restrictions in special cases. Early examples include Baumeister and Peersman (2013), Baumeister and Benati (2013), and Benati and Lubik (2014). Of special interest is the use of subrotations to generate partially restricted draws of B_0^{-1} from block recursive models. A more general algorithm for combining sign restrictions, short-run exclusion restrictions, and long-run exclusion restrictions (including linear restrictions on the structural parameters themselves and on Q) has been proposed by Arias, Rubio-Ramírez, and Waggoner (2013). Closely related work includes Binning (2013).

Subrotations for Preserving Blocks of Zero Restrictions on B_0^{-1} . Block recursive models such as the small open economy example above are often solved using subrotations. Consider the small open economy example. Partition the K model variables into K_1 global variables and K_2 domestic variables. Note that there is no contemporaneous feedback from the K_2 domestic variables to the K_1 global variables in (13.9.1). The purpose of using subrotations is to preserve these exclusion restrictions in generating draws for the rotation matrix Q . The algorithm involves four steps.

Algorithm (Subrotations)

1. Start with the $K \times K$ lower-triangular Cholesky decomposition P such that $PP' = \Sigma_u$.
2. Generate a $K_1 \times K_1$ dimensional subrotation \bar{Q} by drawing the columns of a $K_1 \times K_1$ matrix \bar{W} from the $\mathcal{N}(0, I_{K_1})$ distribution and apply the QR decomposition $\bar{W} = \bar{Q}\bar{R}$.
3. Form Q by placing \bar{Q} in the upper left corner of the identity matrix I_K :

$$Q = \begin{bmatrix} \bar{Q} & 0 \\ 0 & I_{K_2} \end{bmatrix}.$$

4. Then a draw for B_0^{-1} consists of PQ . □

A numerical example for $K_1 = 3$ and $K_2 = 3$ illustrates this procedure. Let

$$P = \begin{bmatrix} 2.2295 & 0 & 0 & 0 & 0 & 0 \\ -2.3258 & 12.5955 & 0 & 0 & 0 & 0 \\ 0.1360 & 0.0746 & 0.8563 & 0 & 0 & 0 \\ -0.0383 & 0.1642 & 0.0222 & 0.3596 & 0 & 0 \\ -0.0019 & 0.0871 & 0.2364 & 0.1466 & 0.6272 & 0 \\ 0.0240 & 0.1288 & 0.4134 & -0.0344 & 0.5960 & 2.4835 \end{bmatrix}$$

and

$$\bar{W} = \begin{bmatrix} -0.7937 & 2.1457 & 0.0984 \\ -0.4165 & -0.7860 & -0.6739 \\ -0.7826 & 0.4481 & 0.2004 \end{bmatrix}.$$

Then

$$\bar{Q} = \begin{bmatrix} -0.6670 & 0.6467 & -0.3699 \\ -0.3500 & -0.7104 & -0.6106 \\ -0.6577 & -0.2778 & 0.7002 \end{bmatrix}.$$

This allows the construction of Q and hence

$$PQ = \begin{bmatrix} -1.4872 & 1.4418 & -0.8248 & 0 & 0 & 0 \\ -2.8575 & -10.4513 & -6.8309 & 0 & 0 & 0 \\ -0.6800 & -0.2030 & 0.5037 & 0 & 0 & 0 \\ -0.0465 & -0.1476 & -0.0705 & 0.3596 & 0 & 0 \\ -0.1847 & -0.1288 & 0.1130 & 0.1466 & 0.6272 & 0 \\ -0.3329 & -0.1908 & 0.2020 & -0.0344 & 0.5960 & 2.4835 \end{bmatrix}.$$

Partitioning the matrix $P = [P_1, P_2]$, where P_i is $K \times K_i$, $i = 1, 2$, we have $PQ = [P_1 \bar{Q}, P_2]$. Thus, the elements of the $K_2 \times K_2$ block on the lower right of PQ remain unaffected by the rotations because they only depend on domestic coefficients that are exactly identified. The algorithm also preserves the $K_1 \times K_2$ block of zeros on the upper right that reflects the block recursive structure of the structural impact multiplier matrix. In contrast, the $K_2 \times K_1$ block on the lower left of PQ that relates to the global variables changes across rotations, as does the $K_1 \times K_1$ block on the upper left of PQ .

A More General Algorithm. One important limitation of the use of subrotations is that this approach relies on a block-recursive structure of the structural model. This restriction can be relaxed, as shown in Arias, Rubio-Ramírez, and Waggoner (2013). Express the stationary, K -dimensional structural VAR model as

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t. \quad (13.9.2)$$

The reduced-form representation of this model is

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where $A_i = B_0^{-1} B_i$, $i = 1, \dots, p$, $u_t = B_0^{-1} w_t$, and $\mathbb{E}(u_t u_t') = \Sigma_u = B_0^{-1} B_0^{-1'}$. Thus, the reduced form parameters are $A \equiv [A_1, \dots, A_p]$ and Σ_u .

The impulse response function for the i^{th} variable with respect to the j^{th} structural shock at a finite horizon h corresponds to the element in row i and column j of

$$\Theta_h = (J \mathbf{A}^h J') B_0^{-1},$$

where \mathbf{A} denotes the companion matrix of the reduced-form model,

$$\mathbf{A} \equiv \begin{bmatrix} A_1 & A_2 & \cdots & A_{p-1} & A_p \\ I_K & 0 & \cdots & 0 & 0 \\ 0 & I_K & & 0 & 0 \\ \vdots & & \ddots & \vdots & \vdots \\ 0 & 0 & \cdots & I_K & 0 \end{bmatrix}$$

and $J \equiv [I_K, 0_{K \times K(p-1)}]$. When imposing restrictions on the infinite horizon, it is assumed that the i^{th} variable is expressed in first differences. The long-run structural impulse response function corresponds to the element in row i and column j of the matrix

$$\Theta_\infty \equiv \left(I_K - \sum_{i=1}^p A_i \right)^{-1} B_0^{-1}.$$

Candidate draws for Θ_∞ and Θ_h , denoted L_∞ and L_h , respectively, are constructed from the reduced-form estimates, given an initial guess of the structural impact multiplier matrix B_0^{-1} of $L_0 = \text{chol}(\widehat{\Sigma}_u)$. It is convenient to stack the structural impulse response matrices to be restricted into a single matrix denoted by \mathbf{L} . For example, if the sign restrictions are imposed at horizons zero, two, and infinity, then

$$\mathbf{L} = \begin{bmatrix} L_0 \\ L_2 \\ L_\infty \end{bmatrix}$$

has K columns corresponding to the K structural shocks and $3K$ rows because we consider restrictions at three horizons for K variables.

Sign restrictions on these structural impulse response functions can be represented by matrices S_j for $j = 1, \dots, K$, where the number of columns in S_j equals the number of rows in \mathbf{L} . Usually S_j will be a selection matrix with one non-zero entry in each row. Let $s_j = \text{rank}(S_j)$. Then s_j is the number of sign restrictions on the impulse response functions associated with the j^{th} structural shock. The total number of sign restrictions is $s = \sum_{j=1}^K s_j$. Let ι_j denote the j^{th} column of the identity matrix I_K . Then the structural parameters satisfy the sign restrictions if and only if $S_j \mathbf{L} \iota_j > 0$ for $j = 1, \dots, K$. As before, let Q denote the solution to the QR decomposition of a $K \times K$ matrix whose columns are independent random draws from $\mathcal{N}(0, I_K)$.

The challenge is to convert draws from the posterior of the reduced-form parameters, together with a draw for Q from the uniform distribution over the space of orthogonal matrices $\mathcal{O}(K)$, to a draw from the posterior distribution of the impulse responses before any identifying restrictions are imposed. The

following algorithm from Arias, Rubio-Ramírez, and Waggoner (2013) can be used for that purpose.

Algorithm (*Sign Restrictions*)

1. Draw (A, Σ_u) from the posterior distribution of the reduced-form parameters.
2. Draw an orthogonal matrix Q from $\mathcal{O}(K)$.
3. Then $L_0 Q$ will be a draw from the posterior distribution of B_0^{-1} before any identifying restrictions are imposed on the implied structural impulse responses.
4. Recompute \mathbf{L} with $L_0 Q$ replacing L_0 . Retain the draw if $S_j \mathbf{L} \boldsymbol{\iota}_j > 0$ is satisfied for $j = 1, \dots, K$.
5. Having repeated steps 2, 3, and 4 as often as desired, return to step 1 until the desired number of draws from the posterior of the structural impulse responses conditional on the sign restrictions has been obtained. \square

This algorithm has to be modified to allow for additional exclusion restrictions on B_0^{-1} . Because the set of structural parameters conditional on zero restrictions has probability measure zero in the space of all structural parameters generated by the algorithm of Rubio-Ramírez, Waggoner, and Zha (2010), it is necessary to generate draws from the posterior of the structural impulse responses conditional on the zero restrictions already having been imposed, as outlined in Arias, Rubio-Ramírez, and Waggoner (2013).

Similar to the case of sign restrictions, we stack the structural impulse response functions at the horizons of interest. \mathbf{L} now contains impulse response functions subject to both sign and exclusion restrictions. Zero restrictions are represented by matrices E_j for $j = 1, \dots, K$. The number of columns in E_j is equal to the number of rows in \mathbf{L} . If the rank of E_j is e_j , then e_j is the number of zero restrictions associated with the j^{th} shock. The total number of zero restrictions is $e = \sum_{j=1}^K e_j$. The structural parameters satisfy the zero restrictions if and only if $E_j \mathbf{L} \boldsymbol{\iota}_j = 0$ for $1 \leq j \leq K$. In what follows, let E_j represent the zero restrictions with the equations of model (13.9.2) ordered such that $e_j \leq K - j$ for $j = 1, \dots, K$.

Algorithm (*Combining Sign Restrictions and Exclusion Restrictions*)

1. Draw (A, Σ_u) from the posterior distribution of the reduced-form parameters.
2. Draw an orthogonal matrix Q such that $L_0 Q$ satisfies the exclusion restrictions.
3. Recompute \mathbf{L} with $L_0 Q$ replacing L_0 . Retain the draw if $S_j \mathbf{L} \boldsymbol{\iota}_j > 0$ is satisfied for $j = 1, \dots, K$.

4. Having repeated steps 2 and 3 as often as is desired, return to step 1 until the desired number of draws from the posterior of the structural impulse responses conditional on the sign restrictions has been obtained. \square

The modification in the algorithm involves step 2. Let X be a $K \times K$ matrix of independent $\mathcal{N}(0, 1)$ draws and choose Q by a QR decomposition of X . Define $q_j \equiv Q\iota_j$ and $x_j \equiv X\iota_j$ for $1 \leq j \leq K$, where ι_j denotes the j^{th} column of the identity matrix I_K . Moreover, define $R_1 = E_1\mathbf{L}$ and, for $j = 2, \dots, K$,

$$R_j = \begin{bmatrix} E_j\mathbf{L} \\ Q_{j-1} \end{bmatrix},$$

where \mathbf{L} denotes the matrix of structural impulse responses to be restricted, E_j is the corresponding matrix of zero restrictions, and $Q_{j-1} = [q_1, \dots, q_{j-1}]$. Then drawing Q involves the following recursive subroutine:

- a. Let $j = 1$.
- b. Construct R_j and find the matrix N_{j-1} whose columns form an orthonormal basis for the null space of R_j .⁶
- c. Draw the $K \times 1$ vector x_j from the $\mathcal{N}(0, I_K)$ distribution.
- d. Let $q_j = N_{j-1} \left(N'_{j-1} x_j / \|N'_{j-1} x_j\| \right)$, where $\|\cdot\|$ is the Euclidian norm.⁷
- e. If $j = K$, stop, construct $Q = [q_1, \dots, q_K]$ and move to step 3 of the modified algorithm. Otherwise, let $j = j + 1$ and move to step b.

As Arias, Rubio-Ramírez, and Waggoner (2013) emphasize, when implementing this modified algorithm, it is critical that one exits upon finding a q_i that violates the sign restrictions, and resumes with a new draw from the reduced-form posterior in step 1. Although it is not permissible to draw orthogonal matrices until acceptance, it is permissible to draw a fixed number of orthogonal matrices Q for each reduced-form draw and to retain all rotations that satisfy the sign restrictions.

A Numerical Example. The following numerical VAR example from Arias, Rubio-Ramírez, and Waggoner (2013) illustrates the implementation of the modified algorithm. Let $K = 4$ and $p = 1$. Let A and Σ_u be a particular draw from the posterior of the reduced-form parameters of a four-dimensional

⁶ An orthonormal basis for the null space of the matrix R_j may be obtained from a singular value decomposition applying the MATLAB function `null` to R_j , for example.

⁷ The Euclidian norm of a vector $x = (x_1, \dots, x_n)'$ is $\sqrt{x'x}$ which can be computed using the `norm` function in MATLAB.

VAR(1):

$$A_1 = \begin{bmatrix} 0.7577 & 0.7060 & 0.8325 & 0.4387 \\ 0.7431 & 0.0318 & 0.6948 & 0.3816 \\ 0.3922 & 0.2769 & 0.3171 & 0.7655 \\ 0.6555 & 0.0462 & 0.9502 & 0.7952 \end{bmatrix},$$

$$\Sigma_u = \begin{bmatrix} 0.0281 & -0.0295 & 0.0029 & 0.0029 \\ -0.0295 & 3.1850 & 0.0325 & -0.0105 \\ 0.0029 & 0.0325 & 0.0067 & 0.0054 \\ 0.0029 & -0.0105 & 0.0054 & 0.1471 \end{bmatrix}.$$

Suppose that we want to impose restrictions on the structural impulse response functions at horizons 0, 2, and ∞ . First, without loss of generality, define an initial solution for the structural impulse response functions of interest from the reduced-form parameters as

$$L_0 = \text{chol}(\widehat{\Sigma}_u),$$

$$L_2 = (J\mathbf{A}^2 J') L_0 = A_1^2 L_0,$$

$$L_\infty = (I_4 - A_1)^{-1} L_0,$$

where $\mathbf{A} = A_1$ because we are considering a VAR(1). Hence,

$$\mathbf{L} = \begin{bmatrix} L_0 \\ L_2 \\ L_\infty \end{bmatrix} = \begin{bmatrix} 0.1676 & 0 & 0 & 0 \\ -0.1760 & 1.7760 & 0 & 0 \\ 0.0173 & 0.0200 & 0.0775 & 0 \\ 0.0173 & -0.0042 & 0.0669 & 0.3772 \\ 0.1355 & 1.9867 & 0.1828 & 0.5375 \\ 0.0259 & 1.3115 & 0.0828 & 0.2882 \\ 0.1377 & 2.1813 & 0.2131 & 0.6144 \\ 0.1069 & 2.0996 & 0.1989 & 0.6281 \\ 0.1091 & -0.3783 & -0.0847 & -0.2523 \\ -0.1170 & 1.2928 & -0.0599 & -0.2201 \\ -0.0422 & -0.7342 & 0.0006 & -0.1695 \\ -0.0575 & -1.1662 & 0.0362 & 0.2577 \end{bmatrix}.$$

The identifying restrictions to be imposed on these structural responses are as follows. The response of the third variable to the second structural shock at horizon 2 is negative; the response of the fourth variable to the second structural shock is positive at horizon 2. The response of the second variable to the third structural shock is negative at horizon 0; and the response of the first variable to the fourth structural shock is positive at horizons 0, 2, and ∞ . The responses of the first and third variables to the first shock are zero at horizon

0; and the response of the fourth variable to the second structural shock is zero at horizon ∞ . These restrictions imply that

$$S_2 = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$S_3 = [0 \ -1 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0],$$

$$S_4 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \end{bmatrix},$$

$$E_1 = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

and

$$E_2 = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1].$$

There are no sign restrictions associated with the first structural shock, so we do not need to specify S_1 , and there are no exclusion restrictions associated with the third and fourth structural shock, so there is no need to specify E_3 and E_4 .

To find a rotation matrix Q that satisfies both the sign and zero restrictions, we apply the subroutine described above:

a. Let $j = 1$.

b. Then $R_1 = \begin{bmatrix} 0.1676 & 0 & 0 & 0 \\ 0.0173 & 0.0200 & 0.0775 & 0 \end{bmatrix}$ and $N_0 = \begin{bmatrix} 0 & 0 \\ -0.0982 & 0 \\ 0.2502 & 0 \\ 0 & 1 \end{bmatrix}$.

c. Suppose the draw of the vector x_1 from the $\mathcal{N}(0, I_4)$ distribution is

$$x_1 = [0.4395 \ -0.1190 \ -0.9354 \ 0.0464]'$$

d. Then $q_1 = N_0 (N_0' x_1 / \|N_0' x_1\|) = [0 \ 0.9018 \ -0.2330 \ 0.3638]'$.

e. Let $j = 2$ and move to step b.

By repeating this routine for $j = 2, \dots, 4$, given the random draws

$$x_2 = \begin{bmatrix} -0.6711 \\ 1.5332 \\ -0.1836 \\ 0.3509 \end{bmatrix}, \quad x_3 = \begin{bmatrix} -0.5941 \\ 0.5901 \\ -1.4499 \\ -0.2632 \end{bmatrix}, \quad x_4 = \begin{bmatrix} 0.6713 \\ -0.4112 \\ 0.7989 \\ -0.0868 \end{bmatrix},$$

we obtain the following solutions:

$$N_1 = \begin{bmatrix} -0.2648 & 0.9609 \\ 0.1095 & -0.0053 \\ 0.9068 & 0.2271 \\ 0.3093 & 0.1586 \end{bmatrix}, N_2 = \begin{bmatrix} 0.1704 & -0.0322 \\ 0.2180 & -0.3697 \\ 0.9582 & 0.0188 \\ 0.0733 & 0.9284 \end{bmatrix},$$

$$N_3 = \begin{bmatrix} -0.0854 \\ -0.4203 \\ -0.2913 \\ 0.8551 \end{bmatrix}.$$

Upon completing this subroutine, we obtain the following solution for the restricted rotation matrix:

$$Q = [q_1, \dots, q_K] = \begin{bmatrix} 0 & -0.9849 & -0.1509 & 0.0854 \\ 0.9018 & 0.0498 & -0.0871 & 0.4203 \\ -0.2330 & 0.1651 & -0.9130 & 0.2913 \\ 0.3638 & -0.0177 & -0.3189 & -0.8551 \end{bmatrix}.$$

The implied posterior draw of the structural impact multiplier matrix is

$$L_0 Q = \begin{bmatrix} 0 & -0.1651 & -0.0253 & 0.0143 \\ 1.6016 & 0.2617 & -0.1281 & 0.7313 \\ 0 & -0.0033 & -0.0751 & 0.0325 \\ 0.1179 & -0.0129 & -0.2025 & -0.3034 \end{bmatrix}.$$

The restrictions may be verified using $S_j \mathbf{L} q_j$ and $E_j \mathbf{L} q_j$, where \mathbf{L} is evaluated at $L_0 Q$.

Caveats. Although the algorithm of Arias, Rubio-Ramírez, and Waggoner (2013) represents an important step forward in implementing mixed restrictions, several open questions remain. One limitation of this algorithm is that we can impose at most $K - j$ zero restrictions for each shock j , where $j = 1, \dots, K$. In addition, this approach cannot be directly extended to imposing nonzero restrictions on specific elements of B_0 . In contrast, additional inequality restrictions on structural impulse responses could easily be imposed in step 3 by generating draws for the structural impulse responses in question and dropping draws that do not satisfy these additional restrictions. One example of such additional restrictions would be the restriction that the impact response of one variable should be larger than the impact response of another variable to the same structural shock, as in Peersman (2005). Another example are bounds on impact elasticity parameters as in Kilian and Murphy (2014). A third example is the use of shape restrictions as in Scholl and Uhlig (2008).

A second limitation relates to the evaluation of the set of admissible models. Arias, Rubio-Ramírez, and Waggoner (2013) rely on posterior median

response functions and quantile bands. This approach suffers from the shortcomings already discussed earlier. It therefore would be useful to generalize the approach of constructing the most likely structural model, as discussed in Section 13.6.2, to the framework of Arias et al. and to derive the analogous expression for the posterior density of the structural impulse responses. Such an extension of the analysis in Inoue and Kilian (2013) should be straightforward in principle.

13.9.3 Discussion

As in the case of models identified only by sign restrictions, the modified algorithms discussed in this section will in general be informative for the structural impulse responses. Four proposals have been made to address this concern. One proposal has been to adapt the procedure of Giacomini and Kitagawa (2015) that is robust to the choice of the rotation matrix.

An alternative proposal has been to focus on priors that are conditionally agnostic. Arias, Rubio-Ramírez, and Waggoner (2015) observe that if a researcher begins with an agnostic prior over a particular parameterization of the model, as defined earlier, and then conditions on the zero restrictions, the resulting prior will be conditionally agnostic over that parameterization of the model. Unlike in the case of imposing sign restrictions only, however, one has to choose which parameterization of the model one wishes to be conditionally agnostic about. Being conditionally agnostic under one parameterization of the structural model does not imply being conditionally agnostic under a different parameterization. Arias et al. propose using an importance sampler to draw from the conditionally agnostic prior for a given parameterization of the structural VAR model. They also show how to construct conditionally flat priors by choosing an appropriate conjugate prior over the orthogonal reduced-form representation, generalizing the approach discussed earlier for the case of sign restrictions only. The choice of this conjugate prior depends on whether one wishes to be flat across the structural impulse responses that satisfy the zero restrictions or flat across the structural model representations that satisfy the zero restrictions.

Yet another proposal has been to embrace the fact that priors for structural impulse responses are informative and to make that information explicit, possibly drawing on extraneous economic information. Plagborg-Møller (2016) suggests such an algorithm that facilitates the imposition of zero restrictions on elements of B_0^{-1} . A complementary proposal has been to specify the prior on the elements of B_0 rather than B_0^{-1} , taking the structural VAR model representation as the object of primary interest (see Baumeister and Hamilton 2015a, 2015b). This procedure also allows us to interpret exclusion restrictions as degenerate priors.

13.10 Empirical Illustrations

We conclude this chapter with two empirical illustrations of the use of sign-identified VAR models.

13.10.1 A Model of the Global Oil Market

The first example is based on the analysis in Inoue and Kilian (2013). Inoue and Kilian adapt the monthly recursively identified structural VAR model of the global oil market in Kilian (2009) for the use of sign restrictions, building on Kilian and Murphy (2012). A negative flow supply shock initially lowers global oil production and global real economic activity, but raises the real price of oil. A positive flow demand shock initially raises oil production, real economic activity, and the real price of oil. Finally, other oil demand shocks (such as precautionary demand shocks) stimulate global oil production and the real price of oil, but lower real economic activity on impact. Formally:

$$\begin{pmatrix} u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} - & + & + \\ - & + & - \\ + & + & + \end{bmatrix} \begin{pmatrix} w_t^{\text{flow supply}} \\ w_t^{\text{flow demand}} \\ w_t^{\text{other demand}} \end{pmatrix}.$$

In addition, the model numerically bounds the price elasticity of oil supply by 0.025, building on Kilian and Murphy (2012).

This elasticity corresponds to the ratio of the oil production response in the impact period triggered by an exogenous demand shock to the price response in the impact period triggered by the same shock. This restriction rules out models with implausibly high price elasticities of oil supply compared with conventional wisdom and extraneous empirical evidence. Finally, Inoue and Kilian (2013) restrict the real price of oil to be positive for the first year in response to positive demand and negative supply shocks, following Baumeister and Peersman (2013).

Following Inoue and Kilian (2013) and related studies in the literature, we specify a VAR(24) model with intercept. The model is estimated on monthly data for 1973m2-2008m9 using a diffuse Gaussian-inverse Wishart prior. Figure 13.7 plots the structural responses. The responses have been normalized such that each structural shock implies an increase in the real price of oil. The response of oil production is obtained by cumulating the responses of its growth rate.

All structural response function estimates are consistent with standard economic intuition. For example, a negative flow supply shock is associated with a persistent decline in oil production, a modest increase in the real price of oil, and a short-lived decline in global real economic activity. A positive flow demand shock is associated with a persistent and hump-shaped response

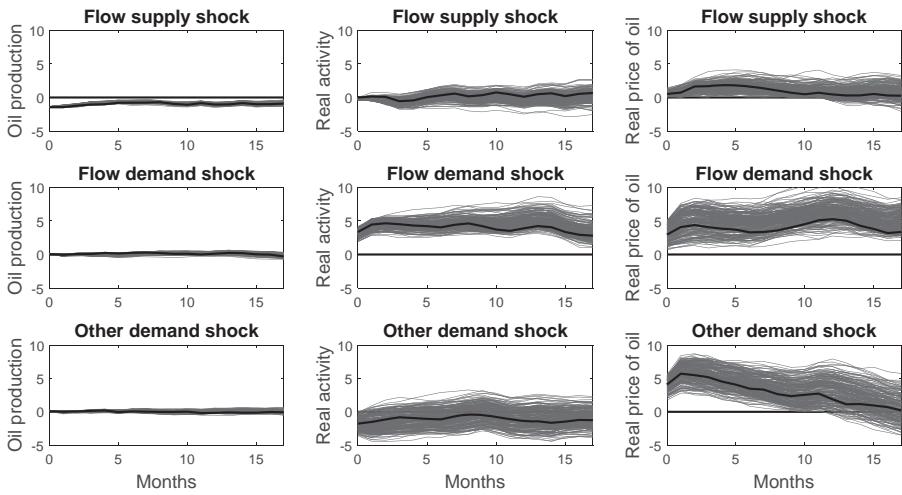


Figure 13.7. Sign-identified oil market model impulse response functions in the modal model and 68% joint HPD regions.

in both global real activity and the real price of oil and with little response in global crude oil production. Other demand shocks (such as shocks to oil inventory demand) cause a temporary increase in the real price of oil, a persistent decline in global real economic activity and little response in global crude oil production. The corresponding credible sets indicate considerable uncertainty about the price responses and to a lesser extent for the responses in real activity, whereas the credible sets for oil production responses are quite narrow. Nevertheless, several response functions are precisely enough estimated to conclude that the response differs from zero. Figure 13.7 also illustrates that the responses of the most likely model need not be near the center of the credible set.

There are also important differences between the most likely estimates provided by the modal model and the conventional median response functions. Median response functions may be closer to zero or further away from zero than the responses of the modal model. For example, median response functions can be shown to overestimate the magnitude of the price response to flow demand shocks, as illustrated in the left panel of Figure 13.8. Moreover, pointwise 68% posterior error bands provide little protection against mischaracterizing the impulse response dynamics, as shown in the right panel of Figure 13.8. At several horizons, the response functions of the modal model are outside the pointwise error bands. The right panel of Figure 13.8 also illustrates that pointwise intervals tend to misrepresent the estimation and identification uncertainty compared with the credible set shown in Figure 13.7 that captures the joint uncertainty over all impulse responses. This example illustrates that

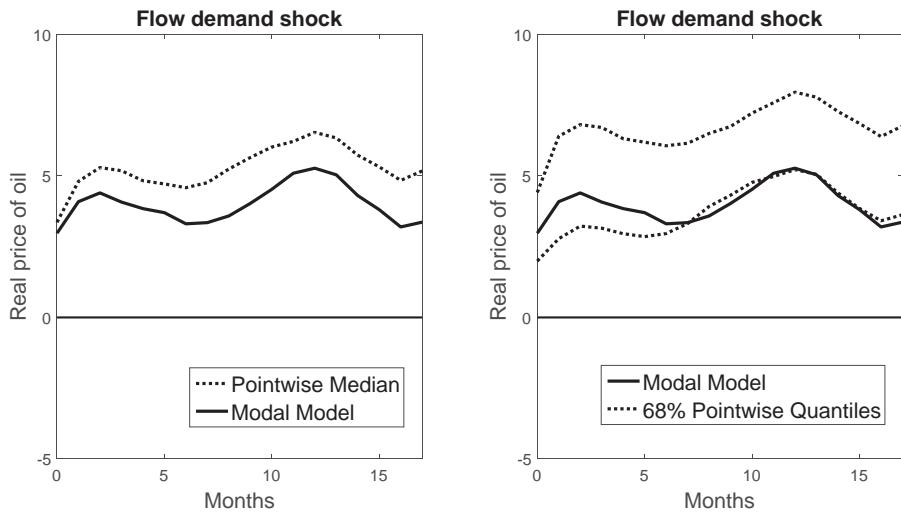


Figure 13.8. Structural impulse responses in the sign-identified oil market model.

the way estimates of sign-identified VAR models are represented matters for the interpretation of the estimated model.

13.10.2 A Model of Monetary Policy

Whereas the first example involved a fully identified model based on sign restrictions, the second example based on Uhlig (2005) involves a sign-identified model that is only partially identified. The set of variables consists of monthly U.S. data for the log of interpolated real GDP (gdp_t) and of its deflator ($defl_t$), the log of a commodity price index ($pcom_t$), total reserves (tr_t), nonborrowed reserves (nbr_t), and the federal funds rate (i_t). The identification is deliberately agnostic. An unanticipated monetary policy tightening is assumed to cause an increase in the interest rate and to lower the GDP price deflator, the commodity price index and nonborrowed reserves all for the first six months, including the impact period:

$$\begin{pmatrix} u_t^{gdp} \\ u_t^{defl} \\ u_t^{pcom} \\ u_t^{tr} \\ u_t^{nbr} \\ u_t^i \end{pmatrix} = \begin{bmatrix} * & * & * & * & * & * \\ - & * & * & * & * & * \\ - & * & * & * & * & * \\ * & * & * & * & * & * \\ - & * & * & * & * & * \\ + & * & * & * & * & * \end{bmatrix} \begin{pmatrix} w_t^{\text{monetary policy}} \\ w_{2t} \\ w_{3t} \\ w_{4t} \\ w_{5t} \\ w_{6t} \end{pmatrix},$$

where * denotes unrestricted elements.

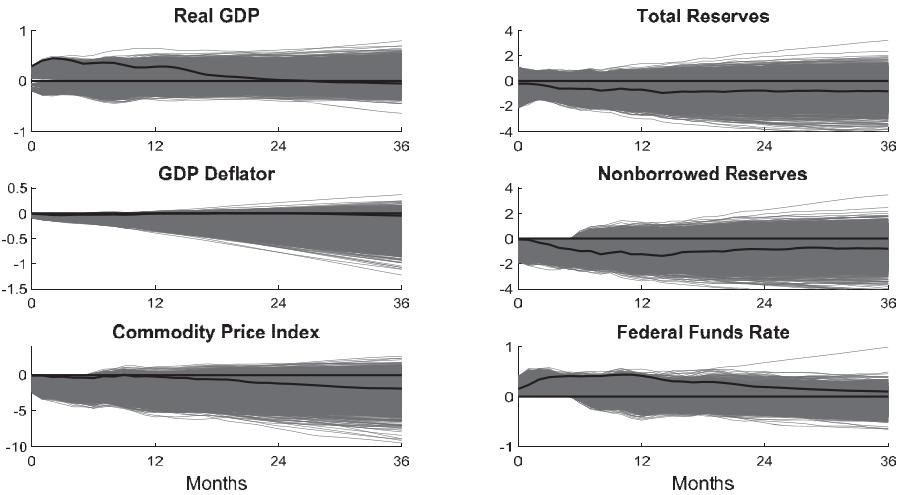


Figure 13.9. Responses to a monetary policy tightening in the original sign-identified model: Response functions in the modal model and 68% joint HPD regions.
Source: Inoue and Kilian (2013).

The sample period is 1965m1–2003m12 to ensure compatibility with Uhlig's original analysis. The VAR(12) model without intercept is estimated using a diffuse Gaussian-inverse Wishart prior on the same data as in Uhlig (2005). We evaluate the posterior of the model as in Inoue and Kilian (2013). The numerical stability of the results requires a fairly large number of draws, especially for M and M^\dagger , where M^\dagger is the number of draws used to approximate the marginalized posterior distribution. The posterior distribution of the impulse response estimates is based on $M = 5,000$ draws from the reduced-form posterior distribution with $N = 500$ rotations each. We set $M^\dagger = 20,000$. Figure 13.9 summarizes the responses of the variables of interest to an unexpected monetary policy tightening.

Figure 13.9 illustrates that there are important differences between the median response estimates of the response of real output and the response in the modal model. Whereas Uhlig reports a peak median output response of 0.15 percentage points, for the same data, the modal model implies a peak response of almost 0.5 percentage points. Moreover, that peak value is near the upper end of the credible set and outside the conventional pointwise posterior error band. It should be noted that both the median estimate and the response estimate based on the modal model are counterintuitive in that a monetary tightening would be expected to cause a decline in real output over time rather than an increase. This outcome reflects the fact that the identifying assumptions are not overly informative. Even in Uhlig's original analysis, there was substantial pointwise probability mass on both negative and positive responses

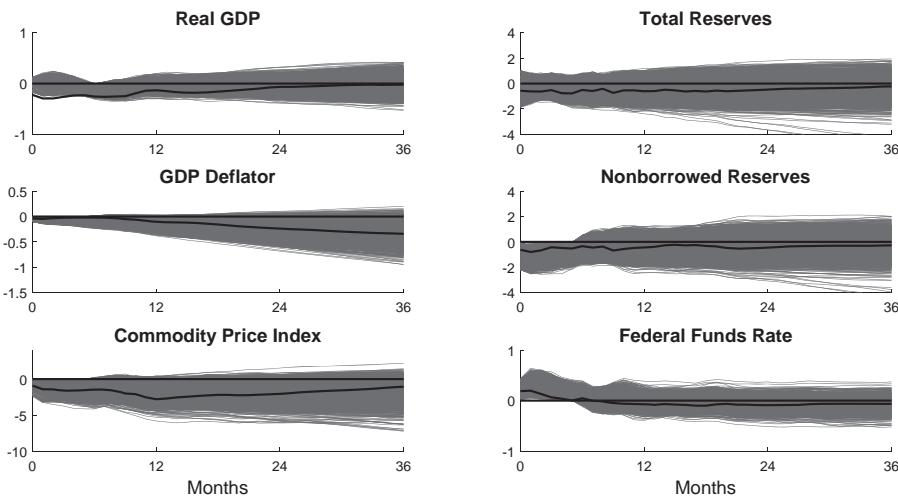


Figure 13.10. Responses to a monetary policy tightening in the modified sign-identified model: Response functions in the modal model and 68% joint regions of highest posterior density.

Source: Inoue and Kilian (2013).

of real output. The 68% credible set further widens the set of probable response functions.

The explicit reason why Uhlig (2005) did not impose further restrictions is that he wished to be as agnostic as possible about the response of real output. This approach is appropriate only to the extent that we view models in which real output increases in response to a monetary tightening as economically plausible a priori (see Kilian and Murphy 2012). Many economists would disagree with this view at least for intermediate horizons. Hence, in Figure 13.10 we consider an alternative set of results for models that impose an additional sign restriction on the response of real GDP to an unexpected monetary policy tightening after 6 months (and only at that horizon):

$$\frac{\partial gdp_{t+6}}{\partial w_t^{\text{monetary policy}}} < 0.$$

This identifying assumption leaves the short-run as well as the longer-run response of real output unrestricted, preserving the spirit of Uhlig's original exercise.

The resulting modal model produces substantially different and more economically plausible results, including a cumulative drop in real GDP of -0.3 percentage points in the second quarter. The response estimate for the modal model is at the lower end of the credible set and again outside the conventional pointwise posterior error band. It also is substantially different from the response estimate obtained from the traditional Cholesky decomposition (see

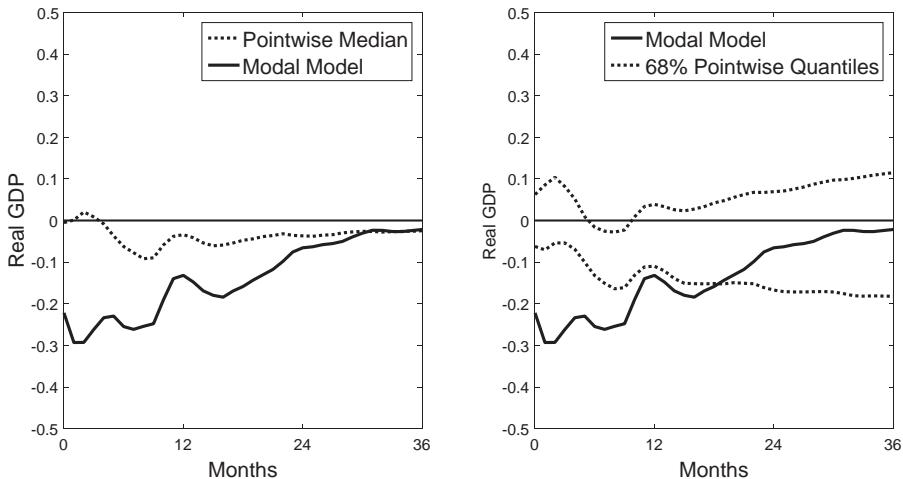


Figure 13.11. Responses to a monetary policy tightening in the modified sign-identified model.

Source: Inoue and Kilian (2013).

Chapter 12). One difference is that the reduction in real GDP in Figure 13.10 is temporary, whereas traditional Cholesky models imply a much more persistent decline in real GDP. Even in this alternative model, however, the 68% credible set includes many positive real output responses, suggesting that the data are not informative about the response of real output. Likewise the other response functions are estimated only very imprecisely. We conclude that there remains substantial uncertainty about the effects of monetary policy shocks on real output.

Figure 13.11 elaborates further on the results in Figure 13.10. The left panel again illustrates that there can be substantial differences between the median response function estimates and the response function estimates based on the modal model. For example, the decline in real GDP caused by an unanticipated monetary contraction is much larger in the modal model, at least in the short-run. In some cases, the median and the modal response of real GDP differ not only in magnitude but in sign. The right panel demonstrates that the modal model responses may be outside the conventional pointwise 68% error bands. This is true in particular for the response of real GDP and to a lesser extent for the response of commodity prices and the own-response of the federal funds rate.

13.11 Concluding Remarks

The advantage of sign-identified models is that sign restrictions are much more easily derived from economic models than alternative identifying restrictions.

At the same time, working with set-identified models poses special challenges. Of particular concern is the possibility that the conventional priors used in estimating sign-identified models may unduly influence the posterior of the structural impulse responses.

Making explicit prior information about the parameters of the structural VAR representation, as proposed by Baumeister and Hamilton (2015a, 2015b, 2015c) is one possible response to this concern. This approach also provides an opportunity for the researcher to bring to bear additional information that otherwise could not have been imposed. However, it is not always clear how to derive within this framework economically motivated priors that have broad appeal. Moreover, this approach typically does not allow the user to replicate the identifying restrictions used in other studies. For example, the methodology proposed by Baumeister and Hamilton (2015a) does not allow for dynamic sign restrictions, shape restrictions, and cross-equation restrictions. Even more importantly, the posterior draws generated by the approach of Baumeister and Hamilton (2015a) currently can only be evaluated under restrictive loss functions that postulate that the user is not concerned with the shapes of the impulse response functions. One way of addressing the latter problem would be to adapt the methodology of Inoue and Kilian (2013) to this new setting, but no such results exist at this point.

Another possible solution to the problem of unintentionally informative priors in sign-identified structural VAR models has been proposed by Plagborg-Møller (2016). His proposal is to estimate the structural moving average coefficients directly by Bayesian methods, allowing one to explicitly specify the prior densities for the structural impulse response parameters not only on impact but also at longer horizons. This approach avoids some of the drawbacks of the proposal in Baumeister and Hamilton (2015a). It not only follows the existing literature on sign-identified VAR models in specifying the prior on the structural impact multiplier matrix, but it also facilitates the imposition of restrictions on the shape and smoothness of the structural impulse response functions. In this sense Plagborg-Møller's approach may be viewed as the complement to Baumeister and Hamilton's proposal to impose priors on the parameters of the structural VAR model. In both cases, the central idea is to replace implicitly informative priors by explicitly informative priors. The main drawback of Plagborg-Møller's approach is that the resulting structural impulse response estimates are likely to be sensitive to the choice of the prior in finite samples and asymptotically. Moreover, as in Baumeister and Hamilton's work, it is rarely clear what independent information the specification of the prior density can be based on.

An alternative response to the concern that conventional priors may be unintentionally informative about the structural impulse responses is the use of the robust Bayesian approach of Giacomini and Kitagawa (2015), which provides bounds on the structural impulse responses. Those bounds, however, may be wide and uninformative.

A third response is to specify the prior to be agnostic (or conditionally agnostic) in the technical sense defined by Arias, Rubio-Ramírez, and Waggoner (2015) or to modify existing priors to imply a flat prior distribution over either the structural VAR representation or the structural impulse response representation. As in Baumeister and Hamilton (2015a), the posterior draws generated by these alternative algorithms to date have only been evaluated under restrictive loss functions.

A fourth response is to dispense with all priors and to rely on frequentist confidence sets for sign-identified structural impulse responses, as discussed in Moon, Schorfheide, and Granziera (2013), Gafarov, Meier, and Montiel Olea (2015a, 2015b), and Kitagawa, Montiel Olea, and Payne (2015), but the latter approach also tends to produce bounds that are too wide to be informative in practice.

14 Identification by Heteroskedasticity or Non-Gaussianity

14.1 Introduction

As we have seen in Chapters 8 and 10, the identification of structural VAR models typically relies on economically motivated identifying restrictions. Another strand of the literature exploits certain statistical properties of the data for identification. In particular, changes in the conditional or unconditional volatility of the VAR errors (and hence of the observed variables) can be used to assist in the identification of structural shocks. For example, Rigobon (2003), Rigobon and Sack (2003), and Lanne and Lütkepohl (2008) rely on unconditional heteroskedasticity, whereas Normandin and Phaneuf (2004), Bouakez and Normandin (2010), and Lanne, Lütkepohl, and Maciejowska (2010) exploit conditional heteroskedasticity.

In this chapter, we explain the principle of identification by heteroskedasticity.¹ In Section 14.2, the general modeling strategy is presented and its advantages and limitations are discussed. The central idea is that in a conventional structural VAR analysis the structural shocks are recovered by transforming the reduced-form residuals. As we have seen in previous chapters, this is typically done through exclusion restrictions. The current chapter considers the question of how changes in the volatility of the model errors can be used for this purpose. We show that the assumption that the structural impulse responses are time invariant, as the volatility of the reduced-form shocks changes, provides additional restrictions that can be used to uniquely pin down mutually uncorrelated shocks. There is nothing in this purely statistical identification procedure, however, that ensures that these shocks are also economically meaningful, making it difficult to interpret them as structural VAR shocks.

One way to assess whether all or some of the shocks identified by heteroskedasticity correspond to economic shocks and, hence, can be interpreted

¹ This chapter is partly adapted from Lütkepohl (2013b). Other papers from which we draw are Lütkepohl and Velinov (2016) and Lütkepohl and Netšunajev (2017).

as proper structural shocks is to treat conventional identifying restrictions as overidentifying restrictions within the heteroskedastic model, facilitating formal tests of these restrictions.

In some cases it may also be possible to infer the economic interpretation of these shocks informally from comparisons of the implied impulse responses with impulse response estimates based on conventional structural VAR models, as illustrated in Lütkepohl and Netšunajev (2014). A necessary condition for such comparisons is that the structural impulse responses in the VAR model based on conventional identifying restrictions can be estimated consistently under the assumptions maintained in the explicitly heteroskedastic VAR model.

In Section 14.3, we consider some specific models of changes in volatility. First, models of heteroskedasticity with extraneously generated shifts in the variance are discussed. We then introduce models in which the timing of the volatility change is determined by the data. One example is models in which volatility changes are governed by a Markov regime-switching mechanism. Another example is the smooth-transition model for capturing changes in the error volatility. Third, we examine a model with multivariate generalized autoregressive conditionally heteroskedastic (GARCH) errors. Detailed examples are provided for each of these models.

Throughout Section 14.3, it is assumed that the impact effects of the structural shocks remain constant, as the volatility of the structural shocks changes. In contrast, in Section 14.4 we allow for time-varying impact effects and consider models that permit the primitive shocks of the model to be mutually correlated. Finally, in Section 14.5 we examine an alternative identification strategy that exploits the non-Gaussianity of the VAR model errors in many applications. Section 14.6 summarizes the pros and cons of each approach.

14.2 The Model Setup

14.2.1 The Baseline Model

The point of departure is a K -dimensional reduced-form $\text{VAR}(p)$ process,

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t. \quad (14.2.1)$$

As usual, A_j , $j = 1, \dots, p$, are $K \times K$ VAR coefficient matrices and u_t is a serially uncorrelated zero-mean error term subject to conditional or unconditional heteroskedasticity. The model may contain cointegrated variables and may in fact be set up as a vector error correction model.² In addition, the VAR

² In the latter case, one may estimate the cointegration relations in a first step without accounting for heteroskedasticity and fix the cointegration parameters at those estimates in the subsequent analysis.

model may include additional unmodeled right-hand side variables. As long as time-invariance of all parameters apart from those in the error covariance structure can be justified, this extension is straightforward.

In earlier chapters, the structural shocks, w_t , were obtained from the reduced-form errors by a linear transformation, $w_t = B_0 u_t$ or, equivalently, $B_0^{-1} w_t = u_t$. The matrix B_0 was chosen such that the components of w_t are instantaneously uncorrelated. Thus, depending on the normalization assumptions,

$$\Sigma_u = B_0^{-1} B_0^{-1\prime} \quad \text{or} \quad \Sigma_u = B_0^{-1} \Sigma_w B_0^{-1\prime},$$

where $\mathbb{E}(u_t u_t') = \Sigma_u$ was assumed to be time-invariant. The identification of the structural shocks relied on normalizing assumptions and additional economically motivated identifying restrictions on the elements of B_0 or B_0^{-1} .

An alternative approach to identification becomes feasible if the unconditional variances of the reduced-form shocks (and, hence, also the unconditional variances of the structural shocks) change during the sample period. To see this suppose that $\mathbb{E}(u_t u_t') = \Sigma_1$ for $t = 1, \dots, T_1$, and $\mathbb{E}(u_t u_t') = \Sigma_2$ for $t > T_1$, where $\Sigma_1 \neq \Sigma_2$. Moreover, suppose that all other VAR parameters in (14.2.1) remain time-invariant. Then there is a matrix decomposition result that ensures the existence of a matrix G and a diagonal matrix $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_K)$ such that

$$\Sigma_1 = GG' \quad \text{and} \quad \Sigma_2 = G\Lambda G' \tag{14.2.2}$$

(see Lütkepohl 1996, chapter 6). This G matrix may be viewed as a solution for the structural impact multiplier matrix B_0^{-1} , which implies that the structural shocks $w_t = G^{-1} u_t$ have variance

$$\mathbb{E}(w_t w_t') = \begin{cases} I_K, & t = 1, \dots, T_1, \\ \Lambda, & t > T_1. \end{cases}$$

Since Λ is diagonal, the structural errors obtained in this way satisfy the basic requirement of being instantaneously uncorrelated across the whole sample. Lanne, Lütkepohl, and Maciejowska (2010) show that the matrix B_0^{-1} obtained from this decomposition is unique apart from changes in the signs and permutations of the columns, if the diagonal elements of Λ are all distinct. If the latter condition is satisfied, we may obtain unique shocks by just imposing the basic requirement that the structural shocks are instantaneously uncorrelated. The possible changes in sign just mean that we consider negative instead of positive shocks and vice versa. The fact that the columns of B_0^{-1} can be permuted means that the ordering of the shocks can be chosen freely. Of course, using the same transformation matrix B_0^{-1} for the whole sample period implies that the impact effects of the shocks are time-invariant as well and only the variances change.

14.2.2 An Illustrative Example

It is useful to review an illustrative example in some detail. Consider the bivariate system

$$u_t = \begin{pmatrix} u_{1t} \\ u_{2t} \end{pmatrix} = \begin{bmatrix} b_0^{11} & b_0^{12} \\ b_0^{21} & b_0^{22} \end{bmatrix} \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix},$$

where b_0^{ij} denotes the ij^{th} element of B_0^{-1} . Substituting $G = B_0^{-1}$ into the relations (14.2.2) yields

$$\begin{bmatrix} \sigma_{1,1}^2 & \sigma_{12,1} \\ \sigma_{12,1} & \sigma_{2,1}^2 \end{bmatrix} = \begin{bmatrix} (b_0^{11})^2 + (b_0^{12})^2 & b_0^{11}b_0^{21} + b_0^{12}b_0^{22} \\ b_0^{11}b_0^{21} + b_0^{12}b_0^{22} & (b_0^{21})^2 + (b_0^{22})^2 \end{bmatrix}$$

and

$$\begin{bmatrix} \sigma_{1,2}^2 & \sigma_{12,2} \\ \sigma_{12,2} & \sigma_{2,2}^2 \end{bmatrix} = \begin{bmatrix} \lambda_1(b_0^{11})^2 + \lambda_2(b_0^{12})^2 & \lambda_1b_0^{11}b_0^{21} + \lambda_2b_0^{12}b_0^{22} \\ \lambda_1b_0^{11}b_0^{21} + \lambda_2b_0^{12}b_0^{22} & \lambda_1(b_0^{21})^2 + \lambda_2(b_0^{22})^2 \end{bmatrix}.$$

Thus, we have six distinct equations

$$\begin{aligned} \sigma_{1,1}^2 &= (b_0^{11})^2 + (b_0^{12})^2, \\ \sigma_{12,1} &= b_0^{11}b_0^{21} + b_0^{12}b_0^{22}, \\ \sigma_{2,1}^2 &= (b_0^{21})^2 + (b_0^{22})^2, \\ \sigma_{1,2}^2 &= \lambda_1(b_0^{11})^2 + \lambda_2(b_0^{12})^2, \\ \sigma_{12,2} &= \lambda_1b_0^{11}b_0^{21} + \lambda_2b_0^{12}b_0^{22}, \\ \sigma_{2,2}^2 &= \lambda_1(b_0^{21})^2 + \lambda_2(b_0^{22})^2, \end{aligned}$$

from which we can solve for the six structural parameters $b_0^{11}, b_0^{12}, b_0^{21}, b_0^{22}, \lambda_1, \lambda_2$. The last two parameters are strictly greater than zero. They are the diagonal elements of Λ . The solution is unique (up to sign) if λ_1 and λ_2 are distinct and are ordered such as $\lambda_1 < \lambda_2$. Note that the variances of the structural shocks are normalized to one in the first part of the sample. Hence, the λ_i indicate the changes in the variances from the first to the second volatility regime. In other words, they can be interpreted as the relative variances in the second part of the sample. The condition $\lambda_1 \neq \lambda_2$ which ensures uniqueness just means that the change in variance is not the same for both variables. In fact, for a unique solution it suffices that the variance of one of the variables changes.

It should be clear, however, that this is a purely statistical approach to uniquely identifying mutually uncorrelated shocks that does not necessarily result in economically meaningful shocks, much like orthogonalizing the reduced-form errors using a Cholesky decomposition does not automatically lead to economically meaningful shocks. A case in point is the bivariate

demand and supply model considered in Rigobon (2003). Rigobon exploits heteroskedasticity to identify a unique set of mutually uncorrelated shocks. Without further economic identifying assumptions, however, we cannot infer which shock, if any, corresponds to the demand shock and which to the supply shock. Thus, the estimates are not structural in the sense discussed in Chapters 8 and 10.

If the statistical identification conditions for the shocks in the heteroskedastic model are satisfied, all we can say is that there are two mutually uncorrelated shocks that induce time-invariant impact responses (and hence more generally time-invariant impulse response functions) throughout the sample. The assumption of time-invariant impact responses is, of course, also maintained in conventional structural VAR analysis with unmodeled heteroskedasticity. While this assumption may be unrealistic, as discussed in Chapter 18, it is required for the methods discussed in the current chapter.

Although this illustrative example involved only one volatility shift, the idea of identification by heteroskedasticity generalizes easily to more than two volatility regimes. In fact, the volatility may change in every period.

14.2.3 The General Model

Identification by heteroskedasticity was originally proposed in the context of shifts in the unconditional variance, but may also be extended to shifts in the conditional variance as illustrated in Sections 14.3.2 and 14.3.4. In general,

$$\mathbb{E}(u_t u_t') = \Sigma_t = B_0^{-1} \Lambda_t B_0^{-1'} \quad \text{or} \quad \mathbb{E}_t(u_t u_t') = \Sigma_t = B_0^{-1} \Lambda_t B_0^{-1'} \quad (14.2.3)$$

with diagonal matrices $\Lambda_t = \text{diag}(\lambda_{1t}, \dots, \lambda_{Kt})$, where the Σ_t may alternatively denote the unconditional covariance matrix or the conditional covariance matrix. Of course, a decomposition of Σ_t as in (14.2.3) may not exist. In particular, there may not exist a time-invariant matrix B_0^{-1} such that $\Sigma_t = B_0^{-1} \Lambda_t B_0^{-1'}$. In fact, the existence of such a decomposition imposes testable restrictions on Σ_t , allowing us to test whether the data are compatible with this decomposition. In the latter case, we may use B_0^{-1} to transform the reduced-form errors into structural errors with time-invariant impact effects.

In order to estimate and analyze the volatility structure we first need to parameterize and estimate the time-varying diagonal covariance matrices. In Section 14.3 we discuss a number of volatility models that have been used for this purpose in the literature. Some of these models assume a finite number of volatility states, whereas others allow for infinitely many volatility states. In each case we discuss the identification conditions required for the uniqueness of the transformation matrix B_0^{-1} and, hence, of the structural shocks.

If unique structural shocks are obtained via heteroskedasticity, then any further restrictions on B_0^{-1} become overidentifying. In particular, restrictions on

the impact or the long-run effects imposed in conventional SVAR analyses are overidentifying and, hence, can be tested against the data. This implication of the present framework is convenient if there are additional competing identifying restrictions such as the exclusion restrictions commonly used in the homoskedastic model. In this case the data can potentially speak against such restrictions or confirm that they are compatible with the data. Of course, it can happen that even this framework does not allow us to discriminate between competing economic theories. In other words, two competing structural forms may both be in line with the data in the present framework. Moreover, rejecting a particular model in the present framework could be the result of the underlying assumptions of the statistical model not being compatible with the data. For example, the assumption of regime-invariant impact effects may be problematic. In that case a particular set of exclusion restrictions may be rejected, although it is actually the modeling framework that is inappropriate.

14.3 Alternative Volatility Models

This section discusses a range of different proposals for modeling changing volatility in this setup, where $\Sigma_t = B_0^{-1} \Lambda_t B_0^{-1'}$ with B_0^{-1} being the matrix of impact effects of the structural shocks and Λ_t being a diagonal matrix.

14.3.1 Structural VAR Models with Extraneously Specified Volatility Changes

General Setup. In the original article on identifying mutually uncorrelated VAR shocks by heteroskedasticity, Rigobon (2003) considers changes in the unconditional variances of the shocks at discrete points in time. In the baseline model the date of the change in volatility is deterministic and known. Suppose that the reduced-form error covariance matrix in time period t is

$$\mathbb{E}(u_t u_t') = \begin{cases} \Sigma_1 & \text{for } t \in \mathcal{T}_1, \\ \vdots & \\ \Sigma_M & \text{for } t \in \mathcal{T}_M, \end{cases} \quad (14.3.1)$$

where $\mathcal{T}_m = \{T_{m-1} + 1, \dots, T_m\}$, $m = 1, \dots, M$, and there are M mutually exclusive volatility regimes. T_m , for $m > 0$, denotes the period of a change in the volatility regime. It is assumed that $T_0 = 0$ and $T_M = T$. At least two of the error covariance matrices, Σ_m , are assumed to be distinct.

Further suppose that we can decompose the error covariance matrices according to

$$\Sigma_1 = B_0^{-1} B_0^{-1'}, \quad \Sigma_m = B_0^{-1} \Lambda_m B_0^{-1'}, \quad m = 2, \dots, M, \quad (14.3.2)$$

with diagonal matrices $\Lambda_m = \text{diag}(\lambda_{1m}, \dots, \lambda_{Km})$. Clearly, such a decomposition may not exist if $M > 2$ and arbitrary covariance matrices Σ_m , $m = 1, \dots, M$, are allowed for. Since the decomposition imposes restrictions on the reduced form, however, its validity can be tested against the data as we will see shortly.

Identification. In the context of this model, the shocks w_t are statistically identified if B_0^{-1} is statistically identified. Lanne, Lütkepohl, and Maciejowska (2010, Proposition 1) show that uniqueness of B_0^{-1} (apart from column ordering and column sign changes) is ensured if for any two subscripts $k, l \in \{1, \dots, K\}$, $k \neq l$, there exists an $m \in \{2, \dots, M\}$ such that $\lambda_{km} \neq \lambda_{lm}$. In other words, there must be at least one regime, in which the change in volatility in variable k is different from that in variable l , relative to what it is in regime 1. Although this condition for exact (local) identification is more complicated than in the 2-state case, it is easy to check. Moreover, the identification condition can in principle be examined with statistical tests because, if there are M distinct volatility regimes, the diagonal elements of the Λ_m matrices are identified and, hence, can be estimated consistently with a proper asymptotic distribution under common assumptions. We discuss this issue in more detail later in the chapter.

So far we have assumed that only the restrictions from the covariance decomposition are imposed on B_0^{-1} . In practice, sometimes the elements on the main diagonal of B_0 are also restricted to be one. In that case the variances of the structural shocks in the first volatility regime cannot be normalized to one, and we consider a decomposition of the reduced-form covariance matrices of the form

$$\Sigma_m = B_0^{-1} \Lambda_m^* B_0^{-1'}, \quad m = 1, \dots, M, \quad (14.3.3)$$

where the $\Lambda_m^* = \text{diag}(\lambda_{1m}^*, \dots, \lambda_{Km}^*)$ are diagonal matrices. It can be shown that for $M \geq 2$ the matrix B_0 is unique, if the following condition is satisfied:

$$\begin{aligned} &\forall k, l \in \{1, \dots, K\} \text{ with } k \neq l, \exists m \in \{2, \dots, M\} \\ &\text{such that } \lambda_{km}/\lambda_{k1} \neq \lambda_{lm}/\lambda_{l1}. \end{aligned} \quad (14.3.4)$$

This condition even ensures global identification of B_0 if this matrix has no further unit elements apart from those on its main diagonal. In that case column or row permutations are not possible, while maintaining the unit diagonal, ensuring the uniqueness of the solution. Enforcing a unit diagonal of B_0 also rules out sign changes of the shocks. Note, however, that a unit diagonal of B_0 does not ensure a unit diagonal of the matrix of impact effects B_0^{-1} . Only the signs of the impact effects are fixed by requiring B_0 to have a unit main

diagonal. Moreover, the fact that $w_t = B_0 u_t$ implies that the structural-form covariance matrices

$$\mathbb{E}(w_t w_t') = \Lambda_m^* \quad \text{for } t \in \mathcal{T}_m,$$

are diagonal, as required by the assumption of instantaneously uncorrelated structural shocks.

One requirement for condition (14.3.4) is that the variances of the model variables do not change proportionately. For example, if there are just two volatility states and all variances change proportionally such that for some scalar c , $\Lambda_1 = c\Lambda_2$, then $\lambda_{k2}/\lambda_{k1} = \lambda_{l2}/\lambda_{l1}$, so that condition (14.3.4) is not satisfied.

It is important to reiterate that the identification conditions for B_0 given so far are purely statistical conditions that allow consistent estimation of all the elements of B_0 without imposing any further restrictions on the model. Hence, the implied shocks w_t may not have an economic interpretation. Still, the statistical uniqueness is useful because it may enable us to perform statistical tests of additional economic identifying restrictions conditional on the heteroskedastic model being correctly specified.

Estimation and Inference. Assuming that $u_t \sim \mathcal{N}(0, \Sigma_t)$ with a covariance structure that satisfies (14.3.1) and (14.3.2), the model may be estimated by ML. The log-likelihood function is

$$\log l(\boldsymbol{\alpha}, \boldsymbol{\sigma}) = -\frac{KT}{2} \log 2\pi - \frac{1}{2} \sum_{t=1}^T \log(\det(\Sigma_t)) - \frac{1}{2} \sum_{t=1}^T u_t' \Sigma_t^{-1} u_t, \quad (14.3.5)$$

where $\boldsymbol{\sigma}$ contains all unknown covariance parameters and the time-invariant VAR parameters including deterministic terms are collected in $\boldsymbol{\alpha} = \text{vec}[\nu, A_1, \dots, A_p]$ (see Lütkepohl (2005, chapter 17) for further details). The log-likelihood (14.3.5) can also be used for quasi-ML estimation if u_t (and hence y_t) is not normally distributed. Alternatively one may estimate the VAR model by equation-by-equation LS and use the residuals, \hat{u}_t , for estimating the covariance matrices as

$$\widehat{\Sigma}_m = \frac{1}{\#\mathcal{T}_m} \sum_{t \in \mathcal{T}_m} \hat{u}_t \hat{u}_t',$$

where $\#\mathcal{T}_m$ denotes the number of observations in \mathcal{T}_m . In the next step these estimates can then be used to form the GLS estimator

$$\widehat{\boldsymbol{\alpha}} = \left(\sum_{t=1}^T Z_{t-1} Z_{t-1}' \otimes \widehat{\Sigma}_t^{-1} \right)^{-1} \left(\sum_{t=1}^T (Z_{t-1} \otimes \widehat{\Sigma}_t^{-1}) y_t \right), \quad (14.3.6)$$

where $Z_{t-1} = (1, y'_{t-1}, \dots, y'_{t-p})'$ and $\widehat{\Sigma}_t = \widehat{\Sigma}_m$ for $t \in \mathcal{T}_m$. If the VAR process is stable, these estimators have standard asymptotic properties. Although this type of model typically requires nonlinear estimation, these estimation methods are simple enough to allow the use of bootstrap methods for inference on the structural impulse responses.

The structural parameters can be estimated by substituting $B_0^{-1}B_0^{-1'}$ for Σ_1 , $B_0^{-1}\Lambda_m B_0^{-1'}$ for Σ_m , $m = 2, \dots, M$, and $y_t - (Z'_t \otimes I_K)\alpha$ for u_t in (14.3.5) and maximizing the log-likelihood. Alternatively, the α parameters can be estimated in a first step, and in the second step the concentrated log-likelihood is optimized only with respect to the structural parameters B_0^{-1} and Λ_m , $m = 2, \dots, M$.

GMM is an alternative estimation method for the structural parameters. In that case, the structural parameters are estimated by minimizing the objective function

$$J = \begin{pmatrix} \text{vech}_1 \\ \vdots \\ \text{vech}_M \end{pmatrix}' W \begin{pmatrix} \text{vech}_1 \\ \vdots \\ \text{vech}_M \end{pmatrix}, \quad (14.3.7)$$

where

$$\begin{aligned} \text{vech}_1 &= \text{vech} \left(\frac{1}{\#\mathcal{T}_1} \sum_{t \in \mathcal{T}_1} \hat{u}_t \hat{u}'_t - B_0^{-1} B_0^{-1'} \right), \\ \text{vech}_m &= \text{vech} \left(\frac{1}{\#\mathcal{T}_m} \sum_{t \in \mathcal{T}_m} \hat{u}_t \hat{u}'_t - B_0^{-1} \Lambda_m B_0^{-1'} \right) \quad \text{for } m = 2, \dots, M, \end{aligned}$$

and W is a positive definite weighting matrix. For example, W may be constructed as a block-diagonal matrix from the inverse covariance matrices of the $\text{vech}(\hat{u}_t \hat{u}'_t)$ as

$$W = \begin{bmatrix} W_1 & & 0 \\ & \ddots & \\ 0 & & W_M \end{bmatrix},$$

with

$$\begin{aligned} W_m &= \left(\frac{1}{\#\mathcal{T}_m} \sum_{t \in \mathcal{T}_m} \left(\text{vech}(\hat{u}_t \hat{u}'_t) - \overline{\text{vech}(\hat{u}_t \hat{u}'_t)} \right) \right. \\ &\quad \times \left. \left(\text{vech}(\hat{u}_t \hat{u}'_t) - \overline{\text{vech}(\hat{u}_t \hat{u}'_t)} \right)' \right)^{-1} \end{aligned}$$

and

$$\overline{\text{vech}(\hat{u}\hat{u}') = \frac{1}{\#\mathcal{T}_m} \sum_{t \in \mathcal{T}_m} \text{vech}(\hat{u}_t \hat{u}'_t)} \quad \text{for } m = 1, \dots, M$$

(see, e.g., Ehrmann, Fratzscher, and Rigobon (2011) and Wright (2012) for applications of GMM estimators for related models).

The GMM and ML estimators of the identified parameters have standard asymptotic properties under common assumptions. In fact, Rigobon (2003) shows that consistent estimators of the structural parameters are obtained, even if the change points for the volatility regimes are misspecified.

Testing the Model Assumptions. In principle, one can test the crucial assumptions underlying the heteroskedastic model. Three types of tests are of particular interest. The first test addresses the question of the existence of the decomposition (14.3.2). An LR test of this null hypothesis can be constructed by comparing the maxima of the log-likelihood with and without the restriction imposed. This LR statistic under usual assumptions has an asymptotic χ^2 distribution with

$$\frac{1}{2}MK(K+1) - K^2 - (M-1)K$$

degrees of freedom under the null hypothesis (see Lanne, Lütkepohl, and Maciejowska 2010).

The second test assesses the validity of the identifying restrictions conditional on the existence of the decomposition (14.3.2). The central idea is that the estimated diagonal elements of the Λ_m matrices can be used to investigate whether the diagonal elements are distinct and, hence, whether the shocks are identified. Unfortunately, developing formal statistical tests of the relevant null hypotheses is not straightforward. To see this, consider a bivariate system with $M = 2$ such that there are two volatility regimes. Decomposing the two covariance matrices as in (14.2.2), we have $\Sigma_1 = B_0^{-1}B_0^{-1'}$ and $\Sigma_2 = B_0^{-1}\Lambda B_0^{-1'}$ with $\Lambda = \text{diag}(\lambda_1, \lambda_2)$. We are interested in testing

$$\mathbb{H}_0 : \lambda_1 = \lambda_2 \quad \text{versus} \quad \mathbb{H}_1 : \lambda_1 \neq \lambda_2.$$

Note that the elements of B_0^{-1} are not identified under \mathbb{H}_0 , but only under \mathbb{H}_1 when $\lambda_1 \neq \lambda_2$. It is well known that, when there are unidentified parameters under \mathbb{H}_0 , standard tests such as LR and Wald tests may not have their usual asymptotic χ^2 distributions. This point is not fully appreciated in this literature.

Finally, it is possible to treat conventional short-run or long-run exclusion restrictions on B_0^{-1} (or on B_0) as overidentifying restrictions in heteroskedastic VAR models. The validity of these overidentifying restrictions may be formally tested by comparing the maxima of the log-likelihood with and without the restriction imposed. This LR test under general conditions has an asymptotic

$\chi^2(n)$ distribution, where n denotes the number of overidentifying restrictions. Note that there are no formal tests of overidentifying sign restrictions, but we may assess whether the estimate of B_0^{-1} from the heteroskedastic VAR model satisfies the sign restrictions (see Lütkepohl and Netšunajev 2014).

The asymptotic properties of these estimators and test statistics may differ substantially from their small-sample properties because the model is highly parameterized and macroeconomic time series are typically relatively short. In addition, any given volatility regime may be present only for a relatively short time span. Hence, the interpretation of the statistical results may not be straightforward. Moreover, the volatility regimes were assumed to be known, which is not the case in practice. Instead there will be uncertainty regarding the change points and perhaps some pretesting or other statistical methods for determining the change points may have been used (see, e.g., Ehrmann, Fratzscher, and Rigobon 2011). All these problems do not add to the reliability of inference in this context. Still, the fact that the data may speak to the crucial issue of the identification of the shocks is a potential advantage over having no information on these issues.

A Detailed Empirical Example. To illustrate how unconditional heteroskedasticity can be used to aid in the identification of structural shocks in a structural VAR analysis we consider an example from Lanne and Lütkepohl (2014) who compare different identification schemes for the monetary policy models discussed in Christiano, Eichenbaum, and Evans (1999). The latter models were already discussed in Chapter 8. They are just-identified and their identifying assumptions cannot be tested within the conventional structural VAR framework.

Let gdp_t be the log of real GDP, p_t the log of the GDP deflator, $pcom_t$ the smoothed change in an index of sensitive commodity prices, nbr_t the log of nonborrowed reserves plus extended credit, tr_t the log of total reserves, i_t the federal funds rate, and m_t the log of M1. Thus, we are dealing with a seven-dimensional system. We follow Christiano, Eichenbaum, and Evans (1999) in using monthly data for the period 1965m7–1995m6. The sample size is $T = 360$. The reduced-form model is a VAR(12) with an intercept.

As in Section 8.4.5, the vector of variables, y_t , is partitioned as

$$y_t = \begin{pmatrix} y_{1t} \\ z_t \\ y_{2t} \end{pmatrix},$$

where y_{1t} is $K_1 \times 1$, z_t is a scalar variable, and y_{2t} is $K_2 \times 1$. The vector y_{1t} contains variables whose contemporaneous values appear in the monetary authority's information set. They are assumed to be orthogonal to the monetary policy shock. The variable z_t is the policy instrument of the monetary authority, and the variables in y_{2t} appear with a lag in the information set.

Identification of the monetary policy shock is achieved by assuming that the matrix B_0 is lower block-triangular:

$$B_0 = \begin{bmatrix} B_{11,0} & 0_{K_1 \times 1} & 0_{K_1 \times K_2} \\ K_1 \times K_1 & & \\ B_{21,0} & B_{22,0} & 0_{1 \times K_2} \\ 1 \times K_1 & 1 \times 1 & \\ B_{31,0} & B_{32,0} & B_{33,0} \\ K_2 \times K_1 & K_2 \times 1 & K_2 \times K_2 \end{bmatrix}.$$

The monetary policy shock is the $(K_1 + 1)^{\text{th}}$ element of w_t . The other shocks are not of direct interest and can therefore be identified arbitrarily by choosing B_0 to be lower-triangular, which means that $B_{11,0}$ and $B_{33,0}$ are restricted to be lower triangular. This choice does not affect the monetary policy shock.

Lanne and Lütkepohl (2014) compare the three alternative identification schemes already discussed in Chapter 8, Section 8.4.5. They differ by the choice of y_{1t} , z_t , and y_{2t} as follows:

FFR policy shock: $y_{1t} = (gnp_t, p_t, pcom_t)', z_t = i_t$, and $y_{2t} = (nbr_t, tr_t, m_t)'$.

NBR policy shock: $y_{1t} = (gnp_t, p_t, pcom_t)', z_t = nbr_t$, and $y_{2t} = (i_t, tr_t, m_t)'$.

NBR/TR policy shock: $y_{1t} = (gnp_t, p_t, pcom_t, tr_t)', z_t = nbr_t$, and $y_{2t} = (i_t, m_t)'$.

There are a number of reasons for structural breaks in the model during the sample period 1965m7–1995m6. Following Christiano, Eichenbaum, and Evans (1999) and Bernanke and Mihov (1998a), Lanne and Lütkepohl (2014) consider two structural breaks in 1979m10 and 1984m2. Thus, we work with three volatility regimes. Notice that the period 1979m10–1984m2 roughly corresponds to the era of Fed chairman Paul Volcker. This period is often regarded as special from the point of view of monetary policy. Assuming that the structural breaks affect only the error variances, Lanne and Lütkepohl (2014) report the estimated relative variances given in Table 14.1.

The estimated λ_{km} in Table 14.1 are of particular interest because they contain the crucial information for the identification of the B_0 matrix. For each pair k, l either $\lambda_{k2} \neq \lambda_{l2}$ or $\lambda_{k3} \neq \lambda_{l3}$ has to hold for full identification. For example, given the standard errors in Table 14.1, $\lambda_{12} \neq \lambda_{22}$ may not hold, whereas $\lambda_{13} \neq \lambda_{23}$ may well hold. Of course, it is not clear that such results hold for all pairs of relative variances, as required for full identification. This hypothesis should ideally be tested by formal statistical procedures. Such tests are not provided in Lanne and Lütkepohl (2014) who point out, however, that even if B_0 is not fully identified by heteroskedasticity, the available information may be sufficient to test the conventional restrictions implied by the three identification schemes presented earlier. Clearly, if some of the restrictions are rejected,

Table 14.1. *Estimation Results for Parameters of VAR(12) Model with Unconditionally Heteroskedastic Errors for Sampling Period 1965m7–1995m6 and Variance Changes in 1979m10 and 1984m2*

Regime	$m = 2$		$m = 3$	
Parameter	Estimate	Std Error	Estimate	Std Error
λ_{1m}	0.8735	0.2761	0.4623	0.0801
λ_{2m}	1.2161	0.4405	0.7458	0.1284
λ_{3m}	0.9791	0.2845	0.7992	0.1397
λ_{4m}	5.2158	1.1729	0.6916	0.1179
λ_{5m}	1.5875	0.4137	1.8507	0.3110
λ_{6m}	1.7791	0.5160	1.4448	0.2481
λ_{7m}	0.6639	0.3452	0.3619	0.0757

Source: Adapted from Table 1 of Lanne and Lütkepohl (2014).

this implies that there is sufficient identifying information to enable the data to speak against the restrictions.

Table 14.2 reports test results from Lanne and Lütkepohl (2014) for the zero restrictions on B_0 implied by the three competing identification schemes. The p -values reported in the table are determined using the degrees of freedom in the asymptotic χ^2 distribution of the test statistics that would have been obtained if the null hypotheses were true and the B_0 matrix were fully identified

Table 14.2. *LR Type Tests of Identification Schemes Based on Unconditionally Heteroskedastic Models with Regime Changes in 1979m10 and 1984m2*

	H_0	df	LR	p -value
FFR	$B_{12,0} = 0_{3 \times 1}$	3	2.7770	0.4273
	$B_{13,0} = 0_{3 \times 3}$	9	9.1872	0.4202
	$B_{23,0} = 0_{1 \times 3}$	3	1.5451	0.6719
	$B_{12,0} = 0_{3 \times 1}, B_{13,0} = 0_{3 \times 3}, B_{23,0} = 0_{1 \times 3}$	15	17.4070	0.2951
NBR	$B_{12,0} = 0_{3 \times 1}$	3	1.2528	0.7404
	$B_{13,0} = 0_{3 \times 3}$	9	28.8631	0.0007
	$B_{23,0} = 0_{1 \times 3}$	3	1.3433	0.7189
	$B_{12,0} = 0_{3 \times 1}, B_{13,0} = 0_{3 \times 3}, B_{23,0} = 0_{1 \times 3}$	15	37.7858	0.0010
NBR/TR	$B_{12,0} = 0_{4 \times 1}$	4	26.3575	2.6802e-5
	$B_{13,0} = 0_{4 \times 2}$	8	20.9426	0.0073
	$B_{23,0} = 0_{1 \times 2}$	2	0.6334	0.7286
	$B_{12,0} = 0_{4 \times 1}, B_{13,0} = 0_{4 \times 2}, B_{23,0} = 0_{1 \times 2}$	14	60.8736	8.3617e-8

Source: Adapted from Table 2 of Lanne and Lütkepohl (2014).

by heteroskedasticity. As this is not necessarily the case, the actual degrees of freedom may be smaller and, hence, also the p -values may be smaller. In other words, whenever a restriction in Table 14.2 is rejected based on its p -value, it would also be rejected if the actual number of degrees of freedom were smaller than assumed in the table. Thus, there is strong evidence that some of the restrictions associated with the NBR and NBR/TR identification schemes can be rejected. For example, $B_{13,0} = 0$ and the joint hypothesis $B_{12,0} = 0$, $B_{13,0} = 0$, $B_{23,0} = 0$ are clearly rejected for the NBR scheme at conventional significance levels. Likewise, $B_{12,0} = 0$, $B_{13,0} = 0$, and the joint hypothesis $B_{12,0} = 0$, $B_{13,0} = 0$, $B_{23,0} = 0$ are clearly rejected for the NBR/TR scheme. In other words, the only identification scheme not rejected in our framework is the FFR scheme. Hence, the changes in volatility help discriminate between these three identification schemes.

It is perhaps worth emphasizing that not rejecting the assumptions underlying the FFR scheme does not necessarily imply that this scheme is correct. Of course, if a statistical test does not reject the null hypothesis of interest, this fact can always be due to low power and does not necessarily confirm the validity of the null hypothesis. In the current example, it is even more problematic to view a non-rejection as support for the FFR model. The reason is that a non-rejection may simply reflect the absence of overidentifying information, since the B_0 matrix may not be fully identified by heteroskedasticity.

Other Empirical Examples. Similar models with extraneously assigned volatility regimes have also been used in a number of other studies. For example, Rigobon (2003) applies this approach to investigate the relationship between returns on sovereign bonds in Argentina, Brazil, and Mexico, using daily bond returns for the period January 1994 to December 2001. Identification by heteroskedasticity may seem appealing in this application because bond yields are simultaneously determined, rendering conventional exclusion restrictions clearly inappropriate, but the resulting shock and impulse response estimates have no obvious economic interpretation.

A similar problem of interpretation arises in Ehrmann, Fratzscher, and Rigobon (2011) who rely on identification through heteroskedasticity when investigating the linkages between the U.S. and euro area money, bond, and equity markets. They analyze a system of short-term interest rates, long-term bond rates, stock returns from both regions, and an exchange rate based on two-day returns for the period 1989–2008. Changes in volatility are determined based on estimates of the variance from a rolling window.

These examples illustrate a tendency in applied studies to rely on identification by heteroskedasticity to compensate for the lack of economic identification conditions. It should be clear that shocks obtained in this way need not correspond to the economically interpretable structural shocks that economists

are interested in. Nor does the implied estimate of B_0 correspond to an economically meaningful structure.

There are exceptions to this rule, however. Notably, identification by heteroskedasticity may allow us to identify selected elements of B_0 in certain economic models. For example, Rigobon and Sack (2003) propose a structural model in which the slope parameter of the monetary policy reaction function with respect to stock returns can be identified by shifts in the variance of shocks to investor risk preferences relative to monetary policy shocks. They distinguish between four different volatility regimes that are determined by 30-day rolling variance estimates. Their analysis is based on daily U.S. data from March 1985 to December 1999. They find that monetary policy reacts to stock returns. For a related application see also Rigobon and Sack (2004).

Finally, Wright (2012) is another study that bridges the gap between statistical and economic identification. Wright observes that monetary shocks have larger than usual variance on days of FOMC meetings and on days when important speeches are delivered by FOMC members. He uses daily data from November 3, 2008 to September 30, 2011 on six U.S. interest rates and identifies monetary policy shocks by utilizing the difference in the error variance of normal days and days with larger variance. Wright is only interested in the monetary policy shock and, hence, in partial identification of the system. The assumption that only the volatility of the monetary policy shock changes on days of larger variance allows him to label the identified shock as a monetary policy shock.

Even if identification by heteroskedasticity does not suffice to estimate the parameters of the underlying economic model, we may use the model estimates obtained under these assumptions to test the validity of alternative structural VAR models based on conventional identifying restrictions. For example, Lanne and Lütkepohl (2008) use identification by heteroskedasticity to assess the validity of several models for the U.S. money market that have been considered in the literature. Their results are in line with those discussed in our earlier detailed example.

14.3.2 Structural VAR Models with Markov Switching in the Variances

In practice, the timing of the volatility changes is rarely known. Typically it has to be determined from the data. One approach to modeling volatility changes as an endogenous process was proposed by Lanne, Lütkepohl, and Maciejowska (2010). They model volatility as state dependent with the state of the system evolving according to a Markov process. The methodology for using this approach in the structural VAR framework is partly due to Herwartz and Lütkepohl (2014). We first present the model setup and then discuss identification conditions, statistical inference, and applications.

Model. It is assumed that the volatility changes are driven by a discrete Markov process $s_t, t \in \mathbb{Z}$, with M states. In other words, $s_t \in \{1, \dots, M\}$. The transition probabilities between the states are

$$p_{ij} = \mathbb{P}(s_t = j | s_{t-1} = i), \quad i, j = 1, \dots, M.$$

Lanne, Lütkepohl, and Maciejowska (2010) postulate that the conditional distribution of u_t , given the state s_t , is normal, i.e.,

$$u_t | s_t \sim \mathcal{N}(0, \Sigma_{s_t}), \quad (14.3.8)$$

where all $\Sigma_m, m = 1, \dots, M$, are distinct, i.e., $\Sigma_m \neq \Sigma_n$ for $m \neq n$. The assumption of conditional normality allows one to employ ML estimation. Note that the implied unconditional distribution of u_t is not Gaussian in general and can capture a wide range of non-Gaussian distributions.

We emphasize that only the error covariance matrices depend on the Markov process, whereas the VAR coefficients are assumed to be time-invariant. VAR models in which the VAR coefficients change over time according to a Markov process have been considered by a number of authors (see, e.g., Rubio-Ramírez, Waggoner, and Zha 2005; Sims and Zha 2006b; Sims, Waggoner, and Zha 2008). Such models are discussed in Chapter 18. In this chapter, we consider VAR models in which only the error covariance structure changes. More precisely, a model with VAR lag order p and M volatility states is denoted as an MS(M)-VAR(p) model, where MS stands for Markov switching.

Although there is only a finite number of volatility states, the model can mix these states by assigning probabilities strictly between zero and one to the states in any particular period t . Thus, the model may capture gradual transitions from one state to another, and may be interpreted as a model with a continuum of states. The state covariance matrices $\Sigma_1, \dots, \Sigma_M$ are used for the identification of shocks, as in the case of known change points. In other words, the state covariance matrices are decomposed as in (14.3.2) or (14.3.3).

Identification. Uniqueness of the B_0^{-1} or B_0 matrices and, hence, of the structural shocks holds under the same conditions as for the case of known change points. The crucial condition is that there is enough heterogeneity in the volatility changes. In the literature, the variance normalization implied by the decomposition (14.3.2) rather than (14.3.3) is typically used. In other words, we normalize the variances of the structural shocks in the first state to be one. Thus the λ_{km} in the other states can be interpreted as variances relative to the first state. If they satisfy the condition

$$\forall k, l \in \{1, \dots, K\} \text{ with } k \neq l, \exists m \in \{2, \dots, M\} \text{ such that } \lambda_{km} \neq \lambda_{lm}, \quad (14.3.9)$$

then B_0^{-1} is unique up to column sign changes and column permutations. This condition must be satisfied for all of the elements of B_0 to be identified by the volatility changes.

Estimation and Inference. The statistical analysis of the MS-VAR model involves the choice of the number of volatility states and the estimation of the parameters of the overall model including the transition probabilities of the Markov process, the diagonal elements of the Λ_m matrices, and the B_0^{-1} or B_0 matrix.

Under assumption (14.3.8), which assigns a conditional normal distribution to the reduced-form errors, the log-likelihood function can be set up as

$$\begin{aligned} & \log l(\boldsymbol{\alpha}, B_0^{-1}, \boldsymbol{\lambda}, P | \mathbf{y}) \\ &= \sum_{t=1}^T \log \left(\sum_{m=1}^M \mathbb{P}(s_t = m | Y_{t-1}) f(y_t | s_t = m, Y_{t-1}) \right), \end{aligned} \quad (14.3.10)$$

where $\boldsymbol{\lambda}$ is the vector of all diagonal elements of $\Lambda_2, \dots, \Lambda_M$, P is the matrix of transition probabilities, \mathbf{y} is the full sample, $Y_{t-1} \equiv (y'_{t-1}, \dots, y'_{t-p})'$ and

$$f(y_t | s_t = m, Y_{t-1}) = (2\pi)^{-K/2} \det(\Sigma_m)^{-1/2} \exp \left\{ -\frac{1}{2} u_t' \Sigma_m^{-1} u_t \right\}.$$

The maximization of this log-likelihood function involves a highly nonlinear optimization problem that poses a number of computational challenges, even when the parameters are identified. Herwartz and Lütkepohl (2014) adopt an expectation-maximization (EM) algorithm from Krolzig (1997) for estimating this model.

They also discuss a number of problems that have to be addressed in maximizing the likelihood or log-likelihood. In particular, they point out that the likelihood function is strictly speaking unbounded and has many local maxima, the largest of which implies the preferred estimate. Moreover, the variances have to remain strictly positive and the covariance matrices positive definite in each stage of the iterative optimization algorithm. Moreover, the same ordering of the shocks has to be enforced in each iteration step. In other words, the λ_{km} have to be ordered in some way.

In practice, the ML-estimation algorithm works reliably only for reasonably small models with a small number of variables K , a small number of volatility states M , and a moderate number of lags p . If the conditional normality of the errors underlying assumption (14.3.8) does not hold, the estimation procedure may be interpreted as a quasi-ML procedure. Unfortunately, as shown by Campbell (2002), quasi-ML estimators for MS models may be inconsistent. Thus, the method has to be used with caution if the assumption of conditional normality cannot be justified. Unless a specific alternative conditional distribution can be justified, working with ML estimators based on other distributions may be preferable.

Given that the ML parameter estimators have standard asymptotic properties, inference on the λ parameters can in principle be conducted as in the case of known volatility change points. In other words, the data can be used to learn about the validity of the identification conditions. Unfortunately, given the problems with unidentified parameters under the null hypothesis mentioned in Section 14.3.1, the asymptotic properties of standard LR and Wald tests are currently unknown in general. The χ^2 distributions that have been used in this context in some of the related literature are not likely to be valid.

If the shocks can be identified by the volatility structure, conventional identifying restrictions become overidentifying in this model and can be tested, as discussed earlier. For example, zero restrictions on the impact effects or the long-run effects can be tested by LR or Wald tests conditional on the MS-VAR model specification.

It still remains to be shown how the number of volatility states in MS-VAR models may be chosen. Standard model selection criteria can be used for that purpose. Based on simulation experiments, Psaradakis and Spagnolo (2003, 2006) conclude that this approach may not be very reliable, however. Testing models with different numbers of states against the data is not easy either. The usual tests have nonstandard properties because of unidentified parameters under the null hypothesis (e.g., Hansen 1992; Garcia 1998).

After identifying the structural shocks, the model can be used for structural analysis. As discussed in earlier chapters, impulse responses are the standard tool for this purpose. Confidence intervals or joint confidence bands are usually generated with bootstrap methods. Because even a single estimation of the structural MS-VAR model is difficult, it is clear that standard bootstrap approaches are computationally challenging in this context. To mitigate the computational problems, Herwartz and Lütkepohl (2014) proposed a fixed-design wild bootstrap procedure.

An alternative to the use of frequentist methods of estimation and inference for heteroskedastic MS-VAR models is Bayesian methods. The estimation of structural MS-VAR models is discussed in Sims, Waggoner, and Zha (2008), Sims and Zha (2006b), and Rubio-Ramírez, Waggoner, and Zha (2005). These methods are discussed in Chapter 18, where more general MS-VAR models are considered that allow for time-varying slope coefficients as well. Bayesian methods for the specific MS-VAR models of this section, in which only the innovation covariance matrices vary, are developed in Kulikov and Netšunajev (2013) for the case of just-identified models and in Woźniak and Droumaguet (2015) for partially identified and overidentified models.

Empirical Examples. Lütkepohl and Velinov (2016) investigate to what extent stock prices reflect their underlying economic fundamentals, building on work by Velinov and Chen (2015). According to the dividend discount model, an asset's price is the sum of its expected future discounted dividends. Since these

Table 14.3. *Estimates and Standard Errors of Relative Variances of MS(3)-VAR(2) Model for $y_t = (\Delta q_t, \Delta r_t, \Delta s_t)'$*

Regime	$m = 2$		$m = 3$		
	Parameter	Estimate	Standard Error	Estimate	Standard Error
λ_{1m}	0.267	0.094		0.845	0.252
λ_{2m}	0.277	0.089		3.777	1.004
λ_{3m}	3.564	1.051		14.878	4.117

Source: Extracted from Table 1 of Lütkepohl and Velinov (2016).

dividends are related to real economic activity, this variable may serve as a proxy for the fundamentals.

For illustrative purposes, consider a three-dimensional model for the U.S. economy. Let $y_t = (\Delta gdp_t, \Delta r_t, \Delta sp_t)'$, where Δgdp_t is the growth rate of real GDP, Δr_t is the change in a real interest rate, and Δsp_t denotes real stock returns. We use quarterly data for the period 1947q1–2012q3.³ The model is set up in first differences because tests provide no evidence for cointegration. Related models have been used by a number of authors (see, e.g., Lee 1995; Rapach 2001; Binswanger 2004; and Lanne and Lütkepohl 2010).

Based on model selection criteria, Lütkepohl and Velinov (2016) choose an MS(3)-VAR(2) model. An LR test does not reject the existence of the decomposition (14.3.2) of the three covariance matrices. The crucial identification condition is linked to the diagonal elements of the Λ_m matrices, which have to be sufficiently heterogeneous. The estimated quantities are shown in Table 14.3. Note that in a model with three volatility states we have to investigate for each pair k, l with $k \neq l$ whether $\lambda_{k2} \neq \lambda_{l2}$ or $\lambda_{k3} \neq \lambda_{l3}$. Thus, ideally one would want to test null hypotheses

$$\mathbb{H}_0 : \lambda_{k2} = \lambda_{l2}, \quad \lambda_{k3} = \lambda_{l3}$$

for all pairs k, l . If all null hypotheses can be rejected, we have statistical evidence that the identification condition (14.3.9) holds. In the absence of suitable asymptotically valid tests, the estimated λ_{im} and their standard errors help determine whether assuming full identification by changes in volatility is reasonable in this model.

Next we focus on the support for conventional identifying restrictions for this class of models. The central question of interest is how important shocks to fundamentals are in driving stock prices. Suppose that there are two nonfundamental shocks and one fundamental shock. In a conventional analysis these structural shocks may be identified by imposing restrictions on the long-run

³ See Velinov (2013) for the details on the data.

effects such that the long-run multiplier matrix can be written as

$$\Upsilon = \begin{bmatrix} * & 0 & 0 \\ * & * & 0 \\ * & * & * \end{bmatrix}. \quad (14.3.11)$$

Each column in Υ corresponds to the effects of a specific shock on the level of each of the three variables (see Chapter 10). Thus, the first shock can have permanent effects on all variables and is hence classified as a fundamental shock. Notably its potential permanent effect on GDP distinguishes this shock from the other two shocks which are not regarded as fundamental. The second shock captures all other components that may have long-run effects on the real interest rate and does not have a specific economic interpretation. The last shock is not allowed to have long-run effects on any variable but stock prices.

These long-run identifying restrictions can be treated as overidentifying restrictions in the MS-VAR model. Lütkepohl and Velinov (2016) report a p -value of 0.207 for the LR test of the null that Υ is lower triangular. They cannot reject this restriction at conventional significance levels.

Since the long-run identification restrictions are not rejected, one could impose them in addition to the MS-VAR structure in estimating the structural impulse responses. This modification would require estimating (14.3.10) subject to the additional restrictions on the elements of Υ (see also Chapter 11 for a similar approach in a simpler context). Alternatively, one could proceed by dropping the MS-VAR structure and fitting the structural VAR model subject to the long-run restrictions only. The point estimates of the structural impulse responses from this model typically remain consistent even in the presence of unmodeled conditional (or for that matter unconditional) heteroskedasticity in the error term. Only the construction of the impulse response confidence intervals may have to be modified to allow for unmodeled volatility.

There are also a number of other studies that use the structural MS-VAR model. Lanne, Lütkepohl, and Maciejowska (2010) consider a small U.S. macroeconomic model due to Primiceri (2005) consisting of inflation, unemployment, and an interest rate. They find that previously used identification restrictions are not supported by the data when changes in volatility are taken into account. They also consider a model of the U.S. economy based on Sims, Waggoner, and Zha (2008) for log GDP, inflation, and a short-term interest rate. Using quarterly data from 1959q1–2005q4 they find support for conventional identification restrictions, as used by Sims, Waggoner and Zha.

Herwartz and Lütkepohl (2014) consider a model from Peersman (2005) that investigates the causes of the early millennium economic slowdown. Peersman uses a four-variable system consisting of the price of oil, output, a consumer price index, and a short-term interest rate. He imposes zero restrictions on the impact effects and the long-run effects of the shocks for identification. Using quarterly U.S. data from 1980q1 to 2002q2 and models with two and

three volatility states, Herwartz and Lütkepohl find that some, but not all of the identifying restrictions used by Peersman are in line with the data, when changes in volatility are accounted for. For a related analysis of the Peersman model see also Chapter 11.

Lütkepohl and Netšunajev (2014) illustrate how the structural MS-VAR approach can be used for testing models identified by sign restrictions. They do not find evidence against the identifying restrictions used by Kilian and Murphy (2012), thereby supporting the findings of the latter article. Finally, Netšunajev (2013) reconsiders alternative approaches to identifying technology shocks based on the structural MS-VAR methodology. His analysis sheds new light on the conflicting evidence on the impact on hours worked.

The MS-VAR model has appeal because it lets the data assign the observations to different volatility regimes. Unlike the model in Section 14.3.1, the MS-VAR model does not require the user to know when volatility shifts occurred. Moreover, its identification conditions can in principle be tested. Finally, the model may be estimated by Gaussian ML, if the model is small enough, allowing the use of existing results for Gaussian ML estimators of Markov-switching models. The main drawback of the model is that for larger models with long lag orders, for a larger set of variables, or for many different volatility regimes, estimation is difficult and potentially unreliable. Moreover, even for small models that can be estimated easily, little is known about the asymptotic and finite-sample properties of the proposed methods for constructing confidence intervals about structural impulse responses, for example.

14.3.3 Structural VAR Models with Smooth Transitions in the Variances

Model Setup. Another model with endogenously changing volatility postulates a smooth change in the unconditional error covariance matrix (see Lütkepohl and Netšunajev 2015). More specifically, the change in the covariance structure is modeled as a smooth transition from a volatility regime characterized by a positive definite covariance matrix Σ_1 to a regime with a different positive definite covariance matrix Σ_2 . The transition is described by a smooth transition function $G(\gamma, c, s_t)$ that depends on parameters γ and c as well as a transition variable s_t . Lütkepohl and Netšunajev use a logistic transition function,

$$G(\gamma, c, s_t) = (1 + \exp[-\exp(\gamma)(s_t - c)])^{-1}. \quad (14.3.12)$$

Clearly, $0 < G(\gamma, c, s_t) < 1$. Since $\exp(\gamma) > 0$ for positive and negative values of γ , $G(\gamma, c, s_t)$ is close to zero when s_t is much smaller than c and close to one if s_t is much larger than c (see also Chapter 18). The reduced-form error covariance matrix is specified as

$$\mathbb{E}(u_t u'_t) = (1 - G(\gamma, c, s_t))\Sigma_1 + G(\gamma, c, s_t)\Sigma_2. \quad (14.3.13)$$

This matrix is positive definite because it is a convex combination of two positive definite matrices, Σ_1 and Σ_2 . Given that $G(\gamma, c, s_t)$ is a continuous function, there is a continuum of potential covariance matrices, corresponding to an infinite number of volatility states. The number of states depends on the transition variable s_t . A transition variable may assign different volatility regimes to different parts of the sample.

Extensions of this smooth-transition VAR (ST-VAR) model to more transition terms are possible in principle, but are more difficult to handle. Such extensions may not be necessary, if a suitable transition variable can be found.

Identification. Identification conditions for structural shocks are linked to the two limiting covariance matrices Σ_1 and Σ_2 . Choosing B_0^{-1} such that

$$\Sigma_1 = B_0^{-1}B_0^{-1\prime} \quad \text{and} \quad \Sigma_2 = B_0^{-1}\Lambda_2 B_0^{-1\prime}, \quad (14.3.14)$$

where $\Lambda_2 = \text{diag}(\lambda_{12}, \dots, \lambda_{K2})$ is a diagonal matrix with positive diagonal elements, B_0^{-1} is unique up to column sign and column permutation, if the diagonal elements of Λ_2 are all distinct. Under this condition the structural shocks $w_t = B_0 u_t$ are identified and their covariance matrix is diagonal,

$$\mathbb{E}(w_t w_t') = (1 - G(\gamma, c, s_t))I_K + G(\gamma, c, s_t)\Lambda_2.$$

Estimation and Inference. If u_t is normally distributed, the log-likelihood function of the model is

$$\log l = \text{constant} - \frac{1}{2} \sum_{t=1}^T \log(\det(\Sigma_t)) - \frac{1}{2} \sum_{t=1}^T u_t' \Sigma_t^{-1} u_t, \quad (14.3.15)$$

where $\Sigma_t = \mathbb{E}(u_t u_t')$. Maximizing this function with respect to the model parameters requires iterative optimization techniques.

For a given transition function, however, the maximization of the log-likelihood with respect to the remaining parameters is straightforward. Lütkepohl and Netšunajev (2015) propose using a grid search over γ and c . They point out that, for the logistic transition function, the effective range of these parameters depends on the transition variable, but is limited to a bounded range, outside of which the transition function will not change noticeably any more. The range is chosen such that the values of the transition function cover the unit interval. In a first round a slightly wider grid can be used, which is then refined in a second round in the neighbourhood of the values maximizing the log-likelihood function in the first round.

For given values of the transition parameters γ and c , Lütkepohl and Netšunajev (2015) propose to estimate the other parameters by iterating the following steps:

Step 1: Given starting values of the reduced-form VAR parameters v, A_1, \dots, A_p , the structural parameters B_0^{-1} and Λ_2 are estimated by maximizing the log-likelihood function (14.3.15) using nonlinear maximization.

Step 2: Given the structural parameter estimates from step 1, the parameters of the reduced-form VAR model are reestimated. For given structural parameters B_0^{-1} and Λ_2 the model is linear in v, A_1, \dots, A_p . Hence, the vectorized VAR coefficients $\alpha = \text{vec}[v, A_1, \dots, A_p]$ can be estimated by generalized least squares,

$$\hat{\alpha} = [(Z \otimes I_K)\Sigma^{-1}(Z' \otimes I_K)]^{-1}(Z \otimes I_K)\Sigma^{-1}\mathbf{y},$$

where

$$\Sigma^{-1} \equiv \begin{bmatrix} \mathbb{E}(u_1 u_1')^{-1} & & 0 \\ & \ddots & \\ 0 & & \mathbb{E}(u_T u_T')^{-1} \end{bmatrix}$$

is a $KT \times KT$ block-diagonal covariance matrix, $\mathbf{y} \equiv \text{vec}[y_1, \dots, y_T]$ is a $KT \times 1$ data vector, and the t^{th} column of the $(1 + Kp) \times T$ data matrix Z is $Z_{t-1} \equiv (1, y'_{t-1}, \dots, y'_{t-p})'$.

The resulting estimates of v, A_1, \dots, A_p are then used as inputs in step 1. These steps are iterated until convergence.

This estimation procedure is technically straightforward and numerically robust even for larger VAR models, provided the transition variable is chosen such that the transition parameters and the matrices Σ_1 and Σ_2 are identified. Note that γ and c are not identified if $\Sigma_1 = \Sigma_2$. Thus, the two volatility regimes associated with these two covariance matrices must be clearly distinct. If identification of the volatility model is ensured, the asymptotic properties of the parameter estimators are standard, even if u_t is not normally distributed. In that case, the estimators obtained by maximizing the Gaussian likelihood function may be interpreted as quasi-ML estimators.

The identification of B_0 or B_0^{-1} , in principle, may be assessed based on the asymptotic properties of the diagonal elements of Λ_2 . The null hypothesis of no identification is equivalent to

$$\mathbb{H}_0 : \lambda_{k2} = \lambda_{l2}$$

for all $k, l \in \{1, \dots, K\}$ with $k \neq l$. General asymptotically valid tests are currently not available. As in the previous models, if the identification conditions are satisfied and all diagonal elements of Λ_2 are distinct, any further restrictions imposed on B_0^{-1} become overidentifying and can be tested using the LR test discussed earlier.

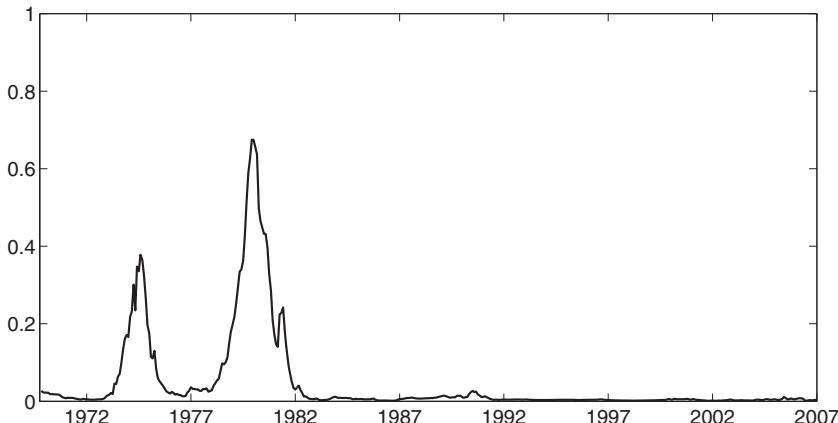


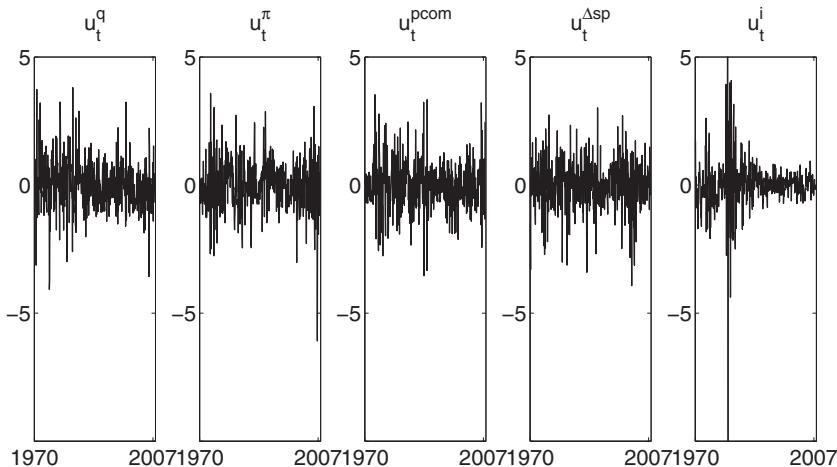
Figure 14.1. Transition function of ST-VAR(3) model with transition variable $s_t = \pi_{t-2}$.

Source: Lütkepohl and Netšunajev (2015).

Empirical Examples. We consider an example from Lütkepohl and Netšunajev (2015) who investigate the interdependence between U.S. monetary policy and stock prices. They reconsider a study of Bjørnland and Leitemo (2009). Let $y_t = (q_t, \pi_t, pcom_t, \Delta sp_t, i_t)'$, where q_t is the linearly detrended log of an industrial production index, π_t is the annual change in the log of consumer prices scaled by 100, $pcom_t$ is the annual change in the log of the World Bank (non energy) commodity price index scaled by 100, Δsp_t is the monthly returns of the real S&P500 stock price index deflated by the consumer price index to measure the real stock prices, and i_t is the federal funds rate.

Lütkepohl and Netšunajev (2015) use monthly data for the period 1970m1-2007m6 and fit an ST-VAR(3) model with a logistic transition function and two different transition variables. We focus on the results obtained when lagged inflation is used as the transition variable. More precisely, the transition variable is $s_t = \pi_{t-2}$. Lütkepohl and Netšunajev justify this choice by observing that inflation is an important driving force of monetary policy and show that model selection criteria favor the second lag over the first.

They estimate the transition function shown in Figure 14.1 which indicates two periods of higher volatility, one in the mid-1970s and the other one around 1980. Clearly, when the transition function is close to zero, the error variance is close to Σ_1 . For the periods in the mid-1970s and around 1980, the error variance is a convex combination of Σ_1 and Σ_2 with higher weight on the latter matrix for larger values of the transition function. It is also apparent in Figure 14.1 that the volatility in the two periods of higher volatility is quite different. The weights assigned to Σ_1 and Σ_2 are quite different in the two periods. This example shows the flexibility of this simple model.



(a) Standardized residuals of VAR(3) model

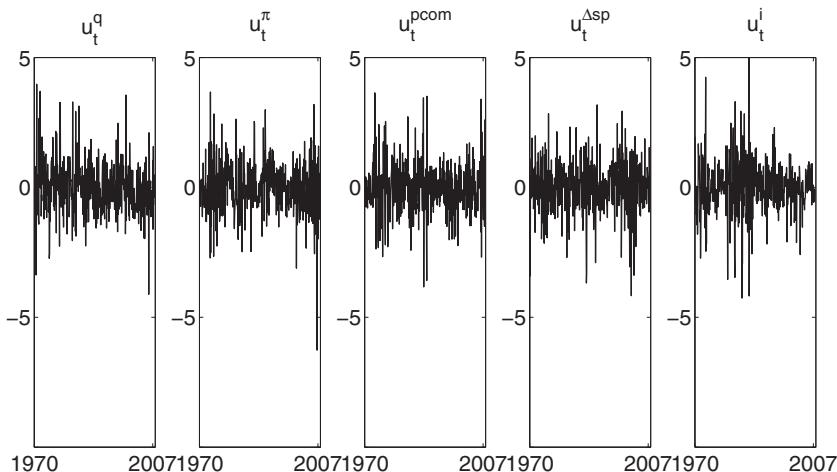
(b) Standardized residuals of ST-VAR(3) model, $s_t = \pi_{t-2}$

Figure 14.2. Residuals of the VAR(3) and ST-VAR(3) models.

Source: Lütkepohl and Netšunajev (2015).

In Figure 14.2 the standardized residuals of a reduced-form VAR(3) model that does not account for heteroskedasticity and the reduced-form ST-VAR(3) model are shown. The residuals are standardized by dividing them by the estimated standard deviation. For example, for the ST-VAR(3) model the k^{th} estimated residual, \hat{u}_{kt} , is divided by the square root of the k^{th} diagonal element of the estimated Σ_t covariance matrix. The standardized residuals of

Table 14.4. *Estimates of Relative Variances of Structural ST-VAR(3) Model with Transition Variable $s_t = \pi_{t-2}$*

Parameter	Estimate	Std.Dev.
λ_{12}	0.899	0.957
λ_{22}	2.739	1.608
λ_{32}	4.176	2.357
λ_{42}	8.091	3.108
λ_{52}	299.562	72.414

Source: Extracted from Table 2 of Lütkepohl and Netšunajev (2015).

the ST-VAR(3) model are seen to be overall more homogeneous throughout the sample than those of the standard VAR(3) model. Thus, the smooth-transition model captures the changes in the volatility at least to some extent.

Since the relative variances in Λ_2 are of central importance for the identification of the shocks, they are presented in Table 14.4 in increasing order. Except for the first element, all the estimated λ_{k2} are clearly larger than 1. Hence, the regime associated with Σ_2 represents a high volatility regime, confirming our previous interpretation of the transition function.

Full identification of all shocks by heteroskedasticity is achieved if all diagonal elements of Λ_2 are distinct. Based on the results in Table 14.4, it is difficult to argue that this condition is satisfied. Thus, we may not have a fully identified set of shocks. We may still use this model to test the validity of conventional identifying restrictions, however, because, under the null that these restrictions are valid, the model will be overidentified and the LR test will retain its asymptotic χ^2 distribution. The only difference is that the asymptotic distribution of the test statistic may have fewer degrees of freedom.

Bjørnland and Leitemo (2009) placed the stock price and monetary policy shocks last in the vector of structural shocks such that

$$w_t = (w_{1t}, w_{2t}, w_{3t}, w_{4t}^{sp}, w_{5t}^m)',$$

where w_{4t}^{sp} and w_{5t}^m denote the stock market and monetary shocks, respectively. They used a combination of short-run and long-run identifying restrictions for the structural shocks. More precisely, they imposed the following restrictions on the matrix of impact effects, B_0^{-1} , and the matrix of long-run effects, Υ :

$$B_0^{-1} = \begin{bmatrix} * & 0 & 0 & 0 & 0 \\ * & * & 0 & 0 & 0 \\ * & * & * & 0 & 0 \\ * & * & * & * & * \\ * & * & * & * & * \end{bmatrix} \quad \text{and} \quad \Upsilon = \begin{bmatrix} * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & * \\ * & * & * & * & 0 \\ * & * & * & * & * \end{bmatrix}. \quad (14.3.16)$$

Table 14.5. *Tests for Identifying Restrictions in Structural ST-VAR Models*

\mathbb{H}_0	\mathbb{H}_1	df	LR statistic	p-value
R1	unrestricted B_0^{-1} , Υ	10	25.811	0.004
R1	R2	1	7.494	0.006

Source: Extracted from Table 3 of Lütkepohl and Netšunajev (2017).

As before, asterisks indicate unrestricted elements and zeros denote elements restricted to zero. Thus, the stock market and monetary shocks are both assumed to have no impact effect on industrial production (q_t), inflation (π_t), and commodity prices ($pcom_t$). Moreover, the zero entry in the last column of Υ means that monetary shocks have no long-run effects on the stock market. Since the first three shocks are not of interest for their analysis, Bjørnland and Leitemo (2009) identified them arbitrarily by the zero restrictions in the second and third columns of B_0^{-1} . Their identification does not have an impact on the last two shocks which are of central importance for the analysis (see Chapter 10).

Following Lütkepohl and Netšunajev (2015), we focus on the set of restrictions:

R1: Bjørnland-Leitemo identification with restrictions as in (14.3.16).

R2: Only B_0^{-1} restricted as in (14.3.16),

where the restrictions in R2 are a subset of the restrictions in R1. They use LR tests to test R1 against an unrestricted ST-VAR model and against R2. The *p*-values in Table 14.5 are smaller than 5%. The first test shows that the Bjørnland-Leitemo restrictions are rejected in favor of an unrestricted model. The second test reveals that imposing the additional long-run restriction on Υ is problematic.

Despite the limited identifying information from heteroskedasticity, both null hypotheses are rejected. Of course, in that situation it is not justified to use these restrictions for impulse response analysis. On the other hand, using the shocks obtained by heteroskedasticity is not justified either, because, as we have seen before, they are not fully identified and because they have no natural economic interpretation. Thus, in this model it is difficult to disentangle the effects of stock market and monetary shocks. The only firm conclusion is that the previously found results have little empirical support when changes in volatility are taken into account.

14.3.4 Structural VAR Models with GARCH Errors

Yet another approach to modeling the conditional volatility of the VAR errors relies on a multivariate GARCH process for the VAR innovations. The

GARCH model is relevant for empirical work, given that the conditional heteroskedasticity of many economic and financial time series is well approximated by GARCH processes. Unlike models of discrete volatility changes, the GARCH model usually does not provide a clear partitioning of the sample period into periods of high and low volatility.

Model. In this setup, the structural shocks are assumed to be orthogonal and their conditional variances are modeled by individual GARCH(1, 1) processes. It is assumed that the conditional covariance matrix of u_t , given information up to period $t - 1$, \mathcal{F}_{t-1} , is

$$\mathbb{E}(u_t u_t' | \mathcal{F}_{t-1}) = B_0^{-1} \Sigma_{t|t-1} B_0^{-1'}, \quad (14.3.17)$$

where $\Sigma_{t|t-1} = \text{diag}(\sigma_{1,t|t-1}^2, \dots, \sigma_{K,t|t-1}^2)$ is a diagonal matrix. The individual conditional variances of the structural shocks $w_t = B_0 u_t$ are assumed to have a GARCH(1,1) structure of the form

$$\sigma_{k,t|t-1}^2 = (1 - \gamma_k - g_k) + \gamma_k w_{k,t-1}^2 + g_k \sigma_{k,t-1|t-2}^2, \quad k = 1, \dots, K, \quad (14.3.18)$$

where $\gamma_k, g_k \geq 0$. Higher-order GARCH processes can be considered in principle. This is typically not done in practice, however. The GARCH processes in (14.3.18) are set up such that the unconditional variances of the structural shocks are 1. Hence, $\mathbb{E}(w_t w_t') = I_K$ and the unconditional covariance matrix of the reduced-form errors is

$$\mathbb{E}(u_t u_t') = \Sigma_u = B_0^{-1} B_0^{-1'}.$$

A multivariate GARCH model of this type was proposed by van der Weide (2002) under the name of generalized orthogonal GARCH (GO-GARCH). It has been used by Normandin and Phaneuf (2004) and others for structural VAR analysis. There are also studies using the VAR-GARCH approach based on other types of GARCH models (e.g., Weber 2010; Strohsal and Weber 2015). We focus on the GO-GARCH setup because it has advantages in specifying and testing the identifying restrictions.

Identification. Sentana and Fiorentini (2001) and Milunovich and Yang (2013) provide identification conditions for this type of model. Let G be a $T \times K$ matrix with k^{th} column consisting of the conditional variances of the k^{th} structural shock, $(\sigma_{k,1|0}^2, \dots, \sigma_{k,T|T-1}^2)'$. Sentana and Fiorentini (2001) show that B_0^{-1} is identified up to column signs and permutations, if G has full column rank. This condition is easy to verify in the unlikely case that the true conditional variances are known.

Milunovich and Yang (2013) show that, equivalently, identification requires at least $K - 1$ of the GARCH processes being nontrivial in that $\gamma_k \neq 0$ for

at least $K - 1$ of the K processes in (14.3.18). In other words, identification requires sufficient heterogeneity in the conditional variances. To achieve full identification, at most one of the structural shocks may be homoskedastic. The latter condition may be investigated by formal statistical tests.

Since the GARCH structure offers additional identifying information, one may even give up the requirement of instantaneously uncorrelated shocks and replace it by other assumptions, as discussed in Section 14.4.2. For example, Weber (2010) assumes constant conditional correlations instead.

Estimation and Inference. To allow explicitly for the possibility that only $r < K$ structural shocks have a nontrivial GARCH(1, 1) structure, we write the volatility model as

$$u_t = B_0^{-1} \begin{bmatrix} \Lambda_{t|t-1}^{1/2} & 0 \\ 0 & I_{K-r} \end{bmatrix} \varepsilon_t, \quad (14.3.19)$$

where $\varepsilon_t \stackrel{iid}{\sim} (0, I_K)$ and

$$\Lambda_{t|t-1} = \begin{bmatrix} \sigma_{1,t|t-1}^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_{r,t|t-1}^2 \end{bmatrix} \quad (14.3.20)$$

is an $r \times r$ diagonal matrix with nontrivial univariate GARCH(1,1) processes on the diagonal. Thus, the structural shocks are

$$w_t = \begin{bmatrix} \Lambda_{t|t-1}^{1/2} & 0 \\ 0 & I_{K-r} \end{bmatrix} \varepsilon_t$$

and, conditionally on \mathcal{F}_{t-1} , the reduced-form errors have a distribution with zero mean and conditional covariance matrix

$$\Sigma_{t|t-1} = B_0^{-1} \begin{bmatrix} \Lambda_{t|t-1} & 0 \\ 0 & I_{K-r} \end{bmatrix} B_0^{-1'}$$

Postulating a normal distribution for ε_t , Lanne and Saikkonen (2007) show how the log-likelihood function can be set up such that ML estimation of the model becomes feasible. They use a polar decomposition of B_0^{-1} defined as a decomposition such that $B_0^{-1} = CR$, where C is a symmetric, positive definite $K \times K$ matrix and $R = [R_1, R_2]$ is an orthogonal $K \times K$ matrix partitioned such that R_1 is $K \times r$ and R_2 is $K \times (K - r)$. Note that

$$\Sigma_u = B_0^{-1} B_0^{-1'} = CRR'C' = CC.$$

Thus, C is the unique square root matrix of Σ_u . Lanne and Saikkonen (2007) decompose the conditional covariance matrices as

$$\Sigma_{t|t-1} = \Sigma_u + CR_1(\Lambda_{t|t-1} - I_r)R_1'C$$

and show that

$$u_t' \Sigma_{t|t-1}^{-1} u_t = u_t' \Sigma_u^{-1} u_t + u_t' C^{-1} R_1 (\Lambda_{t|t-1}^{-1} - I_r) R_1' C^{-1} u_t.$$

The Gaussian log-likelihood function is

$$\log l = \sum_{t=1}^T \log f_{t|t-1}(u_t).$$

Using the previously given decomposition, the conditional densities have the form

$$\begin{aligned} f_{t|t-1}(u_t) &= (2\pi)^{-K/2} \det(\Sigma_{t|t-1})^{-1/2} \exp\left(-\frac{1}{2} u_t' \Sigma_{t|t-1}^{-1} u_t\right) \\ &= (2\pi)^{-K/2} \det(\Sigma_u)^{-1/2} \exp\left(-\frac{1}{2} u_t' \Sigma_u^{-1} u_t\right) \prod_{k=1}^r \sigma_{k,t|t-1}^{-1} \\ &\quad \times \exp\left(-\frac{1}{2} u_t' C^{-1} R_1 (\Lambda_{t|t-1}^{-1} - I_r) R_1' C^{-1} u_t\right). \end{aligned}$$

If there are just r nontrivial GARCH components, the matrix R_2 from the polar decomposition of B_0^{-1} is not identified. The log-likelihood does not depend on R_2 , however, but only on the identified parameters $v, A_1, \dots, A_p, C, R_1$, and the GARCH parameters γ_k, g_k ($k = 1, \dots, r$). Hence, for given r , the model can be estimated.

Of course, estimation of high-dimensional multivariate GARCH models by iterative optimization is not easy. In the present situation, the log-likelihood function factors into a component that is a function of C , and a component that is a function of C, R_1 and $\Lambda_{t|t-1}$. Therefore the estimation can be broken down into two steps. First, C is estimated as the square root matrix of an estimator of Σ_u , for example, based on the residuals from the LS estimation of the VAR part. In the second step the rows of R_1 and the GARCH equation parameters are estimated conditional on the estimated C and separately for $k = 1, \dots, r$. Each equation $k+1$ is estimated conditional on the previously estimated equation k . Thus, effectively only univariate GARCH models have to be estimated in this step. Details of this procedure can be found in Section 4 of Lanne and Saikkonen (2007). This approach results in inefficient estimators, but can be used to obtain initial values for a full likelihood maximization.

Clearly, the choice of r is crucial in this procedure. It is also important for the identification of the structural shocks. If $r < K - 1$ the matrix B_0^{-1} and, hence, the structural shocks w_t are not fully identified by the GARCH structure. Since in many macroeconomic studies variables are involved that may not be conditionally heteroskedastic, it cannot be taken for granted that r is at least equal to $K - 1$. In other words, it is important to have formal statistical

procedures for exploring the number of nontrivial GARCH components that drive the volatility changes in the errors.

Suitable tests based on proposals by Lanne and Saikkonen (2007) are discussed in Lütkepohl and Milunovich (2016). For a given $r_0 < K$ the following pair of hypotheses is tested:

$$\mathbb{H}_0 : r = r_0 \quad \text{versus} \quad \mathbb{H}_1 : r > r_0.$$

Test statistics may be constructed as follows. Under \mathbb{H}_0 , the $K \times r_0$ submatrix consisting of the first r_0 columns of B_0^{-1} can be estimated consistently. Note also that

$$B_0 = R'C^{-1} = \begin{bmatrix} R'_1 \\ R'_2 \end{bmatrix} C^{-1}.$$

Thus, the last $K - r$ components of w_t are $w_{2t} = R'_2 C^{-1} u_t$.

Lanne and Saikkonen (2007) propose to estimate R_2 as an orthogonal complement of an estimator \widehat{R}_1 of R_1 . More precisely,

$$\widehat{R}_2 = \widehat{R}_{1\perp} (\widehat{R}'_{1\perp} \widehat{R}_{1\perp})^{-1/2}.$$

Clearly, since R_2 is not identified, this estimator may not be consistent for R_2 , but estimates some linear transformation of R_2 . Thus, the estimator

$$\widehat{w}_{2t} = \widehat{R}'_2 \widehat{C}^{-1} \widehat{u}_t,$$

is not necessarily an estimator of w_{2t} , but of some linear transformation of w_{2t} . However, if w_{2t} is not driven by GARCH components, the same is true of any linear transformation. Hence, tests for remaining GARCH in w_{2t} can be based on \widehat{w}_{2t} .

Under \mathbb{H}_0 , the last $K - r_0$ components of w_t do not exhibit conditional heteroskedasticity. Therefore Lanne and Saikkonen (2007) propose test statistics based on the autocovariances of the univariate mean-adjusted sum of squared components of \widehat{w}_{2t} ,

$$\xi_t = \widehat{w}'_{2t} \widehat{w}_{2t} - T^{-1} \sum_{i=1}^T \widehat{w}'_{2i} \widehat{w}_{2i},$$

and a related vector quantity consisting of the mean-adjusted squares and cross-products of the \widehat{w}_{2t} components,

$$\vartheta_t = \text{vech}(\widehat{w}_{2t} \widehat{w}'_{2t}) - T^{-1} \sum_{i=1}^T \text{vech}(\widehat{w}_{2i} \widehat{w}'_{2i}).$$

The test statistics are

$$Q_1(H) = T \sum_{h=1}^H [\tilde{\gamma}(h)/\tilde{\gamma}(0)]^2, \tag{14.3.21}$$

where

$$\tilde{\gamma}(h) = T^{-1} \sum_{t=h+1}^T \xi_t \xi_{t-h},$$

and

$$Q_2(H) = T \sum_{h=1}^H \text{tr}[\tilde{\Gamma}(h)' \tilde{\Gamma}(0)^{-1} \tilde{\Gamma}(h) \tilde{\Gamma}(0)^{-1}], \quad (14.3.22)$$

where

$$\tilde{\Gamma}(h) = T^{-1} \sum_{t=h+1}^T \vartheta_t \vartheta_{t-h}' \quad \text{for } h = 0, 1, \dots$$

Lanne and Saikkonen (2007) show that $Q_1(H)$ and $Q_2(H)$ have asymptotic χ^2 distributions with H and $H(K - r_0)^2(K - r_0 + 1)^2/4$ degrees of freedom, respectively, under the null hypothesis.

Lütkepohl and Milunovich (2016) also propose a closely related LM type test based on the auxiliary model

$$\eta_t = \delta_0 + D_1 \eta_{t-1} + \dots + D_H \eta_{t-H} + e_t, \quad (14.3.23)$$

where $\eta_t = \text{vech}(\widehat{w}_{2t} \widehat{w}_{2t}')$ and e_t is an error term. The null hypothesis $\mathbb{H}_0 : r = r_0$ is equivalent to

$$\mathbb{H}_0 : D_1 = \dots = D_H = 0$$

and the corresponding LM statistic is

$$LM(H) = \frac{1}{2} T(K - r_0)(K - r_0 + 1) - T \text{tr}[\tilde{\Sigma}_e \tilde{\Gamma}(0)^{-1}], \quad (14.3.24)$$

where $\tilde{\Sigma}_e$ is the estimated residual covariance matrix from model (14.3.23). The statistic $LM(H)$ has the same asymptotic χ^2 distribution as $Q_2(H)$.

Lütkepohl and Milunovich (2016) compare the tests based on $Q_1(1)$, $Q_2(1)$, and $LM(1)$ in a simulation study and find that the performance of the tests very much depends on the persistence of the underlying GARCH processes (measured by $\gamma_k + g_k$) and the sample size. None of the tests uniformly dominates its competitors in terms of size and power. In fact, even for moderate sample sizes the power of the tests is very low, if the underlying GARCH processes are not very persistent. This result implies in particular that sequential testing procedures for determining the number of nontrivial GARCH components r are not very reliable and tend to end up with estimates of r that are lower than the true r_0 . A plausible sequential testing procedure for the true number of GARCH components, r_0 , tests null hypotheses $\mathbb{H}_0 : r = i$ for $i = 1, 2, \dots$, until the first hypothesis is rejected. One may be tempted to test a reverse sequence of hypotheses, $\mathbb{H}_0 : r = j$, $j = K - 1, K - 2, \dots$. Such a

test is problematic, however, because the true rank may be lower than the one specified under \mathbb{H}_0 . A test of full identification can be set up by specifying $\mathbb{H}_0 : r = K - 2$. If that null hypothesis is rejected against $\mathbb{H}_1 : r > K - 2$, full identification is supported by the data because this indicates that the number r of nontrivial GARCH components is at least $K - 1$.

A drawback of the VAR-GARCH model is that the log-likelihood function is difficult to maximize globally for large models. The procedure of Lanne and Saikkonen (2007) described earlier can be applied, however, to obtain feasible first-stage estimates which can be used as starting values in the maximization.

Empirical Examples. Bouakez and Normandin (2010) examine the importance of U.S. monetary policy shocks for exchange rate dynamics using structural VAR-GARCH models. They consider eight-dimensional models for the U.S. and the G7 countries. They assume that seven univariate GARCH components drive the multivariate GARCH process and that the GARCH models fully identify the structural shocks. Based on this assumption, they conduct a number of tests of conventional identifying restrictions. Lütkepohl and Milunovich (2016) in turn test the statistical identifying assumptions underlying the VAR-GARCH model used by Bouakez and Normandin (2010).

For illustrative purposes we focus on one example from Lütkepohl and Milunovich (2016) based on an 8-dimensional model for the United States and Canada. Let $y_t = (q_t, p_t, pcom_t, nbr_t, tr_t, i_t, d_t, ex_t)'$, where q_t is the log of a U.S. industrial production index, p_t is the log of the U.S. consumer price index, $pcom_t$ is the log of a world export commodity price index, nbr_t is the log of U.S. nonborrowed reserves, tr_t is the log of U.S. total reserves, i_t is the federal funds rate, d_t is the difference between the Canadian short-term interest rate and the U.S. three-months Treasury bill rate, and ex_t is the log of the exchange rate (U.S. dollars per one Canadian dollar).

In testing conventional identifying restrictions, the question arises of how many GARCH components drive the volatility changes in this system. Financial variables such as i_t and ex_t are likely to exhibit GARCH dynamics, but some other variables may not, so identification cannot be taken for granted. Tests of the validity of the GARCH-based identification may be conducted based on the statistics (14.3.21), (14.3.22), and (14.3.24).

Using monthly data for the period 1982m11–2004m10, Lütkepohl and Milunovich (2016) obtain the test results presented in Table 14.6. Of particular interest is testing

$$\mathbb{H}_0 : r = 6 \quad \text{versus} \quad \mathbb{H}_1 : r > 6.$$

Rejecting that null hypothesis would be strong evidence of the full statistical identification of the shocks by the GARCH structure. However, the p -values of all three test statistics are greater than 0.6 and, hence, do not reject \mathbb{H}_0 at

Table 14.6. *p*-Values for the Identification Tests Applied to U.S./Canadian Data from Bouakez and Normandin (2010)

Test	$\mathbb{H}_0 : r = 1$	$r = 2$	$r = 3$	$r = 4$	$r = 5$	$r = 6$	$r = 7$
$LM(1)$	0.000	0.027	0.027	0.150	0.390	0.689	0.528
$Q_1(1)$	0.174	0.183	0.046	0.139	0.214	0.617	0.445
$Q_2(1)$	0.000	0.024	0.029	0.127	0.335	0.632	0.448

Source: Extracted from Table 9 of Lütkepohl and Milunovich (2016).

conventional significance levels. In fact, in Table 14.6 the last null hypothesis that can be rejected at a 5% significance level is $\mathbb{H}_0 : r = 3$. In other words, the tests provide evidence that there may be four GARCH components driving the eight variables. Even if one accounts for low power, it is difficult to argue that there is evidence for more than five GARCH components. Thus, we do not have support for the hypothesis of a fully identified model. Consequently, tests of conventional identifying restrictions have to be interpreted with caution. If such restrictions are rejected within a model with four or five GARCH components this is, of course, strong evidence against their validity. A failure to reject conventional exclusion restrictions, in contrast, could simply be due to not having enough identifying restrictions.

In another example of the structural VAR-GARCH approach, Normandin and Phaneuf (2004) compare identification schemes for monetary policy shocks based on U.S. monthly data for 1982m11–1998m12. They find that some of the assumptions that have been used in conventional analyses are rejected in the VAR-GARCH setup. However, they do not use formal identification tests and, hence, their results have to be interpreted with caution.

There are a number of other studies that utilize the VAR-GARCH model for structural analysis. Examples are Bouakez, Essid, and Normandin (2013) who study the effects of monetary policy shocks on the stock market and Bouakez, Chihi, and Normandin (2014) who examine the effects of fiscal policy shocks on the U.S. economy.

14.4 Alternative Approaches Using Heteroskedasticity

The approach of using heteroskedasticity in the errors of structural VAR models may be extended in several directions. First, as mentioned before, changes in the volatility of the shocks may coincide with changes in the impact effects of the shocks. Second, heteroskedasticity can also be used to dispense with the assumption of instantaneously uncorrelated shocks, albeit at the cost of the resulting shocks no longer being structural in the sense specified in Chapter 1. These two extensions are discussed next.

14.4.1 Time-Varying Instantaneous Effects

Bacchicocchi, Castelnuovo, and Fanelli (2013) and Bacchicocchi and Fanelli (2015) consider the possibility of time-varying instantaneous effects of the shocks. If there are two volatility states with covariance matrices Σ_1 and Σ_2 , respectively, then the decomposition

$$\Sigma_1 = GG' \quad \text{and} \quad \Sigma_2 = (G + C)(G + C)' \quad (14.4.1)$$

may be used to obtain the structural shocks from $w_t = G^{-1}u_t$, if a given t corresponds to the first volatility regime, and, alternatively, from $w_t = (G + C)^{-1}u_t$, if that t belongs to the second volatility regime. Thus, G and $G + C$ are the impact effects matrices corresponding to the first and second regime, respectively. In other words, C represents the change in the impact effects of the structural shocks from regime 1 to regime 2.

Clearly, the relations in expression (14.4.1) are insufficient to identify all the elements of the two $K \times K$ matrices G and C . Therefore further identifying assumptions or restrictions are required. For example, both matrices may be assumed to be lower triangular. This problem is not alleviated if there are more than two volatility regimes. If the impact effects are allowed to vary freely across the volatility regimes, heteroskedasticity is not helpful for the identification of the shocks, because each additional volatility regime increases the number of restrictions required for full identification.

If such assumptions are easy to justify in some context, then using this approach may be useful. Note, however, that a change in the impact effects results in more general changes in the impulse response functions, even if the reduced-form VAR parameters are not regime dependent. Such a model is quite restrictive, because it will often be difficult to argue that the impact effects of the shocks are regime dependent, but the reduced-form parameters apart from the covariances do not change. If the reduced-form parameters are allowed to vary more generally, then the model becomes a time-varying coefficient model. Such models are discussed in more detail in Chapter 18. They are not considered further in the current chapter.

14.4.2 Correlated Shocks

Weber (2010) and Strohsal and Weber (2015) suggest that allowing for instantaneously correlated shocks in the structural model makes sense when working with low-frequency macroeconomic data. They note that if there are enough volatility regimes, the assumption of uncorrelated shocks can be replaced by some other assumption that limits the number of structural parameters. For example, if there are M volatility regimes, one could postulate a regime-invariant correlation structure and decompose the covariance

matrices as

$$\Sigma_m = G\Lambda_m^{1/2} R \Lambda_m^{1/2} G', \quad m = 1, \dots, M, \quad (14.4.2)$$

where R is a correlation matrix that is invariant across the volatility regimes and the Λ_m are again diagonal matrices with positive diagonal elements. If the structural shocks are obtained as $w_t = G^{-1} u_t$, they have covariance matrices

$$\mathbb{E}(w_t w_t') = G^{-1} \Sigma_m G^{-1'} = \Lambda_m^{1/2} R \Lambda_m^{1/2},$$

if t belongs to the m^{th} volatility regime. Thus, the diagonal elements of Λ_m are seen to be the variances of w_t in regime m .

If M is large enough and the covariance matrices satisfy suitable conditions, R and G can be recovered from equation (14.4.2), in which case the shocks are statistically identified. A drawback of this setup is that a decomposition such as that in (14.4.2) may not be easy to justify from an economic point of view and is unhelpful in answering the type of questions commonly analyzed using structural VAR models. Therefore, we do not elaborate on these models.

14.5 Identification by Non-Gaussianity

Identification by heteroskedasticity uses statistical properties of the data for identification. There have also been attempts to use other statistical properties of the data for this purpose. Notably, if the errors, u_t , of the reduced-form VAR process are not Gaussian, this feature can be used for identification. Studies that consider this device include Siegfried (2002), Lanne and Lütkepohl (2010), Moneta, Entner, Hoyer, and Coad (2013), Gourieroux and Monfort (2014), Lanne, Meitz, and Saikkonen (2017), and Herwartz (2015).

14.5.1 Independent Shocks

Gourieroux and Monfort (2014) state a mathematical result that can be used for uniquely identifying a set of structural shocks in a stationary VAR model with iid shocks. Related results are also provided in Lanne, Meitz, and Saikkonen (2017). Suppose $w = (w_1, \dots, w_K)'$ and $w^* = (w_1^*, \dots, w_K^*)'$ are both K -dimensional random vectors with independent components, and that at most one component is normally distributed, whereas all other components have nonnormal distributions. If there exists a nonsingular matrix C such that $w^* = Cw$, then $w_k^* = \gamma_k w_{\pi(k)}$, $k = 1, \dots, K$. Here $\pi(\cdot)$ denotes a permutation of the numbers $1, \dots, K$. In other words, the k^{th} component of w^* is a multiple of one of the components of w . Thus, if all components of w and w^* have unit variance such that both vectors have an identity covariance matrix, then C must be a permutation matrix that merely reorders the components of w , possibly changing their signs in the process.

This result implies that in a non-Gaussian environment, identification can be obtained by insisting that the structural shocks be stochastically independent rather than just uncorrelated. More precisely, if in a VAR analysis the structural shocks are obtained by a linear transformation of the reduced-form errors, $w_t = B_0 u_t$, and the structural shocks are mutually stochastically independent, have variance one, and at most one of them is normally distributed, then B_0 is unique except for row permutation and row sign changes. This follows by noting that any other K -dimensional random vector with independent components that is obtained by a linear transformation of w_t must be just a reordering of the components of w_t , possibly with reversed sign. Hence, the only linear transformations that preserve the independence of the components are of the form $\mathcal{P}B_0$, where \mathcal{P} is a permutation matrix that permutes the rows of B_0 . Hence, \mathcal{P}^{-1} is also a permutation matrix and $B_0^{-1}\mathcal{P}^{-1}$ is the matrix B_0^{-1} with permuted columns. In other words, the matrix of impact effects is unique apart from column permutations and column sign changes.

Of course, this discussion assumes that there exists a linear transformation of the reduced-form residuals that results in a set of independent structural shocks. As illustrated later in this chapter, there are counterexamples of non-Gaussian processes in which no linear transformation results in independent shocks. Thus, we need to consider both linear and nonlinear transformations. The question then arises of how that transformation can be found. Siegfried (2002) and Gouriéroux and Monfort (2014) point out that a method called independent component analysis (ICA) can be helpful in this context. Further references include Jutten and Herault (1991), Hyvärinen, Karhunen, and Oja (2001), and Stone (2004). Moneta, Entner, Hoyer, and Coad (2013) use this method for identifying mutually independent shocks.

An alternative procedure that restricts attention to linear transformations is proposed by Herwartz (2015). The premise is that structural shocks may be obtained by a linear transformation of the reduced-form shocks. Herwartz examines alternative rotations of the orthogonalized reduced-form shocks. He then tests the mutual independence of the elements of the implied vector of structural shocks for each rotation. Finally, he chooses the rotation that maximizes the p -value of the test. Using the rotation with the largest p -value amounts to choosing the structural errors that are least dependent. A possible drawback of this approach is that rotating uncorrelated shocks does not necessarily lead to independent innovations. In other words, the independent components may not be linear transformations of the uncorrelated residuals. An example of a model where independent errors cannot be determined by linear transformations of the reduced-form errors is considered in Section 14.5.2.

Herwartz and Plödt (2016) use the methodology proposed by Herwartz (2015) to investigate the dynamics in the market for crude oil. As in Kilian (2009) and Kilian and Murphy (2012), the variables are the change in global crude oil production, a measure of global real economic activity, and the real

price of oil, but the data definitions and sample period differ from those used in Kilian (2009). Exploiting the nonnormality of the data and applying the procedure of Herwartz (2015), they find shocks that imply impulse responses similar to those in Kilian (2009). Hence, they label the shocks accordingly as an oil supply shock, an aggregate demand shock, and an oil-market specific demand shock.

Yet another procedure for exploiting non-Gaussianity has been discussed in Lanne, Meitz, and Saikkonen (2017). They propose an ML estimator of the transformation matrix that delivers independent errors w_t , conditional on additional assumptions on the distribution of u_t . For identification, Lanne et al. do not assume that the marginal error processes w_{kt} are serially independent, but just uncorrelated. In other words, $\mathbb{E}(w_{kt} w_{k,t+j}) = 0$ for $j \neq 0$, but w_{kt} and $w_{k,t+j}$ may not be independent. Thereby, their model allows, for example, for GARCH in the marginal processes w_{kt} .

Of course, as in the discussion of identification by heteroskedasticity, the shocks obtained in this way are based on purely mathematical properties of the underlying distributions. They will not have any economic interpretation, in general, without additional economic information. However, this framework may, in principle, be used for testing conventional identifying restrictions from models that do not exploit the nonnormality of the data. In the absence of such tests, one may still be able to infer the economic interpretation of these estimates, if the impulse responses of interest obtained using this procedure match the impulse response estimates from the corresponding VAR model identified by conventional restrictions.

14.5.2 Uncorrelated Shocks

Lanne and Lütkepohl (2010) use a quite different approach to take advantage of the nonnormality of the observations. They assume that the reduced-form residuals have a mixture normal distribution, i.e.,

$$u_t \sim \begin{cases} \mathcal{N}(0, \Sigma_1) & \text{with probability } \gamma, \\ \mathcal{N}(0, \Sigma_2) & \text{with probability } 1 - \gamma, \end{cases} \quad (14.5.1)$$

where $0 < \gamma < 1$ is the mixing probability and $\Sigma_1 \neq \Sigma_2$. The class of normal mixture distributions is very broad and members of this class can approximate many continuous distributions. The mixture distribution in (14.5.1) implies that the reduced-form residuals, u_t , have mean zero and a covariance matrix

$$\mathbb{E}(u_t u_t') = \gamma \Sigma_1 + (1 - \gamma) \Sigma_2.$$

Using the decomposition $\Sigma_1 = B_0^{-1} B_0^{-1'}$ and $\Sigma_2 = B_0^{-1} \Lambda B_0^{-1'}$, where B_0^{-1} is a nonsingular $K \times K$ matrix and Λ is a diagonal matrix with positive diagonal elements, and choosing structural shocks $w_t = B_0 u_t$, implies that w_t has instantaneously uncorrelated components. Moreover, as discussed earlier, B_0^{-1}

is unique up to column sign changes and column permutations if the diagonal elements of Λ are all distinct. Hence, using this device, the structural shocks are identified completely analogously to the case of identification via heteroskedasticity. Actually, the model is a special case of the Markov-switching model considered in Section 14.3.2, where the state of the process in period t is independent of the states in the previous periods. In other words, if s_t denotes the Markov process that specifies the state of the system, then the transition probabilities are

$$\mathbb{P}(s_t = m) = \mathbb{P}(s_t = m | s_{t-1} = 1) = \cdots = \mathbb{P}(s_t = m | s_{t-1} = M).$$

For example, for our model with two states we obtain a transition matrix

$$\begin{bmatrix} \gamma & \gamma \\ 1 - \gamma & 1 - \gamma \end{bmatrix}.$$

It is worth contrasting this model with the model of Gourieroux and Monfort (2014), where independence of the structural shocks is used as the identifying device. If in the normal mixture model the structural shocks are chosen as $w_t = B_0 u_t$, they are instantaneously uncorrelated, but generally not independent. Since independence implies uncorrelatedness, it is tempting to think of them as independent as well. Notice, however, that the w_t have a normal mixture distribution,

$$w_t \sim \begin{cases} \mathcal{N}(0, I_K) & \text{with probability } \gamma, \\ \mathcal{N}(0, \Lambda) & \text{with probability } 1 - \gamma. \end{cases}$$

Clearly, if the state of one component of w_t is given, in general, this fact also determines the state of the other components. Hence, two components w_{kt} and w_{jt} are stochastically dependent. This example also illustrates that the assumption of independent structural errors is more restrictive than it may seem at first sight. In fact, the assumption that independent components can be obtained by a linear transformation of the reduced-form residuals is a very special case when considering nonnormal distributions.

The statistical analysis of the normal mixture model is straightforward. A VAR model is fitted by LS and the residuals are tested for nonnormality, as discussed in Chapter 2. If normality is rejected, the model with mixed normal residuals is fitted by ML. Since we assume a specific distribution, the likelihood function is readily available which makes ML estimation a natural choice. The diagonal elements of Λ are then investigated. If suitable tests suggest that they are all distinct, we have identified the structural shocks. Although such tests do not seem to be currently available, developing them may be possible.

The use of the mixed normal distribution is appealing, if the system consists of two different states that are characterized by two normal distributions and if the underlying shocks are mutually uncorrelated in either state. If instead

the mixture normal distribution is just seen as a description of a general non-normal distribution, then the specific choice of transformation matrix for deriving the structural shocks from the reduced-form errors is arbitrary. In the latter case, this approach is merely a technical device and not a technique for identifying economically interpretable structural shocks. Therefore the interpretation of the shocks is problematic.

Lanne and Lütkepohl (2010) illustrate the use of these models by two examples. One of them considers the system that we already used to illustrate the MS-VAR analysis in Section 14.3.2 consisting of U.S. real output growth, changes in the real interest rate, and real stock returns. The other example investigates interest rate linkages between the U.S. and Europe.

14.6 Discussion

In this chapter we studied the question of how statistical properties of VAR errors can deliver additional information that is sufficient for the identification of mutually uncorrelated, but not necessarily economically interpretable shocks. Our analysis focused primarily on the use of volatility changes in the data. We also briefly discussed data-based identification methods exploiting the non-Gaussianity of many economic time series.

Although mutually uncorrelated shocks identified by heteroskedasticity or non-Gaussianity alone have no economic interpretation without additional assumptions, they can be useful in assessing the validity of conventional identifying restrictions that cannot be tested against the data in standard structural VAR models. Conventional just-identifying restrictions in this setting become overidentifying restrictions that can be formally tested.

We reviewed a number of models for changes in the volatility of the data that are used in the literature. If volatility changes are known to occur at some point in time, volatility breaks can be modeled by extraneously imposing these volatility changes. This situation is rare in practice, however. It is more likely that statistical pretests are used to determine the dates of the volatility changes. In addition, often it is not clear whether the change in volatility is abrupt or smooth. In the latter case, it is preferable to work with models that allow the volatility to evolve endogenously.

We discussed three different models that endogenize the changes in the volatility of the VAR innovations: Markov-switching, smooth transition, and VAR-GARCH models. The MS-VAR model is particularly useful if there is a finite number of volatility regimes and if the transition between these regimes can be described by a Markov process whose parameters can be estimated from the data. Although this MS-VAR model is often economically and empirically plausible, its main drawback is that it is difficult to estimate when there are many variables, long VAR lags, or a large number of volatility states. An alternative model, for which robust estimation algorithms exist, describes the

change in volatility by a smooth transition function. Even if this model is based on two covariance states only, it can describe complex changes in volatility. Finally, we presented a VAR-GARCH model that allows the conditional variance of the errors to evolve over time, while preserving a constant unconditional variance. Estimation of this model can also be challenging for larger models.

If there is no prior knowledge on the structure of the volatility changes, it is difficult to decide between different volatility models. Using model selection criteria for that purpose may be a useful strategy. This approach has been employed by Lütkepohl and Netšunajev (2017), but there do not seem to be formal results on the properties of these criteria. In some cases, it may be advisable to consider different models and to investigate the robustness of the empirical results of interest.

Finally, we discussed how the non-normality of many economic time series may be used as a device for generating additional identifying information about the structural shocks. It can be shown that structural shocks being non-Gaussian implies the uniqueness of these shocks, provided that the structural shocks are assumed to be independent. We emphasized that such independent shocks, in general, cannot be obtained by a linear transformation of non-Gaussian reduced-form errors. There are alternative algorithms in the literature, but at this point no consensus has emerged on how to transform the reduced-form residuals in practice. An additional question that does not appear to have been discussed in the literature to date is how strong the degree of non-Gaussianity has to be for estimators derived in this way to be reliable.

15 Identification Based on Extraneous Data

Yet another approach to identifying structural VAR shocks is to rely on additional data not included among the VAR model variables. This chapter discusses two such approaches. The first approach relies on information from high-frequency futures prices, whereas the second approach relies on external exogenous instruments such as measures of exogenous OPEC oil supply shocks or the narrative measures of exogenous monetary policy shocks and exogenous fiscal policy shocks already discussed in Chapter 7.

15.1 Identification Based on High-Frequency Futures Prices

Discomfort with semistructural VAR models of monetary policy, in which the monetary policy shock is identified based on a recursive ordering of the model variables, has stimulated the development of yet another approach to identification that relies on high-frequency futures market data to identify monetary policy shocks. One motivation for this approach is that the sequences of policy shocks identified by recursive structural VAR models do not always correspond to common perceptions of when policy shocks occurred and indeed vary widely across models (see Rudebusch 1998).¹

The other motivation is that the reduced-form interest rate forecasts of conventional VAR models of monetary policy are difficult to reconcile with financial market measures of interest rate expectations. For example, Rudebusch (1998) documents a low correlation between the quarterly reduced-form prediction errors for the federal funds rate implied by conventional VAR

¹ Sims (1998) stresses that this finding is to be expected in a simultaneous equations model. For example, if we compare estimates of a model of demand and supply using alternative exogenous supply shifters (say, weather and insect density), then there is no reason for these instruments to generate the same sequence of supply shocks. This fact does not preclude them from accurately estimating the slope of the demand curve, however. Indeed, as Sims observes, models that imply very different sequences of monetary policy shocks still tend to agree on the response of the economy to a given monetary policy shock.

models of monetary policy and the quarterly changes in the expected federal funds rate implied by the prices of federal funds futures contracts, defined as

$$\hat{u}_t^{\Delta FFF} \equiv (FFR_i + FFR_{i+1} + FFR_{i+2} - FFF_{i-1}^1 \\ - FFF_{i-1}^2 - FFF_{i-1}^3)/3,$$

where quarter t contains months i , $i+1$, and $i+2$, FFR_i is the actual federal funds rate in month i , and FFF_i^j denotes the j -months ahead expected federal funds rate at the end of month i , as measured by the price of the federal funds futures contract. The financial market shocks statistically explain only between 10% and 25% of the variation in the quarterly VAR prediction errors, \hat{u}_t , based on data since 1988 when federal funds futures contracts were introduced.² Rudebusch interprets this evidence as suggesting that the reduced-form representation of VAR models of monetary policy is inherently misspecified, perhaps reflecting an informational deficiency of these models (see also Chapter 16).

This finding suggests that conventional low-dimensional VAR models of monetary policy are unable to capture the market's interest rate expectations. It also suggests that the timing of quarterly VAR models may be too coarse to separate monetary policy shocks from policy responses to contemporaneous events. These two concerns have spurred the development of an alternative approach to identifying monetary policy shocks in monthly VAR models based on high-frequency futures prices.

Whereas Rudebusch (1998) defines the monetary policy shock as the difference between the realized federal funds rate target and the expected federal funds rate derived from federal funds futures, more recent studies consider monetary policy shocks measured in this way potentially misleading for technical reasons related to the measurement of the realized federal funds rate (see Piazzesi and Swanson 2008). More importantly, these shocks will also be contaminated by risk premia in the futures market. As discussed in Chapter 7, the alternative of measuring monetary policy shocks based on the change in the daily federal funds futures rate around Federal Open Market Committee (FOMC) announcements, as first suggested by Kuttner (2001), tends to be more robust to the presence of risk premia in these contracts. This approach uses daily (or even intradaily) federal funds futures data to narrow the time interval around the FOMC announcement and assumes that risk premia do not change over such a small interval.

Initially, researchers constructed responses to such policy shocks from distributed-lag models, as reviewed in Chapter 7 (see, e.g., Cochrane and Piazzesi 2002). The idea of incorporating such measures of policy shocks into

² In addition, Rudebusch shows that the forecast errors from unrestricted VAR models are more variable than the futures price shocks. The latter problem may be overcome with the use of suitable priors when estimating the model (see Robertson and Tallman 2001).

structural VAR models was pioneered by Faust, Rogers, Swanson, and Wright (2003) and Faust, Swanson, and Wright (2004). This approach allows the user not only to incorporate information from financial markets into the VAR model but also to dispense with the exclusion restrictions used in recursively identified VAR models of monetary policy. A potential drawback of this approach is that the structural impulse responses are not longer point-identified, but only set identified.

15.1.1 A Set-Identified Approach

Faust, Swanson, and Wright (2004) consider a standard closed-economy VAR model of U.S. monetary policy using a benchmark model from Christiano, Eichenbaum, and Evans (1996, 1999). The model includes monthly observations for U.S. industrial production, the CPI, the smoothed change in an index of commodity prices, the federal funds rate, nonborrowed reserves, and total reserves. Unlike Christiano et al., they also incorporate information from the federal funds futures market.

Measuring the Policy Shock. Faust et al. assume that the change in the federal funds target rate on the days of Federal Open Market Committee (FOMC) meetings that was not anticipated by the federal funds futures market represents a monetary policy shock. Specifically, they treat the change in the futures rate on the day, on which a change in the Fed's target federal funds rate is announced, as a measure of the change in market expectations. This interpretation requires the risk premium to remain unchanged. Faust et al. further postulate that this change in expectations is due to the policy shock only. In other words, no other news move the market on that day and the policy announcement itself does not reveal information about other structural shocks.

Faust et al.'s procedure involves two key steps: In the first step, the surprise change in the target rate is measured by the change in the closing price of the federal funds contract at 3:00PM Eastern Standard Time, suitably scaled by the ratio

$$\frac{\text{days in month}}{\text{days left in month}}$$

to take account of how far into the current month the surprise occurs. Near the end of the month, this scaling factor is quite large.

Faust et al. then regress for all FOMC meeting dates the expected future change in the monthly federal funds rate on the monthly surprise change in the target rate. These regressions yield an estimate of the response of the monthly federal funds rate to a policy shock that can be scaled appropriately. In the second step, when estimating the VAR model, Faust et al. impose that the response of the monthly federal funds rate to a monetary policy shock in the monthly

VAR model must match the impulse response of the federal funds rate already estimated from the high-frequency data.

While these two steps are conceptually straightforward, the information from the futures market only set-identifies the responses in the structural VAR model, as explained shortly. This means that one must give up on point estimation of the structural responses and focus on confidence intervals instead, similar to classical inference in sign-identified VAR models.

In closely related work, Faust, Rogers, Swanson, and Wright (2003) extend this algorithm to allow for restrictions on the impulse responses of several model variables. They consider open economy VAR models for selected countries, for which quotes from futures markets are available at daily frequency for the domestic interest rate and the U.S. interest rate as well as the bilateral U.S. dollar exchange rate.

Estimation and Inference. Consider a K -dimensional reduced-form VAR model

$$A(L)y_t = u_t.$$

The corresponding structural MA representation of this model can be written as

$$y_t = \Phi(L)B_0^{-1}w_t,$$

where $u_t = B_0^{-1}w_t$. One of the structural shocks in w_t is assumed to be the monetary policy shock of interest. Without loss of generality, Faust et al. assume that this shock is the first shock and let δ denote the first column of B_0^{-1} . Then the set of response functions of the model variables to the policy shock is

$$\Phi(L)\delta = \sum_{j=0}^{\infty} \Phi_j \delta L^j.$$

This expression is a $K \times 1$ vector of lag polynomials. The coefficients of the k^{th} element of this vector trace out the responses of the k^{th} variable to the policy shock. Given that Φ_j can be computed from the reduced-form model, identifying the structural impulse responses only requires restricting the K elements of δ . There are two sets of restrictions. First, it is essential to normalize the impact response of the federal funds rate to a monetary policy shock. Faust et al. choose 25 basis points as this normalization. Second, in addition, sign and magnitude restrictions on the other elements of δ may be imposed based on a priori economic reasoning, as discussed later in this chapter.

The final requirement is that the response of the federal funds rate to the policy shock must match the corresponding response estimated from the futures market data. Denote the latter response at horizon $t + h$ by r_h^{FFF} ,

$h = 0, 1, \dots, K - 1$, where FFF denotes the federal funds forecast obtained from the futures market. By construction this means that

$$\phi_h^{FF} \delta = r_h^{FFF}, \quad h = 0, 1, \dots, K - 1,$$

where ϕ_h^{FF} denotes the row of Φ_h pertaining to the federal funds rate response. For example, if the federal funds rate is ordered last in the vector of observed variables, y_t , then ϕ_h^{FF} is the last row of Φ_h . After stacking these K equations, we obtain the set of restrictions

$$R\delta = r^{FFF}. \quad (15.1.1)$$

Here $R' = [\phi_0^{FF'}, \dots, \phi_{K-1}^{FF'}]$ so that R is $K \times K$ and $r^{FFF} = (r_0^{FFF}, \dots, r_{K-1}^{FFF})'$ is $K \times 1$. If R is of rank K , then there is a unique solution to this equation and $\delta = R^{-1}r^{FFF}$. A failure of this rank condition can be tested, allowing for the fact that R depends on the reduced-form estimate of the VAR model and r^{FFF} must be estimated from the futures price data. If R is rank deficient, as is typically the case, then there are many solutions for δ and the structural impulse responses are only set-identified by the restriction that $R\delta = r^{FFF}$, which greatly complicates inference, as we have already seen in Chapter 13.

Testing the Rank Condition. Following Faust, Swanson, and Wright (2004), consider testing $H_0 : \text{rank}(R) = l$ against $H_1 : \text{rank}(R) > l$. In other words, the null hypothesis is that the model is rank deficient. Let α denote the vector of reduced-form VAR parameters. Provided that $\sqrt{T}(\hat{\alpha} - \alpha) \xrightarrow{d} \mathcal{N}(0, \Sigma_\alpha)$, we can write R as a nonlinear function of α with consistent plug-in estimator \hat{R} . By the delta method, $\sqrt{T}(\text{vec}(\hat{R}) - \text{vec}(R)) \xrightarrow{d} \mathcal{N}(0, \Sigma_R)$, where

$$\Sigma_R = \frac{\partial \text{vec}(R)}{\partial \alpha'} \Sigma_\alpha \frac{\partial \text{vec}(R)'}{\partial \alpha}.$$

Hence, the test statistic is

$$T \min_{S \in \pi(l)} (\text{vec}(\hat{R}) - \text{vec}(S))' \hat{\Sigma}_R^{-1} (\text{vec}(\hat{R}) - \text{vec}(S)),$$

where

$$\hat{\Sigma}_R = \frac{\partial \text{vec}(\hat{R})}{\partial \alpha'} \hat{\Sigma}_\alpha \frac{\partial \text{vec}(\hat{R})'}{\partial \alpha}$$

is an estimator of Σ_R and

$$\pi(l) = \{S \mid \text{rank}(S) = l, S\delta = r^{FFF}\}$$

is the space of all conformable matrices S of rank l . Assuming that Σ_R has full rank, the test statistic by Theorem 1 in Cragg and Donald (1997) has a χ^2 null distribution. In practice, it is common to test for successively higher ranks l sequentially until some null cannot be rejected (or until all rank deficient models have been rejected).

In the empirical application of Faust et al. there are six futures contracts and six variables in the VAR model, so if $\text{rank}(R) = 6$, the model would be exactly identified. The normalization of the impact response of the federal funds rate to a policy shock means that R has at least rank 1. Faust et al. proceed by testing for successively higher ranks $l = 1, 2, 3, 4, 5, 6$. While ranks of 1 and 2 are clearly rejected, the null hypothesis of a rank 3 is not rejected, so the model is considered set identified.

Impulse Response Confidence Intervals under Set Identification. The construction of confidence intervals for the structural impulse responses involves two steps. First, we need to construct a confidence set D for the vector δ . Second, we need to construct the implied confidence interval for each structural impulse response coefficient for any fixed $\delta \in D$. Third, we need to construct a Bonferroni bound for these intervals across all $\delta \in D$.

Let \mathcal{D} denote the parameter space for δ . Assume that the restrictions $R\delta = r^{FFF}$ hold, that R is consistently estimated by \widehat{R} , that R is rank deficient, that $\sqrt{T}(\text{vec}(\widehat{R}) - \text{vec}(R)) \xrightarrow{d} \mathcal{N}(0, \Sigma_R)$ and that $\sqrt{T}(\widehat{r}^{FFF} - r^{FFF}) \xrightarrow{d} \mathcal{N}(0, \Sigma_{r^{FFF}})$. Consider the objective function

$$S(\delta) = T(\widehat{R}\delta - r^{FFF})'[(\delta' \otimes I_K)\widehat{\Sigma}_R(\delta \otimes I_K) + \widehat{\Sigma}_{r^{FFF}}]^{-1}(\widehat{R}\delta - r^{FFF}).$$

The estimator $\widehat{\delta}$ that minimizes this objective function is not consistent for the true δ because of the rank deficiency of R . However, $S(\delta_0)$ has a χ^2 null distribution regardless of the rank of R , where δ_0 denotes the true value of the vector δ . Hence,

$$D = \{\delta \in \mathcal{D} : S(\delta) \leq \chi^2_{0.95}(K)\}$$

is a confidence set for δ with asymptotic coverage probability of 0.95, regardless of the rank of R , where $\chi^2_{0.95}$ denotes the 95th percentile of a χ^2 distribution with degrees of freedom equal to K , the number of elements in r^{FFF} .

In practice, Faust et al. recommend constructing the confidence set D by drawing 10 million candidate solutions for δ at random uniformly from the parameter space \mathcal{D} and retaining the draws that satisfy the definition of D . In Faust, Swanson, and Wright (2004), δ is a 6×1 parameter vector. The parameter space \mathcal{D} they draw from is restricted as follows. After normalizing the coefficient on the federal funds rate to 0.25 (so a unit monetary policy shock raises the federal funds rate by 25 basis points), Faust et al. propose restricting the impact response of industrial production and of the CPI to this policy shock to lie between 0 and -0.1 and the other impact coefficients to lie between 0 and -0.25 . The magnitude restrictions are somewhat ad hoc, but they are clearly different from the exclusion restrictions traditionally used in recursive structural VAR models of monetary policy.

Having simulated the $(1 - \gamma_1)100\%$ confidence set D , for any fixed $\delta \in D$, one can use conventional bootstrap methods to construct a $(1 - \gamma_2)100\%$ confidence interval for each structural impulse response coefficient conditional on δ . Denote this confidence interval by $[\underline{\theta}(\delta), \bar{\theta}(\delta)]$. Finally, we form the outer envelope of all these intervals across $\delta \in D$ as $[\inf_{\delta \in D} \underline{\theta}(\delta), \sup_{\delta \in D} \bar{\theta}(\delta)]$. The latter pointwise interval has at least $(1 - \gamma_1 - \gamma_2)100\%$ asymptotic coverage by the Bonferroni inequality.

Faust et al. report both confidence intervals conditional on the original reduced-form VAR estimate and allowing for sampling uncertainty in the reduced-from parameters. Given the absence of point estimates for the impulse responses, these confidence bands may be difficult to interpret in practice, but the empirical results in Faust et al. are nevertheless informative and allow us to reject hypotheses of economic interest. Most importantly, the usual recursive identification of monetary policy shocks is clearly rejected, as is any identification that insists on a monetary policy shock having no effect on prices contemporaneously. This confirms the earlier concerns with recursive structural monetary policy VAR models. Faust et al.'s identification also eliminates the price puzzle – the finding in the benchmark recursive identification that the impulse response of prices to an unexpected monetary policy tightening first rises significantly, before falling. Faust, Swanson, and Wright (2004) nevertheless find that only a small fraction of the variance of output can be attributed to monetary policy shocks.

15.1.2 A Point-Identified Approach

D'Amico and Farka's (2011) analysis of the interaction of the stock market and the interest rate presents an alternative approach to using high-frequency financial data for the identification of structural VAR models. They propose a methodology for estimating simultaneously the response of stock returns to policy decisions and the Federal Reserve's contemporaneous reaction to the stock market. Their methodology has broad applicability when modeling asset prices.

Overview. D'Amico and Farka's approach involves two steps. In the first step, the response of the stock market to policy shocks is estimated outside the VAR model by measuring changes in intraday S&P500 futures prices immediately before and after policy announcements. The monthly policy shock is obtained by summing the intraday shocks over the course of a given month. In the second step, D'Amico and Farka impose that external estimate when estimating the response of the federal funds rate to stock returns in a monthly VAR model.

D'Amico and Farka consider a 7-variable monthly VAR model of U.S. monetary policy. They propose to partition the structural impact multiplier matrix

into two blocks. One block includes the macroeconomic variables (industrial production, the consumer price index, a smoothed index of commodity prices, nonfarm payrolls, and the ISM manufacturing index) and the other block includes the financial variables (federal funds rate, S&P500 stock returns). Their methodology is based on five key identifying assumptions.

1. The structural impact multiplier matrix is recursive except for the VAR variables within the financial block.
2. The relationship between reduced-form and structural errors is the same at the intraday and at the monthly frequency.
3. The monetary policy shock is the only shock at the time of the FOMC announcement.
4. Intraday changes in spot month federal funds futures prices around FOMC announcements reflect the effects of policy shocks.
5. Intraday changes in S&P500 futures prices around FOMC announcements provide a good measure of unexpected changes in stock prices caused by monetary policy shocks.

The assumption of a recursive structure outside of the financial block is innocuous as long as only the simultaneity of stock prices and policy rates is of interest. As in Faust et al., the analysis treats risk premia as approximately constant around FOMC announcements.

Estimation and Inference. Suppose that the economy is described by the K -dimensional structural VAR model

$$B_0 y_t = \tilde{B}(L) y_t + w_t,$$

where $\tilde{B}(L) = B_1 L + \dots + B_p L^p$ and $u_t = B_0^{-1} w_t$ is the vector of reduced-form errors. Suppose that there are two financial variables (y_t^f) and $(K - 2)$ macroeconomic variables (y_t^m). The matrix B_0 is partitioned as

$$B_0 = \begin{bmatrix} B_{0,M}^m & 0_{(K-2) \times 2} \\ B_{0,M}^f & B_{0,F}^f \end{bmatrix},$$

where $B_{0,M}^m$ is a $(K - 2) \times (K - 2)$ lower-triangular matrix, while the $2 \times (K - 2)$ matrix $B_{0,M}^f$ and the 2×2 matrix $B_{0,F}^f$ govern the contemporaneous interaction of the macroeconomic variables and financial variables. Note that the model differs from a conventional semistructural model of monetary policy in that it allows for simultaneous interaction between the financial variables. Because the diagonal elements of B_0 can be normalized to 1,

$$B_{0,F}^f = \begin{bmatrix} 1 & -a \\ -b & 1 \end{bmatrix}$$

and we can partition B_0^{-1} such that

$$B_0^{-1} = \begin{bmatrix} (B_0^{-1})_M^m & 0 \\ (B_0^{-1})_M^f & (B_0^{-1})_F^f \end{bmatrix},$$

where

$$(B_0^{-1})_F^f = (B_{0,F}^f)^{-1} = \frac{1}{1-ab} \begin{bmatrix} 1 & a \\ b & 1 \end{bmatrix},$$

so $a/(1-ab)$ denotes the contemporaneous response of stock prices to a monetary policy shock, if the federal funds rate is ordered last in the VAR model. In conventional recursive structural VAR models of monetary policy, identification is achieved by imposing either $a = 0$ or $b = 0$. Neither assumption is compelling as discussed earlier. D'Amico and Farka dispense with that assumption. This means that they require one more identifying restriction for exact identification. D'Amico and Farka achieve identification by imposing an extraneous estimate of a obtained by regressing monthly stock returns on measures of monetary policy shocks constructed from intraday changes in the federal funds rate near FOMC announcements.

Denote this intraday change by $w_{t,d}$. Then the monthly policy shock is constructed by cumulating the intradaily policy shocks over the course of a given month t ,

$$w_t^{FF} = \sum_{d=1}^D w_{t,d},$$

where d indexes the subintervals within the month. In practice, D'Amico and Farka (2011) rely on 20-minute intervals. Likewise, it is assumed that the reduced-form errors may be decomposed as

$$u_t^{FF} = \sum_{d=1}^D u_{t,d},$$

and it is assumed that $u_{t,d} = B_0^{-1} w_{t,d}$, where B_0^{-1} is the same matrix as in the monthly data. This is, of course, a very strong assumption.

The premise of the analysis is that, in a short interval around an FOMC announcement, the only new information is the monetary policy news. It is therefore reasonable to assume that all elements but the last element of $w_{t,d}$ are zero within that interval. Thus, if u_t^{FF} and u_t^{ret} denote the reduced-form errors for the equations of the federal funds rate and for S&P500 stock returns,

in a narrow window around the policy announcement, we have

$$u_{t,d}^{FF} = \frac{1}{1-ab} w_{t,d}^{policy},$$

$$u_{t,d}^{ret} = \frac{a}{1-ab} w_{t,d}^{policy} = au_{t,d}^{FF}.$$

The estimation proceeds in two steps. First, following Kuttner (2001), it is assumed that changes in the federal funds futures price around the time of FOMC announcements can be used to measure the difference between the announced funds rate and the ex-ante expectation, $u_{t,d}^{FF}$. Likewise, changes in the S&P500 futures prices within the same window of time are assumed to be a good measure of unexpected changes in stock prices, $u_{t,d}^{ret}$. An OLS regression of changes in S&P500 futures and changes in the federal funds rate for all FOMC announcement dates yields an estimate \hat{a} of a .

Second, one substitutes this estimate for a in the monthly structural VAR model

$$\begin{bmatrix} 1 & -\hat{a} \\ -b & 1 \end{bmatrix} \begin{pmatrix} ret_t \\ FF_t \end{pmatrix} = -B_{0,M}^f y_t^m + \tilde{B}^f(L) y_t + \begin{pmatrix} w_t^{ret} \\ w_t^{policy} \end{pmatrix},$$

where $\tilde{B}^f(L)$ denotes the last two rows of $\tilde{B}(L)$. Let \hat{w}_t^{ret} denote the residual from the first equation. Then the second equation can be estimated by regressing the federal funds rate on contemporaneous and lagged values of all variables, using the residuals \hat{w}_t^{ret} as an instrument for ret_t . Because the model proposed by D'Amico and Farka (2011) is point-identified, inference is standard and may be implemented by conventional bootstrap methods.

D'Amico and Farka (2011) conclude that a surprise 25 basis points tightening in policy rates is associated with a decline in stock prices by 1.25%. The response of monetary policy to stock prices is positive and statistically significant. Policy rates increase by about 2 basis points in response to a 1% unanticipated increase in stock prices. The use of intraday data allows D'Amico and Farka to address both the endogeneity issue (because there is no simultaneous reaction within the short window around a policy announcement) and the omitted variable problem (by reducing the likelihood that new information is released during the short window). One drawback of their analysis is that accurate information about the timing of U.S. monetary policy announcements is available only after 1994.

This approach may be extended to other forms of interventions and other financial variables including exchange rates, commodity prices, and other financial instruments to the extent that the precise timing of these interventions is known and to the extent that corresponding intraday futures price data are available. Of course, the premise that the risk premium remains constant during policy interventions must be evaluated on a case-by-case basis.

15.1.3 Discussion

Identifying structural VAR models based on high-frequency futures data offers several advantages over conventional exclusion restrictions, but it also has several limitations. One limitation is that this approach is designed for modeling monetary policy decisions and does not easily generalize to other economic shocks. Another concern is that this approach requires high-frequency futures prices for extended periods, which limits its applicability. A third concern is that this approach is only as good as the information we have about the timing of the underlying policy decisions, which limits its applicability to other countries.

There is yet another and more fundamental concern, however, and that is how to measure policy shocks at monthly frequency that are actually observed at daily or intradaily frequency. Indeed, the studies we reviewed made different choices in this regard. As discussed in Chapter 2, temporal aggregation of variables has a strong impact on the dynamic structure of the DGP. If the actual reaction of markets occurs within a very short period following the policy shock, then ideally one would have to consider the daily or intradaily process generating the set of variables under consideration. The effects on the corresponding monthly variables would have to be assessed by analyzing the properly aggregated process. No in-depth investigation of the aggregation implications on the stochastic properties driving the whole system of variables has been conducted thus far.

An alternative approach to allowing higher frequency data in a monthly model would be to consider explicitly models for mixed frequency data. In other words, the DGP of a system of variables being observed at different frequency is modeled explicitly. This approach formally allows one to take into account the changes in the dynamic structure resulting from time aggregation. The approach is used by Foroni and Marcellino (2016). They find that taking into account the mixed frequency nature of the data explicitly can make a substantial difference to the impulse responses. Thus, simply aggregating the shocks and making the assumptions on the aggregated effects that are used by the research reviewed in this chapter so far is problematic and can lead to conclusions that do not reflect the true responses of the variables in the underlying economic system.

15.2 Identification Based on External Instruments

In some situations, researchers have access to direct measures of exogenous shocks obtained from information not contained in the VAR model. Examples of such exogenous variables are monetary policy and fiscal policy shocks obtained from narrative evidence as in Romer and Romer (2004) or Ramey and Shapiro (1998), the exogenous OPEC oil supply shock measures of Hamilton

(2003) and Kilian (2008b), or the oil supply news shock measure proposed in Arezki, Ramey, and Sheng (2017) (see Chapter 7).

Consider measures of exogenous OPEC oil supply shocks, for example. One way of incorporating this exogenous information is to augment the baseline structural VAR model of Kilian (2009) for the growth in global oil production ($\Delta prod_t$), global real economic activity (rea_t), and the real price of oil ($rpoil_t$) to include the exogenous variable in question, x_t . If this variable is serially uncorrelated, we may treat it both as an observable and as a structural shock. For example, Kilian (2006) incorporates the exogenous OPEC oil supply shock series developed in Kilian (2008b) into a restricted monthly structural VAR(24) model

$$B_0 y_t = B_1 y_{t-1} + \cdots + B_{24} y_{t-24} + w_t,$$

where $y_t = (x_t, \Delta prod_t, rea_t, rpoil_t)'$ and the first row of B_i , $i = 1, \dots, 24$, has been restricted to zero, reflecting the exogeneity of x_t and its lack of serial correlation. The vector w_t denotes the serially and mutually uncorrelated structural shocks. By construction, in this VARX model of the global oil market $y_{1t} = x_t = u_t^x = b_0^{11} w_t^{\text{political OPEC oil supply}}$. The model is estimated by restricted GLS as discussed in Chapter 2. As in Kilian (2009), the identification relies on a recursive structure, but now with the exogenous OPEC oil supply shock ordered first, allowing global oil production to respond endogenously to politically motivated OPEC supply shocks:

$$\begin{pmatrix} u_t^x \\ u_t^{\Delta prod} \\ u_t^{rea} \\ u_t^{rpoil} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} & 0 \\ b_0^{41} & b_0^{42} & b_0^{43} & b_0^{44} \end{bmatrix} \begin{pmatrix} w_t^{\text{political OPEC oil supply}} \\ w_t^{\text{other oil supply}} \\ w_t^{\text{aggregate demand}} \\ w_t^{\text{oil-specific demand}} \end{pmatrix}.$$

Because the distinction between politically motivated OPEC oil supply shocks and other oil supply shocks turns out to be unimportant in practice, Kilian (2009) therefore switched to a simpler version of this model that excluded the OPEC oil supply shock and focused on shocks to aggregate oil production, allowing the model to be estimated by unrestricted least squares.

Similar approaches have also been used in fiscal policy VAR models. For example, Auerbach and Gorodnichenko (2012) augment a (nonlinear) quarterly structural VAR model of fiscal spending to include a measure of news about fiscal spending, defined as the forecast errors implied by professional survey forecasts or other expert forecasts and denoted by $news_t$. The model variables are $y_t = (news_t, gov_t, tax_t, gdp_t)'$, where gov_t is the log of government purchases, tax_t is the log of government receipts, and gdp_t is the log of real GDP. The model imposes the identifying assumption that government spending does not respond to changes in fiscal and macroeconomic conditions

in the short run. The news variable is ordered first, allowing an autonomous shock to fiscal spending to be identified as the second structural shock, with the third and fourth structural shocks remaining unidentified from an economic point of view such that

$$\begin{pmatrix} u_t^{news} \\ u_t^{gov} \\ u_t^{tax} \\ u_t^{gdp} \end{pmatrix} = \begin{bmatrix} b_0^{11} & 0 & 0 & 0 \\ b_0^{21} & b_0^{22} & 0 & 0 \\ b_0^{31} & b_0^{32} & b_0^{33} & 0 \\ b_0^{41} & b_0^{42} & b_0^{43} & b_0^{44} \end{bmatrix} \begin{pmatrix} w_t^{\text{fiscal news}} \\ w_t^{\text{fiscal spending}} \\ w_{3t} \\ w_{4t} \end{pmatrix}.$$

Unlike in Kilian (2006), the model is estimated without imposing the assumption that the news shock variable is serially uncorrelated and strictly exogenous. Another difference from the analysis in Kilian (2006) is that Auerbach and Gorodnichenko (2012) use their exogenous variable to deal with the informational deficiencies of their baseline structural VAR model. A closely related approach was also suggested by Favero and Giavazzi (2012).

More recently, a number of authors have suggested to use such exogenous variables instead as instruments for the structural shock of interest (see, e.g., Stock and Watson 2012; Mertens and Ravn 2012, 2014; Montiel Olea, Stock, and Watson 2015b; Piffer and Podstawska 2016). For example, direct measures of monetary policy shocks, as discussed in Chapter 6, may be used as an external instrument (see Stock and Watson 2012). This approach is also sometimes referred to as using a proxy VAR model. One key advantage of this model is that it allows the central bank to respond to asset prices without restricting the feedback from policy shocks to asset prices. As in the approach discussed in Section 15.1, however, the high-frequency proxy shocks must be scaled to monthly frequency.

Likewise, direct measures for exogenous fiscal policy shocks may be used as instruments in modeling fiscal aggregates. For example, Mertens and Ravn (2014) propose to interpret exogenous tax changes identified by the narrative approach to identification as proxy measures of latent structural tax shocks. The key identifying assumption is that the narrative measure is correlated with the tax shocks in the structural VAR model, but orthogonal to other structural shocks in this model. Similarly, Montiel Olea, Stock, and Watson (2015b) propose using external estimates of exogenous OPEC oil supply shocks as instruments for the reduced-form error of world oil production in a VAR model of the impact of oil price shocks on the U.S. economy.

15.2.1 Estimation and Inference

A comprehensive discussion of this approach is provided in Montiel Olea, Stock, and Watson (2015b). Recall the covariance stationary reduced-form

K -dimensional $\text{VAR}(p)$ model

$$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where $A_i = B_0^{-1} B_i$, $i = 1, \dots, p$, and $u_t = B_0^{-1} w_t$.

Instrumental Variable Estimation. We are interested in estimating the responses of the K model variables to structural shock j, h periods ahead,

$$\theta_{j,h} = \Phi_h B_0^{-1} e_j,$$

where e_j denotes the j^{th} column of I_K , $B_0^{-1} e_j$ is the j^{th} column of B_0^{-1} , and Φ_h refers to the $K \times K$ matrix of h -step ahead reduced-form impulse responses, as defined in Chapter 4. Since $B_0^{-1} e_j = \theta_{j,0}$, $\theta_{j,h} = \Phi_h \theta_{j,0}$. Without loss of generality, let the structural shock of interest be w_{1t} . Thus, we focus on the estimation of $\theta_{1,h}$. Suppose that there is only one external instrument, z_t .³

The key assumption for estimating $\theta_{1,0}$ is that the external instrument, z_t , is correlated with w_{1t} and uncorrelated with all other structural shocks, i.e., $\mathbb{E}(z_t w_{kt}) = a \neq 0$ and $\mathbb{E}(z_t w_{kt}) = 0$ for $k = 2, \dots, K$. Except for the fact that w_{1t} is unobserved, these are the standard conditions required for conventional IV analysis. Using $u_t = B_0^{-1} w_t$, it follows that

$$\mathbb{E}(z_t u_t) = \mathbb{E}(B_0^{-1} z_t w_t) = a\theta_{1,0}.$$

Montiel Olea et al. impose the normalization that the $(1, 1)$ element of the structural impact multiplier matrix B_0^{-1} is unity. Under these assumptions the scalar structural parameter a and the $K \times 1$ vector of structural impact responses $\theta_{1,0} = B_0^{-1} e_1$ are identified by the moment condition

$$\mathbb{E}(z_t u_t) = a\theta_{1,0}$$

or, equivalently, by

$$\mathbb{E}[z_t(y_t - A_1 y_{t-1} - \cdots - A_p y_{t-p}) - a\theta_{1,0}] = 0.$$

In addition, the VAR slope parameters $\alpha = \text{vec}(A_1, \dots, A_p)$ satisfy the moment conditions

$$\mathbb{E}\left[\left(y'_{t-1}, \dots, y'_{t-p}\right)' \otimes (y_t - A_1 y_{t-1} - \cdots - A_p y_{t-p})\right] = 0.$$

³ Montiel Olea, Stock, and Watson (2015b) show that their approach may be generalized to a situation in which up to K structural shocks are identified by as many external instruments.

The GMM estimator corresponding to these two sets of moment conditions is given by

$$\begin{aligned}\widehat{\alpha} &= \left[\left(\frac{1}{T} \sum_{t=1}^T Y_{t-1} Y'_{t-1} \right)^{-1} \otimes I_K \right] \frac{1}{T} \sum_{t=1}^T (Y_{t-1} \otimes y_t), \\ \widehat{\alpha} &= \frac{1}{T} \sum_{t=1}^T z_t \widehat{u}_{1t},\end{aligned}\tag{15.2.1}$$

$$\widehat{\theta}_{1,0} = \frac{1}{\widehat{\alpha}} \frac{1}{T} \sum_{t=1}^T z_t \widehat{u}_t,\tag{15.2.2}$$

where $Y_{t-1} \equiv (y'_{t-1}, \dots, y'_{t-p})'$ and $\widehat{u}_t = y_t - (Y'_{t-1} \otimes I_K) \widehat{\alpha}$, such that the K -dimensional vector of structural responses to shock 1 at horizon h is

$$\widehat{\theta}_{1,h} = \widehat{\Phi}_h \widehat{\theta}_{1,0},$$

where $\widehat{\Phi}_h$ is obtained from $\widehat{\alpha}$, as discussed in Chapter 4. The estimator is \sqrt{T} -consistent and asymptotically normal. Analogous results may be constructed for the structural forecast error variance decomposition with respect to the first shock. Pointwise inference on the structural impulse responses and forecast error variance decompositions may be based on the plug-in estimators of the asymptotic variance derived in Montiel Olea, Stock, and Watson (2015b).

Weak Instrument Diagnostics. An important concern in applied work is that many external instruments are weak in the econometric sense. Conventional GMM procedures are not robust to the use of weak external instruments. The concern is that the correlation between the instrument, z_t , and w_{1t} is low, calling into question the identifying assumption that $\mathbb{E}(z_t w_{1t}) = a \neq 0$. Under the normalizing assumption that the $(1, 1)$ element of the structural impact matrix B_0^{-1} is unity, a low correlation between z_t and the reduced-form residual \widehat{u}_{1t} resulting in a value of $\widehat{\alpha}$ close to zero in expression (15.2.1), renders the estimator $\widehat{\theta}_{1,0}$ in expression (15.2.2) unreliable.

Montiel Olea et al. show that the asymptotic distribution of the GMM Wald test for the scalar structural impulse response $\widehat{\theta}_{1,h}$ under weak-instrument asymptotics is equivalent to the asymptotic distribution of a Wald test statistic for the coefficient of an endogenous regressor in a conventional IV regression model given by

$$e'_i \widehat{\Phi}_h \widehat{u}_t = \kappa_0 \widehat{u}_{1t} + v_{2t},\tag{15.2.3}$$

$$\widehat{u}_{1t} = az_t + v_{1t},\tag{15.2.4}$$

where κ_0 is the value of $\theta_{1,h}$ under the null hypothesis, $e'_i \widehat{\Phi}_h \widehat{u}_t$ is the dependent variable in the second-stage regression (15.2.3), \widehat{u}_{1t} is the endogenous

regressor in the second-stage regression, and z_t is the external instrument used in the first-stage regression (15.2.4). This alternative representation illustrates that one can assess the strength of the instrument z_t based on the correlation between \hat{u}_{1t} and z_t . In practice, an F -statistic for $H_0 : a = 0$ below 10 in the first-stage regression is usually interpreted as an indication of a weak instrument problem.

Weak instrument problems are common in IV regression analysis. For example, Kilian (2008a, 2008b) provides evidence that measures of exogenous OPEC oil supply shocks of the type proposed by Hamilton (2003) and Kilian (2008b) do not have much predictive power for changes in the real price of oil and, hence, must be considered a weak instrument for the real price of oil. In the context of a VAR model identified by external instruments, Stock and Watson (2012) show that a similar weak instrument problem arises when regressing the VAR prediction error for world oil production, \hat{u}_{1t} , on exogenous OPEC oil supply shocks, z_t .

Robust Inference. The key result in Montiel Olea et al. is that inference about structural impulse responses may be conducted in a way that is robust to the strength of the instrument and remains valid even when using a weak instrument. This involves an alternative asymptotic thought experiment than that underlying the derivation of the conventional GMM estimator discussed earlier. In the context of the IV model (15.2.3) and (15.2.4), this thought experiment amounts to postulating that the population parameter a in the first-stage regression of the IV model is not a constant, but is modeled as local to zero in the sense that $a = 0 + \lambda/\sqrt{T}$, as $T \rightarrow \infty$, where λ is a constant. As in the discussion of local-to-unity asymptotics in Chapters 3 and 12, these local asymptotics are merely a device for approximating the finite-sample distribution of the estimator.

Montiel Olea et al. propose an adjusted version of the Anderson-Rubin test statistic for the IV model. This test statistic rejects $H_0 : \theta_{i1,h} = \kappa_0$ if

$$AR(\kappa_0) = \left(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t (e'_i \hat{\Phi}_h \hat{u}_t - \kappa_0 \hat{u}_{1t}) \right)^2 / \hat{\sigma}_T^2(\kappa_0) > \chi^2_{1-\gamma}(1).$$

The statistic $\hat{\sigma}_T^2(\kappa_0)$ is a consistent estimator of the asymptotic variance of

$$\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t (e'_i \hat{\Phi}_h \hat{u}_t - \kappa_0 \hat{u}_{1t}).$$

Montiel Olea, Stock, and Watson (2015b) derive a formula for $\hat{\sigma}_T^2(\kappa_0)$ based on the delta method.

In the just-identified case, this test statistic has three desirable properties. First, the test is asymptotically valid whether the instruments are weak or

strong. Second, the test achieves the same local power as the GMM test under strong asymptotics. Third, the test is asymptotically efficient within the class of tests that are asymptotically valid in the presence of weak instruments.

Robust $(1 - \gamma)100\%$ confidence sets for the structural impulse responses may be obtained by inverting this test statistic such that

$$CS \equiv \{\kappa_0 \in \mathbb{R} \mid AR(\kappa_0) \leq \chi^2_{1-\gamma}(1)\}.$$

Implementation of this robust interval does not require the user to evaluate the $AR(\kappa_0)$ statistic over a grid of κ_0 . Montiel Olea, Stock, and Watson (2015b) show that the confidence set for $\theta_{i1,h}$ may also be computed from the roots that solve

$$\left(\frac{1}{\sqrt{T}} \sum_{t=1}^T z_t (e'_i \hat{\Phi}_h \hat{u}_t - \kappa_0 \hat{u}_{1t}) \right)^2 - \hat{\sigma}_T^2(\kappa_0) \chi^2_{1-\gamma}(1) = 0,$$

which can be shown to be a quadratic equation in κ_0 .

15.2.2 Discussion

The literature on the use of external instruments is evolving at a rapid pace. For example, researchers recently have been exploring the use of external shock measures based on futures prices, as discussed in Section 15.1 (see, e.g., Gertler and Karadi 2015; Rogers, Scotti, and Wright 2015). One context in which this approach may prove particularly useful is the identification of monetary policy shocks in factor-augmented VAR (FAVAR) models, because recursive orderings are not easy to defend in this context (see Chapter 16).

At the same time, the econometric methodology continues to be refined. For example, Lunsford (2015a) proposes a test of the strength of external instruments. In related work, Lunsford (2015b) develops a residual-based block bootstrap method that allows for conditional heteroskedasticity in structural VAR models identified by external instruments, building on Brüggemann, Jentsch, and Trenkler (2016).

16 Structural VAR Analysis in a Data-Rich Environment

Typical VAR models used for policy analysis include only small numbers of variables. One motivation for considering larger VAR models is that policy institutions such as central banks and government organizations consider large panels of time series variables in making policy decisions. If important variables are not included in a VAR model, that model becomes informationally deficient and estimates of the responses to policy shocks will be distorted by omitted-variable bias. Thus, unless a variable is known to be irrelevant, one should ideally include it in the structural VAR model.

Deciding on the relevance of a particular variable for an empirical model is a difficult task because the variables for which data are available may not correspond exactly to the variables used in theoretical models. For example, consider a monetary policy reaction function that includes inflation and the output gap as explanatory variables. It is well known that the output gap is difficult to measure. Hence, one can make the case for including all variables that contain information about the output gap because they could all be important for the analysis of the impact of monetary policy.

Another motivation for considering larger VAR models is that we may wish to examine the impact of monetary policy shocks at a more disaggregate level. For example, one may be interested not only in the response of the overall price level to a monetary policy shock, but also in the response of sub-indices corresponding to specific expenditure components. Such an analysis necessitates the inclusion of disaggregate price level data in the model. Likewise, one may be interested in the output response in specific sectors of the economy. In that case, again, these additional variables have to be included in the analysis.

Conventional unrestricted VAR models do not allow for the situations described above because the inclusion of many additional variables undermines the precision of the model estimates in small samples. Moreover, the extent to which a VAR model can be enlarged is limited by the fact that the

number of regressors cannot exceed the number of observations. This restriction can easily become binding when working with large-dimensional VAR models because the number of parameters in a VAR model increases with the square of the number of variables included.

Thus, the analyst often faces a dilemma in setting up the model. Whereas degrees-of-freedom constraints suggest including only a small number of variables, concerns over omitted-variables bias make it desirable that we include a large number of variables in the model. In response to this problem, several techniques have been developed that make it possible to increase the information content of VAR models. Two such techniques are factor-augmented VAR (FAVAR) models and large-dimensionl Bayesian VAR models. Related methods include panel VAR models, global VAR (GVAR) models, and spatial VAR models.

In Section 16.1, we show how factor models may be used to condense the information in a large panel of variables to a small number of factors. We distinguish between static and dynamic factor models. Section 16.2 illustrates how factors may be incorporated into structural VAR analysis. We consider two classes of models. The class of FAVAR models refers to VAR models that include one or more factors in addition to observables. Identification of the structural shocks within this augmented VAR model may be achieved by conventional methods. An alternative is the class of dynamic factor models. Whereas FAVAR models may include both factors and observed variables, in structural dynamic factor models all observables are expressed as a weighted average of factors. The evolution of these factors depends on the structural shocks. The identification of these shocks is achieved by restricting the implied responses of the observed variables to the structural shocks. These models play an increasingly important role in applied work both in forecasting and in structural analysis (see, e.g., Bernanke, Boivin, and Eliasz 2005; Favero, Marcellino, and Neglia 2005; Del Negro and Otrok 2007).

Imposing Bayesian restrictions on the parameters of a structural VAR model is an alternative approach to dealing with many variables in VAR analysis. Large-scale BVAR models have gained popularity lately, in particular as forecasting models (see, e.g., Carriero, Kapetanios, and Marcellino 2009, 2012; and Koop 2013a). They have also been used for structural analysis, however (see, e.g., Bańbura, Giannone, and Reichlin 2010). In Section 16.3, problems specific to the analysis of large-dimensional structural BVARs are discussed.

Finally, some alternative approaches to fitting large-dimensional VAR models are considered in Section 16.4. In particular, we briefly discuss panel VAR models, GVAR models, and spatial VAR models. Section 16.5 contains a summary of the pros and cons of alternative modeling approaches for large panels of time series.

16.1 Factor Models

This section shows how common factors may be extracted from a large cross-section of time series data. FAVAR models rely on a small number of these factors to summarize the information contained in the original large-dimensional data set. The construction of structural FAVAR models and related models is discussed in Section 16.2.

In Section 16.1.1, we review static factor models. Static factor models were originally designed for cross-sectional analysis. In Section 16.1.2, we present the more general dynamic factor model (DFM) specifically designed for time series analysis. We present alternative versions of the DFM. The problem of determining the appropriate number of factors is reviewed in Section 16.1.3.

There are a number of good surveys on factor analysis for time series data (see, e.g., Stock and Watson 2005, 2016; Breitung and Eickmeier 2006; Bai and Ng 2008; Barhoumi, Darné, and Ferrara 2014). DFM s in particular have been used extensively for forecasting (e.g., Stock and Watson 2002a, 2006, 2011). Some of that literature is also relevant in the present context. Important results on statistical inference for DFM s are available in Forni, Hallin, Lippi, and Reichlin (2000, 2004), Breitung and Tenhofen (2011), Choi (2012), Stock and Watson (2002a, 2005), Bai (2003), and Gonçalves and Perron (2014), among others. A detailed survey of how to conduct inference in DFM s is provided in Bai and Ng (2008).

16.1.1 Static Factor Models

Model Setup. The classical static factor model assumes the form

$$x_t = \Lambda f_t + v_t, \quad (16.1.1)$$

where $x_t \stackrel{iid}{\sim} (0, \Sigma_x)$ is a vector of N observed variables, f_t is an r -dimensional vector of unobserved common factors, and r is typically much smaller than N , $r << N$. Accordingly Λ is a $N \times r$ matrix of factor loadings. Finally, $v_t \stackrel{iid}{\sim} (0, \Sigma_v)$ is an N -dimensional vector of uncorrelated idiosyncratic components such that Σ_v is diagonal. Moreover, the common factors and idiosyncratic components are assumed to be orthogonal, i.e., $\mathbb{E}(f_t v_s') = 0$ for all s and t . Hence,

$$\Sigma_x = \Lambda \Sigma_f \Lambda' + \Sigma_v, \quad (16.1.2)$$

where $\Sigma_f = \mathbb{E}(f_t f_t')$ is the covariance matrix of the factors. If the factors are mutually uncorrelated such that Σ_f is diagonal, the factors are said to be orthogonal. Otherwise they are oblique. This basic model has been in use in statistical analysis for many decades. For a detailed review see Anderson

(2003), for example, who traces such models back to Spearman (1904). Notice that in the basic model (16.1.1) the observed variables are assumed to have mean zero. Thus, in practice, the data have to be transformed prior to the analysis.

Obviously, in model (16.1.1) the factors and factor loadings are not separately identified. For any nonsingular $r \times r$ matrix Q , defining $f_t^* = Qf_t$ and $\Lambda^* = \Lambda Q^{-1}$, implies $\Lambda f_t = \Lambda^* f_t^*$. In practice, it is common to choose the factor loading matrix such that it has orthonormal columns, implying that

$$\Lambda' \Lambda = I_r, \quad (16.1.3)$$

or to choose uncorrelated factors with variances normalized to 1,

$$f_t \sim (0, I_r). \quad (16.1.4)$$

In the latter case, the factors are orthogonal and

$$\Sigma_x = \Lambda \Lambda' + \Sigma_v.$$

Such normalizations are useful for developing estimation algorithms. They are not sufficient for uniquely identifying the model, however. For instance, if we normalize the factors as in (16.1.4), Λ is still not unique without further restrictions. This can be seen by choosing an orthogonal matrix Q and defining $\Lambda^* = \Lambda Q$. Thereby we arrive at the decomposition

$$\Sigma_x = \Lambda^* \Lambda^{*'} + \Sigma_v.$$

Exact identification of the static factor model can be ensured by choosing Λ such that expression (16.1.3) holds and choosing the factors such that Σ_{f_t} is a diagonal matrix with distinct diagonal elements in decreasing order. In other words, the first factor has the largest variance and, hence, explains the largest part of the variance of x_t among the common factors. The second factor, f_{2t} , has the second largest variance, etc. The requirement that the factor variances have to be distinct means that the columns of Λ cannot simply be reordered. Table 16.1.1 summarizes four alternative sets of identification conditions for factors and factor loadings (see Bai and Ng 2013). It should be noted that even when these conditions are satisfied, the Λ matrix is unique only up to the sign of its columns. For a thorough discussion of alternative identification conditions, see Anderson (2003, section 14.2.2). If the model parameters are identified, their estimation is straightforward.

Estimating Static Factor Models. If the factor loadings were known and normalized such that $\Lambda' \Lambda = I_r$, a natural estimator for the factors would be obtained by left-multiplying (16.1.1) with Λ' and dropping the idiosyncratic term,

$$\hat{f}_t = \Lambda' x_t. \quad (16.1.5)$$

Table 16.1. Identification Conditions for Factors and Factor Loadings

Restrictions for Λ	Restrictions for Σ_f
$\Lambda' \Lambda = I_r$	Σ_f diagonal with decreasing diagonal elements
$\Lambda' \Lambda$ diagonal with distinct, decreasing diagonal elements	$\Sigma_f = I_r$
$\Lambda = \begin{bmatrix} \lambda_{11} & 0 & \cdots & 0 \\ \lambda_{21} & \lambda_{22} & & 0 \\ \vdots & \vdots & \ddots & \vdots \\ \lambda_{r1} & \lambda_{r2} & \cdots & \lambda_{rr} \\ \vdots & \vdots & & \vdots \\ \lambda_{N1} & \lambda_{N2} & \cdots & \lambda_{Nr} \end{bmatrix}$ $\lambda_{ii} \neq 0, i = 1, \dots, r$	$\Sigma_f = I_r$
$\Lambda = \begin{bmatrix} I_r \\ \Lambda_2 \end{bmatrix}$	Σ_f unrestricted

In practice, the factor loadings are typically unknown. A possible objective function for estimation in that case is the sum of the squared idiosyncratic errors. Minimizing the variance of the idiosyncratic components amounts to maximizing the part of the variance of the observed variables explained by the common factors. In other words, we may estimate the factor loadings and factors by minimizing the sum of squared errors,

$$\begin{aligned}
& \min_{\Lambda, f_1, \dots, f_T} T^{-1} \sum_{t=1}^T (x_t - \Lambda f_t)' (x_t - \Lambda f_t) \\
&= \min_{\Lambda, f_1, \dots, f_T} \text{tr} \left(T^{-1} \sum_{t=1}^T (x_t - \Lambda f_t)(x_t - \Lambda f_t)' \right). \tag{16.1.6}
\end{aligned}$$

A solution to this minimization problem is obtained in three steps:

1. Find the r largest eigenvalues $\lambda_1 > \dots > \lambda_r$ of $S_x = T^{-1} \sum_{t=1}^T x_t x_t'$ and the corresponding orthonormal eigenvectors $\lambda_1, \dots, \lambda_r$.
2. Choose $\widehat{\Lambda} = [\lambda_1, \dots, \lambda_r]$.
3. Express the factor estimate as $\widehat{f}_t = \widehat{\Lambda}' x_t$.

The estimator $\widehat{\Lambda}$ is the so-called principal components (PC) estimator of Λ . Given the orthogonality of the eigenvectors, it satisfies $\widehat{\Lambda}' \widehat{\Lambda} = I_r$. The factors are the principal components and $\widehat{\Sigma}_f = T^{-1} \sum_{t=1}^T \widehat{f}_t \widehat{f}_t' = \widehat{\Lambda}' S_x \widehat{\Lambda} = \text{diag}(\lambda_1, \dots, \lambda_r)$. The eigenvalues $\lambda_1, \dots, \lambda_r$ are the empirical variances of the factors with λ_1 representing the variance of the principal component with the

largest contribution to the variance of the data, λ_2 represents the variance of the second-most important component, etc.

The asymptotic properties of estimators of static factor models for $T \rightarrow \infty$ and fixed N can be found in Anderson (2003, chapter 14). Results for more general factor models under the assumption that both the number of variables N and the sample size T go to infinity are derived in Stock and Watson (2002a) and Bai (2003), among others. In particular, these authors establish the consistency of $\widehat{\Lambda}$ and its asymptotic normality, provided N and T go to infinity at suitable rates and some further regularity conditions hold (see also Bai and Ng 2008).

The PC estimator is the ML estimator if the observations x_t come from a normal distribution and the idiosyncratic components have equal variances such that $\Sigma_v = \sigma^2 I_N$. In that case, the factors and idiosyncratic components are normally distributed according to $f_t \stackrel{iid}{\sim} \mathcal{N}(0, \Sigma_f)$ and $v_t \stackrel{iid}{\sim} \mathcal{N}(0, \sigma^2 I_N)$ (see Anderson 2003). If the variances of the idiosyncratic components are heterogeneous, i.e., $\Sigma_v = \text{diag}(\sigma_1^2, \dots, \sigma_N^2) \neq \sigma^2 I_K$, the log-likelihood is

$$\begin{aligned} \log l(\Lambda, f_1, \dots, f_T, \Sigma_v) &= \text{constant} - \frac{T}{2} \log(\det(\Sigma_v)) \\ &\quad - \frac{1}{2} \text{tr} \left(\sum_{t=1}^T (x_t - \Lambda f_t)(x_t - \Lambda f_t)' \Sigma_v^{-1} \right). \end{aligned}$$

Anderson (2003, section 14.4) points out that this likelihood function is unbounded in general and, hence, does not have a global maximum. Thus, standard ML estimation cannot be used. The likelihood function has local maxima, however, allowing us to consider local maxima in the neighbourhood of the true parameter vector (e.g., Breitung and Tenhofen 2011; Bai and Li 2012).

If an estimator $\widetilde{\Sigma}_v$ of Σ_v is available, the factor loadings and factors may be estimated by a feasible GLS method based on the minimization problem

$$\min_{\Lambda, f_1, \dots, f_T} T^{-1} \sum_{t=1}^T (x_t - \Lambda f_t)' \widetilde{\Sigma}_v^{-1} (x_t - \Lambda f_t).$$

Feasible GLS estimators for factor models and their asymptotic properties are discussed in Choi (2012).

If the normalization in expression (16.1.4) is used for the common factors and if the observations are normally distributed, ML estimation of the factor loadings and idiosyncratic variances involves maximizing the log-likelihood

$$\begin{aligned} \log l(\Lambda, \Sigma_v) &= \text{constant} - \frac{T}{2} \log(\det(\Sigma_x)) - \frac{1}{2} \text{tr}(TS_x \Sigma_x^{-1}) \\ &= \text{constant} - \frac{T}{2} \log(\det(\Lambda \Lambda' + \Sigma_v)) - \frac{1}{2} \text{tr}[TS_x (\Lambda \Lambda' + \Sigma_v)^{-1}] \end{aligned}$$

by numerical methods. Suitable algorithms are discussed, for instance, by Magnus and Neudecker (1988, chapter 17).

Approximate Static Factor Models. So far we have considered what might be called an exact static factor model in which the idiosyncratic components are clearly separated from each other and from the factors. For economic data such an assumption may be too strict. An alternative approach is to work with an approximate factor model that allows for the possibility that the common factors specified by the researcher do not include all of the common factors in the data. In other words, some common factors will be contained in what the static factor model labels as the idiosyncratic component. In the approximate factor model

$$\Sigma_x = \Lambda \Sigma_f \Lambda' + \Sigma_v,$$

where Σ_v is not necessarily a diagonal matrix. Assuming that the common factors are normalized to have variance one, Chamberlain and Rothschild (1983) define an approximate factor model by the condition that Σ_x has only r unbounded eigenvalues when $N \rightarrow \infty$, where N is the number of variables considered by the researcher. The common factors are defined by the requirement that there exists a sequence of $N \times r$ matrices Λ and positive definite covariances Σ_v such that

$$\Sigma_x = \Lambda \Lambda' + \Sigma_v$$

and the maximum eigenvalue of Σ_v is bounded when $N \rightarrow \infty$. Thus, the relative variance share of each idiosyncratic component is small, when the number of variables is large. Chamberlain and Rothschild (1983) apply this model to a financial market with many assets (see also Connor and Korajczyk 1986, 1993).

Obviously, in that case identification of the model becomes more difficult and conditions different from those stated earlier are required. In fact, it is even possible that Σ_v has a factor decomposition that needs to be clearly separated from the common factor part captured by $\Lambda \Lambda'$, at least asymptotically. Choi (2012) considers estimation of models of that type and provides general asymptotic results.

16.1.2 Dynamic Factor Models

Taking into account the serial dependence in the data is essential for forecasting and for structural analysis. As we have seen earlier, static factor models and approximate static factor models ignore the time series dependence of the data in extracting the common factors. Dynamic factor models are obtained by allowing f_t and v_t in (16.1.1) to be serially correlated or more generally dependent stochastic processes. Which model is obtained depends on the

assumptions made about the stochastic processes f_t and v_t . A number of special cases may be considered.

For example, if v_t is white noise, x_t inherits all its serial dependence from the common factors. An early example of such a model for time series data is considered by Peña and Box (1987) who postulate that the factors have a VARMA-DGP and that the Σ_v matrix is not necessarily diagonal. Inference may be conducted as in the static factor model (e.g., Choi 2012). This case is therefore not considered here. From a practical point of view such models are typically too restrictive.

If f_t and v_t have parametric VAR representations, model (16.1.1) may be viewed as a dynamic factor model (DFM). Some authors, including Boivin and Ng (2005), refer to this model as a static factor model to distinguish it from a model where lagged factors f_{t-j} appear on the right-hand side of (16.1.1) in addition to the contemporaneous factors. We do not use this terminology here because it can be shown that every dynamic factor models in the sense of Boivin and Ng (2005) can also be expressed in static form.

Finally, when the common component and the idiosyncratic components are general stochastic processes, we follow part of the literature in calling this model a generalized dynamic factor model (GDFM). We caution the reader, however, that it is important to check which assumptions are made and which terminology is used in each case, when reading the original literature.

The remainder of this section is structured as follows. DFM may be written equivalently in static form or in dynamic form. First, the static form of the DFM is presented and its estimation is discussed. Then the dynamic form of the DFM is considered. Finally the GDFM is presented.

Throughout this section, we presume that the variables have been transformed to be stationary. It should be noted that factor models can be adapted to allow for integrated variables, but the presence of stochastic trends means that the methods described in this chapter would have to be modified. For example, the estimation procedures discussed in this chapter are based on covariance matrix and spectral density estimators that are not meaningful when dealing with integrated variables. Extensions of factor models to allow for integrated variables can be found in Bai (2004), for example. In fact, there is a close relationship between factor models and the cointegrated VAR models discussed in Chapter 3 in that cointegrated variables share a common trend that may be viewed as a common factor. A detailed analysis of the DFM for integrated and possibly cointegrated processes can be found in Barigozzi, Lippi, and Luciani (2016).

Static Form of the DFM. Consider the model (16.1.1),

$$x_t = \Lambda^f f_t + v_t, \quad (16.1.7)$$

with dynamic factors being generated as

$$f_t = \Gamma_1 f_{t-1} + \cdots + \Gamma_s f_{t-s} + \eta_t$$

and idiosyncratic components being generated as

$$v_t = A_1 v_{t-1} + \cdots + A_p v_{t-p} + u_t,$$

where the A_i , $i = 1, \dots, p$, are diagonal matrices and u_t is white noise with diagonal covariance matrix Σ_u . Using lag operator notation,

$$\Gamma(L)f_t = \eta_t \quad \text{and} \quad A(L)v_t = u_t,$$

where $A(L) = \text{diag}[a_1(L), \dots, a_N(L)]$. This model is called the static form of the DFM because the relation between the observed x_t and the dynamic factors is contemporaneous. No lagged f_t appear in equation (16.1.7).

Estimation. Estimation of the factors and factor loadings in model (16.1.7) for a given number of factors, r , can be carried out by PC estimation, ignoring any serial dependence in the error terms. Bai (2003) derives the properties of this estimator. PC estimation is generally inefficient because the dependence structure of the errors is ignored. Choi (2012) develops a GLS estimation procedure that can accommodate heteroskedastic idiosyncratic components, and Breitung and Tenhofen (2011) propose a GLS estimation procedure that can deal with a more general dependence structure in the error terms. This procedure applies even if the model is just an approximate factor model. Bai and Li (2012) discuss ML estimation of such models.

Dynamic Form of the DFM. A more general formulation of a DFM is obtained if the factors are also allowed to enter in lagged form such that

$$x_t = \Lambda_0^f f_t + \Lambda_1^f f_{t-1} + \cdots + \Lambda_{q^*}^f f_{t-q^*} + v_t. \quad (16.1.8)$$

Assuming the same DGP for f_t and v_t as in the static form (16.1.7) of the DFM, this model can be written in lag operator notation as

$$x_t = \Lambda^f(L)f_t + v_t, \quad \Gamma(L)f_t = \eta_t, \quad A(L)v_t = u_t,$$

where

$$\Lambda^f(L) = \Lambda_0^f + \Lambda_1^f L + \cdots + \Lambda_{q^*}^f L^{q^*},$$

$$\Gamma(L) = I_r - \Gamma_1 L - \cdots - \Gamma_s L^s,$$

$$A(L) = \text{diag}[a_1(L), \dots, a_N(L)],$$

$f_t = (f_{1t}, \dots, f_{rt})'$ are the common factors, $v_t = (v_{1t}, \dots, v_{Nt})'$ is the vector of idiosyncratic components, and η_t is white noise such that $\mathbb{E}(u_t \eta_s') = 0 \forall t, s$.

Defining $F_t = (f'_t, \dots, f'_{t-q^*})'$ and $\Lambda^F = [\Lambda_0^f, \Lambda_1^f, \dots, \Lambda_{q^*}^f]$, model (16.1.8) can equivalently be written in static form as

$$x_t = \Lambda^F F_t + v_t,$$

the only difference being that the dimension of the factor vector is larger. The vector F_t is often referred to as the vector of static factors, whereas the corresponding shorter vector f_t is called the vector of primitive dynamic factors (or, simply, the vector of dynamic factors).

In order to estimate the DFM (16.1.8), it is useful to left-multiply model (16.1.8) by $A(L)$ which yields

$$A(L)x_t = \Lambda(L)f_t + u_t, \quad (16.1.9)$$

where $\Lambda(L) = A(L)\Lambda^f(L)$ is a matrix polynomial of order $q \leq pq^*$. Assuming without loss of generality that $q \geq s$, the model (16.1.9) can be written in static form as

$$A(L)x_t = \Lambda F_t + u_t, \quad F_t = \Gamma F_{t-1} + G\eta_t, \quad (16.1.10)$$

where, using similar notation as before, $F_t = (f'_t, \dots, f'_{t-q})'$, $\Lambda = [\Lambda_0, \Lambda_1, \dots, \Lambda_q]$, and

$$\Gamma = \begin{bmatrix} \Gamma_1 & \Gamma_2 & \cdots & \Gamma_q & \Gamma_{q+1} \\ I_r & 0 & \cdots & 0 & 0 \\ 0 & I_r & & 0 & 0 \\ \vdots & & \ddots & 0 & 0 \\ 0 & 0 & \cdots & I_r & 0 \end{bmatrix}_{R \times R} \quad \text{and} \quad G = \begin{bmatrix} I_r \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{R \times r}.$$

Here $R = r(q + 1)$ and $\Gamma_i = 0$ for $i > s$.

Following Chamberlain and Rothschild (1983), Stock and Watson (2005) refer to model (16.1.8) as an exact DFM if $A(L)$ has a diagonal structure and the error covariance matrix $\mathbb{E}(u_t u_t') = \Sigma_u$ is diagonal, which implies mutually uncorrelated idiosyncratic components. Models of this type were used in the earlier econometrics literature by Sargent and Sims (1977). They are also closely related to the index models considered by Reinsel (1983) and reduced-rank VAR models discussed by Velu, Reinsel, and Wichern (1986), Tso (1981), Ahn and Reinsel (1988), Reinsel (1993), Reinsel and Velu (1998) and Anderson (1999, 2002). Such models differ from the DFM in (16.1.8) by their assumptions about the error term v_t . They assume that v_t is white noise with a general, not necessarily diagonal covariance matrix. In other words, the error term cannot be interpreted easily as a vector of idiosyncratic components. In contrast to exact DFMs, approximate DFMs also allow for dependence between the idiosyncratic components. In the following we treat $A(L)$ and Σ_u as diagonal, unless noted otherwise.

Estimation of DFM. We defer the question of how to specify the model until later and take as given the number of lags and the number of factors in the model.

Before estimating the DFM, it is recommended that the variables are first standardized such that they have zero mean and variance one. Following Stock and Watson (2005), the dynamic form of the DFM can be estimated as follows:

Step 1. Construct an initial estimate $\tilde{A}(L)$ of $A(L) = \text{diag}[a_1(L), \dots, a_N(L)]$, for example, by regressing the individual variables on their own lags.

Step 2. Compute the PC estimator $\widehat{\Lambda}$ of Λ , as discussed earlier, from the model $\tilde{A}(L)x_t = \Lambda F_t + \tilde{u}_t$, where $\tilde{A}(L)x_t$ assumes the role of x_t in equation (16.1.1), and estimate the factors as $\widehat{F}_t = \widehat{\Lambda}'\tilde{A}(L)x_t$.

Step 3. Estimate $A(L)x_t = \Lambda \widehat{F}_t + \hat{u}_t$ by single-equation LS for each equation to update the estimate of $A(L)$.

Step 4. Iterate Steps 2 and 3 until convergence.

Using single-equation LS in Step 3 is justified because the idiosyncratic error terms are assumed to be instantaneously uncorrelated such that Σ_u is a diagonal matrix. If that assumption is incorrect, estimation efficiency can be improved by using a feasible GLS procedure because the regressors in the different equations of the system $A(L)x_t = \Lambda \widehat{F}_t + \hat{u}_t$ are not identical.

So far we have only discussed the estimation of the observation equation of the state-space model (16.1.10). Once the estimated factors \widehat{F}_t are available, the Γ coefficient matrix in the transition equation in model (16.1.10) can be estimated as well. It should be noted, however, that estimates of Γ are not required for many applications of dynamic factor models. The estimation of Γ is complicated by the fact that in practice we can only estimate some linear transformation of the static factors $F_t = (f'_t, \dots, f'_{t-q})'$. The four-step estimator of the DFM discussed above uses a statistical normalization that may not result in the primitive dynamic factors, f_t . Thus, estimating Γ by regressing \widehat{F}_t on \widehat{F}_{t-1} only produces an estimator $\widehat{\Gamma}^*$ of some linear transformation of Γ .

This concern may be addressed by implementing a second-stage estimator which determines the r linearly independent primitive factors f_t underlying $F_t = (f'_t, \dots, f'_{t-q})'$. Let \widehat{W} be the matrix of eigenvectors corresponding to the r largest eigenvalues of the residual covariance matrix $\widehat{\Sigma}_{\varepsilon} = T^{-1} \sum_t \widehat{\varepsilon}_t \widehat{\varepsilon}_t'$, where $\widehat{\varepsilon}_t = \widehat{F}_t - \widehat{\Gamma}^* \widehat{F}_{t-1}$ and $\widehat{\Gamma}^*$ is the estimator obtained by regressing \widehat{F}_t on \widehat{F}_{t-1} . Then $\widehat{\eta}_t = \widehat{W}' \widehat{\varepsilon}_t$ and the primitive factors f_t can be estimated as $\widehat{f}_t = \widehat{W}' \widehat{F}_t$. If estimates of $\Gamma_1, \dots, \Gamma_{q+1}$ are required, they may be obtained by regressing \widehat{f}_t on $\widehat{f}_{t-1}, \dots, \widehat{f}_{t-q-1}$. Finally, the covariance matrix of η_t can be estimated in the usual way using the residual covariance estimator based

on the latter regression. Alternatively, it may be based on $\widehat{\Sigma}_\eta = T^{-1} \sum_t \widehat{\eta}_t \widehat{\eta}_t'$, where $\widehat{\eta}_t = \widehat{W}' \widehat{\mathcal{E}}_t$. Methods for choosing the values of R and r required in this procedure are discussed in Section 16.1.3.

It is also possible to estimate all DFM parameters simultaneously by ML estimation under the assumption that x_t is Gaussian. The Gaussian likelihood may be evaluated with the Kalman filter because model (16.1.10) is in state-space form. Using state-space estimation algorithms is also appealing because it allows one to handle missing observations (see Stock and Watson 2016, section 2.3.4). Missing observations are a common problem in panels of variables, when the beginning of the available time series varies across N , and starting the panel sample at the beginning of the shortest series may result in a substantial loss of sample information for other variables. In addition, the state-space framework can accommodate time series with different sampling frequencies. Computing the full Gaussian ML estimator may be challenging if the panel of variables is large. Doz, Giannone, and Reichlin (2011) propose an alternative two-step estimator based on the Kalman filter that may be helpful in the latter context.

Asymptotic results for the estimators of dynamic factor models can be found in Stock and Watson (2002a), Bai (2003), and Bai and Ng (2008) among others. Despite the fact that the asymptotic properties are derived for T and $N \rightarrow \infty$, small-sample simulation evidence in Boivin and Ng (2006) indicates that including more variables in a factor analysis does not necessarily result in more accurate estimates. Likewise, Boivin and Ng show that including more variables need not improve the forecast accuracy of an approximate factor model.

Rather than using frequentist methods, one may also use Bayesian methods for estimating dynamic factor models. We return to Bayesian estimation in the context of large panels of variables in Section 16.3 and therefore do not discuss these methods here. Bayesian methods of estimating DFMs have been used by Otkrok and Whiteman (1998), Kose, Otkrok, and Whiteman (2003) and Amir Ahmadi and Uhlig (2009), for example.

Generalized Dynamic Factor Models. The generalized dynamic factor model (GDFM) generalizes (16.1.7) by allowing the common and idiosyncratic components to be general stationary processes that may not admit a finite-order VAR representation. We consider the model $x_t = \Lambda f_t + v_t$, where f_t is the process of common factors and v_t is the process of idiosyncratic components. The two processes f_t and v_t can be characterized by their spectral properties (see Forni and Lippi 2001; Forni, Hallin, Lippi, and Reichlin 2000, 2004, 2005). Provided x_t is $I(0)$, a natural generalization of the covariance decomposition (16.1.2) is a decomposition of the spectral density of x_t . Denoting the spectral density functions of x_t , f_t , and v_t by $\Sigma_x(\xi)$, $\Sigma_f(\xi)$, and $\Sigma_v(\xi)$,

respectively,

$$\Sigma_x(\xi) = \Lambda \Sigma_f(\xi) \Lambda' + \Sigma_v(\xi), \quad \forall \xi \in [0, 2\pi]. \quad (16.1.11)$$

The matrix $\Sigma_v(\xi)$ is assumed to be diagonal in the exact GDFM, whereas more general structures are allowed for in the approximate GDFM.

Estimation. Estimation of GDFMs based on the decomposition (16.1.11) is considered in Forni and Reichlin (1998) and Forni, Hallin, Lippi, and Reichlin (2000, 2004, 2005). Since they do not assume a parametric model for the DGP of the observables and the common factors, they use a nonparametric frequency-domain estimator developed by Brillinger (1975). Based on the work of Forni et al., Favero, Marcellino, and Neglia (2005) propose the following procedure for estimating the dynamic principal components:

Step 1. For a sample x_1, \dots, x_T of size T , the spectral density matrix of x_t is estimated as

$$\widehat{\Sigma}_x(\xi_j) = \sum_{s=-S}^S \omega_s \widehat{\Gamma}_x(s) e^{-is\xi_j}, \quad \xi_j = 2\pi j/(2S+1), \quad j = 0, 1, \dots, 2S,$$

where S is the window width, $\omega_s = 1 - |s|/(S+1)$ are the weights of the Bartlett window and $\widehat{\Gamma}_x(s) = T^{-1} \sum_t (x_t - \bar{x})(x_{t-s} - \bar{x})'$ is the sample covariance matrix of x_t for lag s . The window width has to be chosen such that $S \rightarrow \infty$ and $S/T \rightarrow 0$ as $T \rightarrow \infty$. Forni, Hallin, Lippi, and Reichlin (2000) observe that a choice of $S = 2T^{1/3}/3$ worked well in their simulations.

Step 2. For $j = 0, 1, \dots, 2S$, determine the eigenvectors $\lambda_1(\xi_j), \dots, \lambda_r(\xi_j)$ corresponding to the r largest eigenvalues of $\widehat{\Sigma}_x(\xi_j)$.

Step 3. Defining

$$\lambda_{sk} = \frac{1}{2S+1} \sum_{j=0}^{2S} \lambda_s(\xi_j) e^{ik\xi_j}, \quad k = -S, \dots, S,$$

the dynamic PCs of x_t are obtained as

$$\hat{f}_{st} = \sum_{k=-S}^S \lambda'_{sk} x_{t-k}, \quad s = 1, \dots, r,$$

and collected in the vector $\hat{f}_t = (\hat{f}_{1t}, \dots, \hat{f}_{rt})'$.

Step 4. Fit the regression

$$x_t = \Lambda_{-q} \hat{f}_{t+q} + \dots + \Lambda_p \hat{f}_{t-p} + v_t$$

and estimate the common component as the fitted value

$$\hat{x}_t = \widehat{\Lambda}_{-q} \hat{f}_{t+q} + \dots + \widehat{\Lambda}_p \hat{f}_{t-p},$$

where $\widehat{\Lambda}_j$, $j = -q, \dots, p$, are the LS estimators. The leads q and lags p used in the regression could be chosen by model selection criteria. In practice, small numbers of leads and lags are used.

Using leads of the estimated factors and, hence, of the observations in Step 4 to reconstruct the common component is not an option in forecasting. An alternative one-sided procedure has been proposed in Forni, Hallin, Lippi, and Reichlin (2005). For impulse response analysis it is not clear that a one-sided procedure is preferred because impulse responses are expectations that are estimated from the full sample. Hence, we presented the two-sided procedure.

Comparison of DFM s and GDFM s . As we have emphasized, the various forms of the DFM are just different representations of the same DGP. The particular form used in empirical analysis is largely a matter of convenience. Of course, factor models may also differ to some extent in their underlying assumptions. For example, GDFMs in principle allow for more general dynamics of the factors and idiosyncratic components than DFMs. In particular, whereas GDFMs allow the factors and idiosyncratic errors to be general stationary stochastic processes, DFMs focus on finite-order VAR and AR approximations to these processes. How important that distinction is in practice, is not clear because linear stationary processes can be approximated by finite-order autoregressive processes. Thus, for all practical purposes, the choice of the model depends on the preferences of the user. Technical considerations are of secondary importance.

16.1.3 Selecting the Number of Factors

The specification of a DFM requires selecting the number of common factors and the various lag orders. Since PC analysis does not require the specification of a lag order, it is possible to specify the number of static factors before the lag orders of the VAR operators are determined. In classical static factor models, criteria such as choosing as many factors as are necessary to explain a prespecified fraction of the overall variance are commonly used. More precisely, we include additional common factors until the sum of the variances of the common factors exceeds a prespecified fraction of the sum of the eigenvalues of the sample covariance matrix. Another criterion in that spirit is the so-called scree test proposed by Cattell (1966), which increases the number of factors until the variance explained by these factors starts to taper off.

Specification of DFM s . For DFMs more formal criteria have been developed that assume the existence of a true number of factors, R_0 , and allow for consistent estimation of this quantity when both the cross-section and time dimension become large ($N, T \rightarrow \infty$). The most popular criteria are proposed by Bai and

Ng (2002) and take the form

$$IC(R) = \log \mathbb{V}(R) + Rg(N, T), \quad (16.1.12)$$

where $\mathbb{V}(R) = (NT)^{-1} \sum_{t=1}^T (x_t - \widehat{\Lambda}^F \widehat{F}_t)'(x_t - \widehat{\Lambda}^F \widehat{F}_t)$ and $g(N, T)$ is a penalty term. Note the similarity with the information criteria for VAR order selection discussed in Chapter 2. Bai and Ng show that under suitable conditions the estimator $\hat{R} = \operatorname{argmin}_{R=1, \dots, R_{\max}} IC(R)$ is consistent for the true number of factors R_0 . Of course, a necessary condition is that $R_{\max} \geq R_0$. Moreover, the penalty term $g(N, T)$ has to go to zero at a suitable rate with growing T and N . According to Breitung and Eickmeier (2006), the most popular criterion from this class chooses $g(N, T) = \left(\frac{N+T}{NT}\right) \log(\min[N, T])$ such that

$$IC_{p2}(R) = \log \mathbb{V}(R) + R \left(\frac{N+T}{NT} \right) \log(\min[N, T]). \quad (16.1.13)$$

Using this criterion we can estimate the number of static factors, which determines the dimension of F_t .

For structural analysis, the number of primitive dynamic factors, which corresponds to the dimension r of f_t in the dynamic form (16.1.9), is of primary interest. Bai and Ng (2007) propose a procedure for determining r . They utilize the fact that the error term in the transition equation in the static form (16.1.10), $G\eta_t$, has covariance matrix $G\Sigma_\eta G'$ of rank r and devise a procedure for determining that rank. Starting from estimates \widehat{F}_t of the static factors, they propose to fit a VAR model to the \widehat{F}_t . In our current framework that VAR model is of order one because we have assumed that $q \geq s$ (see model (16.1.10)). Thus, fitting

$$\widehat{F}_t = \Gamma \widehat{F}_{t-1} + \mathcal{E}_t$$

generates residuals $\widehat{\mathcal{E}}_t$, $t = 1, \dots, T$ (see Section 16.1.2). Let $\rho_1 \geq \rho_2 \geq \dots \geq \rho_R$ be the eigenvalues obtained from a PC analysis of the estimated residual covariance matrix $T^{-1} \sum_{t=1}^T \widehat{\mathcal{E}}_t \widehat{\mathcal{E}}_t'$ and define

$$\hat{D}_1(r) = \left(\frac{\rho_{r+1}^2}{\sum_{i=1}^R \rho_i^2} \right)^{1/2} \quad (16.1.14)$$

and

$$\hat{D}_2(r) = \left(\frac{\sum_{i=r+1}^R \rho_i^2}{\sum_{i=1}^R \rho_i^2} \right)^{1/2}. \quad (16.1.15)$$

Based on these quantities Bai and Ng (2007) propose to estimate the number of primitive dynamic factors as

$$\hat{r} = \min r \in \left\{ r \left| \hat{D}_1(r) < \frac{1}{\min(N^{1/2-\delta}, T^{1/2-\delta})} \right. \right\} \quad (16.1.16)$$

or as

$$\hat{r} = \min r \in \left\{ r \mid \hat{D}_2(r) < \frac{1}{\min(N^{1/2-\delta}, T^{1/2-\delta})} \right\}, \quad (16.1.17)$$

where δ is a small number between 0 and 1/2. In a simulation study they choose $\delta = 1/4$ which appears to give reasonable results.

Because the lag length of $\Lambda(L)$ in the DFM (16.1.9) cannot be larger than R/r , the number of static factors, R , bounds the lag length of $\Lambda(L)$ for given r . More precisely, we can choose $q + 1 = R/r$, if the latter ratio is an integer. If R/r is not an integer, a plausible value for q is the smallest integer larger than $(R/r) - 1$. Thus, choosing the number of factors is related to selecting the lag length of at least one of the lag polynomials in the DFM. The other lag lengths can be chosen by standard model selection criteria. Jacobs and Otter (2008) propose a procedure for determining the number of dynamic factors and their lags simultaneously.

A number of studies address the problem of estimating the number of factors in DFM. Further important contributions include Amengual and Watson (2007) and Breitung and Pötsch (2013). For a thorough review see Bai and Ng (2008). A more recent contribution is Caner and Han (2014) who propose to choose the number of factors based on a bridge estimator.

Specification of GDFMs. There are also methods specifically designed to estimate the number of factors in GDFMs. For example, Onatski (2009) presents a testing procedure for the number of static factors that may be viewed as a formalization of the scree test mentioned earlier. Hallin and Liška (2007) develop information criteria for estimating the number of dynamic factors based on the eigenvalues $\lambda_1(\xi_j), \dots, \lambda_S(\xi_j)$ of the estimated spectral density matrix $\widehat{\Sigma}_x(\xi_j)$, $\xi_j = 2\pi j/(2S+1)$, $j = 0, 1, \dots, 2S$. These criteria are defined as

$$PCP(k) = \frac{1}{N} \sum_{j=k+1}^N \frac{1}{2S+1} \sum_{s=-S}^S \lambda_j(\xi_s) + k\varphi(N, T)$$

and

$$IC(k) = \log \left(\frac{1}{N} \sum_{j=k+1}^N \frac{1}{2S+1} \sum_{s=-S}^S \lambda_j(\xi_s) \right) + k\varphi(N, T).$$

The bandwidth S has to be chosen such that $S \rightarrow \infty$ and $S/T \rightarrow 0$ for $T \rightarrow \infty$, and the penalty term $\varphi(N, T)$ has to satisfy

$$\varphi(N, T) \rightarrow 0 \quad \text{and} \quad \min\{N, S^{-2}, (ST)^{1/2}\}\varphi(N, T) \rightarrow \infty.$$

The search is done over $k \in \{0, \dots, r_{\max}\}$ with r_{\max} greater than or at least equal to the true rank in order to obtain consistent estimates.

16.1.4 Structural Change

As in other models, there may be parameter instability in dynamic factor models. A number of authors have considered that possibility and have developed statistical tools for diagnosing structural change in factor models (e.g., Stock and Watson 2009; Breitung and Eickmeier 2011; Chen, Dolado, and Gonzalo 2014; Han and Inoue 2015). A survey of related issues and procedures is provided in Stock and Watson (2016, section 2.5).

16.2 Factor-Augmented Structural VAR Models and Related Techniques

Section 16.1 focused on the question of how to extract the common factors from a large panel of time series. The statistical models used to estimate these factors by themselves have no structural economic interpretation. In this section we show how factor structures may be incorporated into structural VAR models and how these models may be used to construct structural impulse responses. Section 16.2.1 examines structural FAVAR models. In Section 16.2.2, we discuss structural analysis with dynamic factor models. Section 16.2.3 reviews empirical examples.

16.2.1 Structural FAVAR Models

One way of obtaining a FAVAR model is to augment the set of observed variables, z_t , in a given VAR model by latent factors, F_t , extracted from a set of observed variables x_t that does not include z_t . This approach is often intended to address the informational deficiencies of conventional small-scale structural VAR models (see Chapter 17). It also allows the researcher to approximate the responses of many more variables to a given structural shock than is possible in conventional VAR models.

Another motivation for this approach is that often the observed economic variables are only proxies for the economic concepts used in economic theory or that these variables may be subject to measurement error. In these situations a more accurate measure of the variable of economic interest may be obtained as a factor extracted from alternative measures of this variable. For example, Aruoba, Diebold, Nalewaik, Schorfheide, and Song (2016) propose viewing ‘true GDP’ as a latent variable for which we have several indicators, the two most obvious being an expenditure-based measure of GDP and an income-based measure of GDP. They propose to extract this latent ‘true GDP’ measure using a DFM. The resulting factors may be used instead of any one of the observed imperfect measures of the variables of interest, possibly in combination with observed variables that are either accurately measured or that are known to matter to economic agents. For example, the European Central Bank

(ECB) is required by law to maintain price stability as measured by the harmonized index of consumer prices in the European Monetary System. Thus, this variable has to enter directly the ECB's policy reaction function when specifying a VAR model of European monetary policy.

An example of such a model is Bernanke, Boivin, and Eliasz (2005). This study proposes a structural FAVAR model for $y_t = (F'_t, z'_t)'$ as

$$B(L) \begin{bmatrix} F_t \\ z_t \end{bmatrix} = w_t, \quad (16.2.1)$$

where the vector of structural shocks w_t is $(R + M)$ -dimensional white noise, $B(L) = B_0 + B_1 L + \dots + B_p L^p$ is a $(R + M) \times (R + M)$ matrix operator, F_t is a vector of R unobserved common factors that are related to the $N \times 1$ vector of informational variables x_t by the observation equation

$$x_t = \Lambda^F F_t + \Lambda^z z_t + e_t, \quad (16.2.2)$$

where Λ^F is the $N \times R$ matrix of factor loadings and Λ^z is an $N \times M$ matrix of coefficients. Bernanke, Boivin, and Eliasz (2005) assume that the upper $R \times R$ block of Λ^F is an identity matrix and the upper $R \times M$ block of Λ^z is a zero matrix, noting that these conditions are sufficient for the identification of the factors. Alternative conditions for identifying FAVAR models are discussed in Bai, Li, and Lu (2016).

The M observed variables z_t are usually a small set of variables of interest from an economic point of view that drive the dynamics of the system together with the unobserved factors F_t . The z_t variables must not be included in x_t because otherwise some of the idiosyncratic components would be zero and, hence, the covariance matrix of the idiosyncratic components would be singular. It should be noted that the factors F_t in model (16.2.1) need not be ordered first. They may also be placed below z_t or in between the elements of z_t .

Estimation. Estimation of model (16.2.1) requires knowledge of F_t . One approach taken in many studies is to first extract F_t by PC analysis from a large set of informational variables, x_t , that do not include the observed variables of interest, z_t , before estimating the FAVAR model (16.2.1) with the estimated factors replacing the true factors. This procedure is inspired by the work of Stock and Watson (2002b). It is also used by Favero, Marcellino, and Neglia (2005), Bernanke and Boivin (2003), and, in modified form, by Bernanke, Boivin, and Eliasz (2005).¹

¹ In a detailed analysis of the Bernanke, Boivin, and Eliasz (2005) procedure, Ouliaris, Pagan, and Restrepo (2014) show that this modified procedure implies a VARMA-DGP with dependent errors for the factors and the key observed variables, necessitating a sieve approach to estimating the FAVAR model.

The unknown factors F_t may be constructed from model (16.1.10) using the four-step procedure for the estimation of DFM discussed in Section 16.1.2. Bai and Ng (2006) show that the uncertainty in the estimates of F_t is negligible when N is large relative to T , allowing one to condition on the estimated factors as though they were known.

For example, Kilian and Lewis (2011) replace the real output variable in an otherwise standard VAR model of U.S. monetary policy by the Chicago Fed National Activity Index (CFNAI), which is constructed as the leading principal component of a wide range of monthly indicators of U.S. real activity, including five categories of data: output and income (21 series); employment, unemployment, and hours (24 series); personal consumption, housing starts, and sales (13 series); manufacturing and trade sales (11 series); and inventories and orders (16 series). They specify a VAR(12) with intercept for $y_t = (\Delta rcom_t, \Delta rpoil_t, F_t, \pi_t, i_t)'$, where $\Delta rcom_t$ is the percentage change in the real price of imported commodities (excluding crude oil), $\Delta rpoil_t$ is the percentage change in the real price of imported crude oil, F_t is the CFNAI real activity index, π_t is the CPI inflation rate, and i_t is the federal funds rate. After fitting the reduced-form VAR model $A(L)y_t = v + u_t$, the structural shocks are recovered by applying a lower-triangular Cholesky decomposition to Σ_u .

This approach does not impose the additional restrictions implied by the observation equation (16.2.2). Boivin and Giannoni (2009) discuss an alternative estimation procedure that is intended to account for all restrictions implied by the structural FAVAR model. They construct the initial value of the factors, $F_t^{(0)}$, by performing a PC analysis on x_t . They then estimate Λ^F and Λ^z by applying equation-by-equation LS to

$$x_t = \Lambda^F F_t^{(i-1)} + \Lambda^z z_t + e_t. \quad (16.2.3)$$

Denoting the estimate of Λ^z by $\widehat{\Lambda}^z$, they then perform a PC analysis of $x_t - \widehat{\Lambda}^z z_t$, resulting in factors $F_t^{(i)}$. The last step is repeated for $i = 1, 2, \dots$, until convergence.

This alternative procedure has its own limitations, however. As Ouliaris, Pagan, and Restrepo (2014) point out, the effects of z_t on the unobserved factors $F_t^{(i)}$ will not tend to zero as i increases. Hence, there is no clear separation in this estimation procedure between the unobserved factors and the observed variables z_t . Thus, this alternative method is not necessarily superior to the simpler approach that ignores the observation equation (16.2.2) in estimating the state-space model.

Bai, Li, and Lu (2016) discuss an alternative two-step estimation procedure for FAVAR models. In the initial stage of the procedure the observed variables z_t are concentrated out of equation (16.2.2) and the unobserved factors F_t are estimated by a quasi-ML procedure rather than by PC estimation. Then a reduced-form version of model (16.2.1) with $B_0 = I_{R+M}$ is estimated

by LS and the estimates are used to update the initial estimates of the factors and parameters of the FAVAR model (16.2.1). Bai et al. prove the consistency and asymptotic normality of this estimator under alternative sets of identifying conditions, provided that suitable moment conditions hold and that the data are stationary. They also show that their quasi-ML procedure is asymptotically more efficient than PC estimation.

Extensions. There are several extensions of this framework. First, Dufour and Stevanović (2013) generalize the FAVAR model to the factor-augmented vector autoregressive moving average (FAVARMA) model. They note that, if the factors are driven by a finite-order VAR process, the informational variable x_t follows a VARMA process. They propose a suitable estimation procedure for such models and provide evidence that the FAVARMA model may have higher forecast accuracy than the FAVAR model.

Second, Banerjee and Marcellino (2008) and Banerjee, Marcellino, and Masten (2014a, 2014b) consider factor-augmented cointegrated VAR models in error correction form. They discuss estimation, forecasting, and structural analysis based on such factor-augmented VECMs (FECMs). The advantage of FECMs is that they explicitly allow for integrated variables, whereas standard dynamic factor models are designed for stationary variables. An obvious advantage of including integrated variables in levels is that the models can also capture cointegration relations.

Third, Balabanova and Brüggemann (2017) consider FAVAR models with factors entering as exogenous variables

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + \Pi_0 F_t + \cdots + \Pi_q F_{t-q} + u_t,$$

resulting in a FAVARX model.

Structural Identification. Structural shocks in the FAVAR model (16.2.1) can be recovered as in the standard VAR model by a linear transformation of the reduced-form errors from the model

$$A(L)y_t = u_t,$$

where $y_t = (F'_t, z'_t)'$. As before, the structural shocks are obtained as $w_t = B_0 u_t$, given suitable identifying restrictions on B_0^{-1} or B_0 . Of course, the number of variables and, hence, structural shocks to be identified may be larger in FAVAR models, rendering the identification potentially more difficult. In many applications the FAVAR model is only partially identified.

For example, Favero, Marcellino, and Neglia (2005) specify a FAVAR model for

$$(F'_t, z_t^{*\prime}, r_t)',$$

where z_t^* contains all directly observed variables apart from the interest rate r_t . They order the monetary policy instrument last in their FAVAR model and impose a recursive ordering because they are only interested in the effects of monetary policy shocks. This ordering implies that a monetary policy shock has no instantaneous impact on any of the observed variables or on the factors.

Of course, the assumption that none of the factors and observed variables reacts to a monetary policy shock within the same period may be questioned, in particular, when fast-moving financial variables are included in the model. Alternatively, one could split up the variables into fast-moving and slow-moving variables, as proposed in Bernanke, Boivin, and Eliasz (2005), and extract factors separately from the two groups of variables. Then one could order the slow-moving factors (f_t^s) above the interest rate and the fast-moving factors (f_t^f) below it, so that they can be instantaneously affected in a lower-triangular recursive identification scheme. In other words, the variables in the FAVAR could be arranged as follows:

$$(f_t^{s'}, z_t^{s'}, r_t, f_t^{f'}, z_t^{f'})',$$

where z_t^s and z_t^f contain the slow- and fast-moving observed variables, respectively, in the system. Of course, this identification prevents the central bank from responding to fast-moving variables such as stock returns or changes in commodity prices and, hence, is not particularly credible either.

Suitable restrictions separating slow- and fast-moving factors can also be implemented by imposing zero restrictions on the factor loadings. In other words, one may specify that some factors load only on fast-moving variables and others load only on slow-moving variables. Imposing such restrictions requires suitable estimation algorithms that allow for a restricted loading matrix such as ML or Bayesian methods.

The responses of the factors to structural shocks in the FAVAR setup (16.2.1) in general differ from the responses of the informational variables, x_t . The latter may be approximated using the observation equation (16.2.2) as

$$x_t \approx [\Lambda^F, \Lambda^z] A(L)^{-1} B_0^{-1} w_t. \quad (16.2.4)$$

The approximation error relates to the omission of the idiosyncratic error term, e_t , in the observation equation. The ability of FAVAR models to generate (approximate) structural responses for the large number of variables contained in x_t makes these models particularly useful for practitioners. For example, we may trace the effect of a monetary policy shock on a large number of disaggregate measures of real output or inflation if the latter are contained in x_t .

Of course, in many applications we may not be interested in the responses of x_t , but are satisfied with reporting the responses of z_t and F_t . One approach in the literature has been to treat F_t like an observed variable. In that case, the identification of the structural shocks and the construction of the structural

impulse responses and related statistics of economic interest is no different than in earlier chapters. For example, Mumtaz and Surico (2009) and Eickmeier and Hofmann (2013) use a mixture of exclusion and sign restrictions in a FAVAR setting. Mumtaz and Surico (2009) also consider recursive and nonrecursive models identified by short-run exclusion restrictions.

It should be noted, however, that the structural impulse responses implied by the system of equations (16.2.1) and (16.2.2) may not coincide with the structural impulse responses $B(L)^{-1}$ implied by model (16.2.1) alone because the DGP is not a conventional VAR model, but a state-space model with latent structure. Although the responses of z_t and F_t to structural shocks in the system of equations may in principle be computed by iteration, this approach does not appear to have been used in applied work. Nor is it clear how to achieve identification of the structural shocks in the latter framework.

Inference on Structural Impulse Responses. If we treat the estimated factors as though they were known, inference in structural FAVAR models may proceed exactly as in Chapter 12. In practice, the use of bootstrap methods is standard. If the factors are subject to estimation uncertainty, in contrast, the bootstrap procedure has to account for the full DGP described by the state-space model (16.2.1) and (16.2.2). Several studies including Bernanke, Boivin, and Eliasz (2005) have used the bootstrap to approximate the distribution of structural impulse responses in FAVAR models, but without stating explicitly their algorithm or establishing its asymptotic validity.

Building on related ideas in Yamamoto (2016), such an algorithm may look as follows. Consider the reduced-form state-space model with transition equation

$$A(L) \begin{bmatrix} F_t \\ z_t \end{bmatrix} = u_t,$$

and observation equation

$$x_t = \Lambda^F F_t + \Lambda^z z_t + e_t,$$

where u_t and e_t are mutually uncorrelated. Then a bootstrap procedure involves the following steps:

Step 1. Estimate the factors and parameters of the state-space model by one of the procedures discussed earlier, denote the estimates of parameters of the observation equation by $\widehat{\Lambda}^F$ and $\widehat{\Lambda}^z$, and denote the corresponding (mean-adjusted) residuals by \widehat{u}_t and \widehat{e}_t .

Step 2. Draw bootstrap samples u_t^* and e_t^* , $t = 1, \dots, T$, with replacement from the sets of residuals \widehat{u}_t and \widehat{e}_t , respectively, and compute a bootstrap sample of factors, F_1^*, \dots, F_T^* , and observed variables, z_1^*, \dots, z_T^* , based on the transition equation in the usual way. Then compute the bootstrap

sample of the informational variables based on the observation equation as $x_t^* = \widehat{\Lambda}^F F_t^* + \widehat{\Lambda}^z z_t^* + e_t^*$ for $t = 1, \dots, T$.

Step 3. Use the bootstrap realizations x_t^*, z_t^* for $t = 1, \dots, T$ to reestimate the factors and the parameters of the state-space model and to recompute the structural impulse responses.

These steps are repeated a large number of times to obtain the bootstrap distribution of the quantities of interest. There are no formal results about the validity of this bootstrap.

16.2.2 Structural Analysis with DFM

Structural impulse response analysis can alternatively be performed within the framework of a DFM or GDFM. There is one important difference. Whereas in FAVAR models we are concerned with structural shocks to either F_t or z_t , in the current setting all observed variables are in x_t and the structural innovations are driving the primitive factors f_t .

Identification. Recall that the DFM (16.1.7) can be written in compact form as

$$x_t = \Lambda^f f_t + v_t,$$

with the dynamic factors being generated as

$$\Gamma(L)f_t = \eta_t \quad \text{and} \quad A(L)v_t = u_t,$$

such that $f_t = \Gamma(L)^{-1}\eta_t$. Substituting the latter expression into equation (16.1.7) yields

$$x_t = \Phi(L)\eta_t + v_t, \tag{16.2.5}$$

where $\Phi(L) = \Lambda^f \Gamma(L)^{-1}$. For our discussion of structural forms and of the identification of the structural shocks we treat the reduced-form parameters $\Phi(L)$ and Σ_η as known. If necessary, they can be estimated as discussed in Section 16.1.2. It should be noted that the representation (16.2.5) remains valid when working with GDFMs. The only difference is how the parameters of representation (16.2.5) are derived from the underlying model.

Assuming as usual that the $r \times 1$ vector of reduced-form residuals η_t is related to the $r \times 1$ vector of structural shocks w_t by a linear transformation $B_0\eta_t = w_t$, the structural form corresponding to model (16.2.5) is

$$x_t = \Phi(L)B_0^{-1}w_t + v_t. \tag{16.2.6}$$

If the structural shocks are instantaneously uncorrelated and the variances are normalized to 1, we know that $w_t \sim (0, I_r)$. In that case, B_0 has to satisfy

$B_0^{-1}B_0^{-1\prime} = \Sigma_\eta$ and we need at least $r(r - 1)/2$ more restrictions for identifying the $r \times r$ matrix B_0 . In other words, identifying the structural shocks requires putting enough restrictions on B_0 or its inverse to obtain uniqueness. Because the primitive factors f_t have no natural economic interpretation, economic identifying restrictions cannot be imposed on the effects on f_t , but only on the effects of the structural shocks on x_t . These restrictions may come in the form of exclusion restrictions on the impact effects or on the long-run effects of the structural shocks on x_t . They may also be available in the form of sign restrictions. Some specific identification strategies are discussed next.

Restrictions on the Impact Effects of Shocks. Note that the impact matrix Φ_0 in $\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i$ will in general not be an identity matrix. In fact, $\Phi(L)$ is $N \times r$ and is typically not a square matrix. Therefore the impact effects of the structural shocks, w_t , on x_t are given by $\Phi_0 B_0^{-1}$, and exclusion restrictions on the impact effects are zero restrictions on the elements of the matrix product $\Phi_0 B_0^{-1}$. For example, one may want to impose a recursive identification scheme on the impact effects, as is often done in conventional structural VAR analysis. This amounts to choosing a suitable nonsingular $r \times r$ submatrix of $\Phi_0 B_0^{-1}$ to be lower triangular. Such restrictions would suffice for identifying B_0 and, hence, the structural shocks. Denoting the $r \times r$ submatrix of Φ_0 that is of interest in the present context by $\bar{\Phi}_0$, the corresponding B_0^{-1} matrix can be obtained by noting that $\bar{\Phi}_0 B_0^{-1} B_0^{-1\prime} \bar{\Phi}'_0 = \bar{\Phi}_0 \Sigma_\eta \bar{\Phi}'_0$. Thus, computing a Cholesky decomposition of this matrix and left-multiplying by $\bar{\Phi}_0^{-1}$ yields a suitable B_0^{-1} matrix.

An illustration of this approach is provided in Forni and Gambetti (2010) who specify a GDFM for a panel of 112 monthly series. In their benchmark model the number of primitive dynamic factors is four. They use industrial production, a consumer price index, the federal funds rate and the Swiss/U.S. real exchange rate as the first four variables in their panel in that order. The structural shocks are identified recursively with the monetary policy shock specified to be the third shock. This policy shock is identified as a shock that does not have an instantaneous effect on industrial production and on the price level, but may induce immediate responses of the real exchange rate. All four shocks are allowed to have impact effects on the other variables.

More generally, exclusion restrictions can be imposed on the impact effects by choosing a suitable $\frac{1}{2}r(r - 1) \times r^2$ selection matrix J such that

$$J \text{vec}(\bar{\Phi}_0 B_0^{-1}) = 0$$

which implies restrictions

$$J(I_r \otimes \bar{\Phi}_0) \text{vec}(B_0^{-1}) = 0$$

for B_0^{-1} .

As in standard structural VAR models identified by exclusion restrictions on the impact effects, we may get away with imposing fewer than $r(r - 1)/2$ structural restrictions if fewer than r shocks are of interest in a particular application. The other shocks can then be identified arbitrarily. For example, in the Forni and Gambetti (2010) study, because only the monetary policy shock is of interest, the restrictions used for making the other shocks unique are not important.

It is also possible in the current framework to identify the shocks by placing restrictions directly on the factor loadings Λ^f in a structural version of model (16.1.7), exploiting the fact that $\Phi_0 = \Lambda^f$ and normalizing Σ_η to be diagonal. This approach may be more natural if the common factors have a direct economic interpretation. For example, a study of the international business cycle by Kose, Otrok, and Whiteman (2003), based on a panel of macroeconomic aggregates for 60 countries from 7 different regions in the world, postulates the existence of one world factor, a factor for each region, and a country-specific factor for each country. Only the world factor is allowed to have a direct impact on all variables in x_t . The effects of the other factors are restricted by imposing suitable zero constraints on the corresponding loadings. Of course, this approach is not structural in the conventional sense.

Restrictions on the Long-Run Effects of Shocks. Long-run restrictions as in Blanchard and Quah (1989) may also be used in the current context, provided x_t includes at least some variables that are expressed in growth rates. The approach of restricting the long-run cumulative effects of the structural shocks on x_t is more natural here than specifying a vector error correction model because x_t typically does not include $I(1)$ variables. This method involves restricting an $r \times r$ submatrix of $\Phi(L)$, say $\bar{\Phi}(L)$, such that $\bar{\Phi}(1)$ is nonsingular and $\bar{\Phi}(1)B_0^{-1}$ is lower-triangular. In this case the corresponding B_0^{-1} is obtained by computing a Cholesky decomposition of $\bar{\Phi}(1)B_0^{-1}B_0^{-1}\bar{\Phi}(1)' = \bar{\Phi}(1)\Sigma_\eta\bar{\Phi}(1)'$ and left-multiplying by $\bar{\Phi}(1)^{-1}$.

Identification Through Instruments. Identification of the $r \times r$ matrix B_0^{-1} and thereby of the structural shocks can also be achieved by using instruments with suitable properties, as discussed in Chapter 15. Suppose that a variable ζ_t is available that is correlated with the k^{th} structural shock, but uncorrelated with all other shocks such that

$$\mathbb{E}(w_{it}\zeta_t) = \begin{cases} \rho \neq 0 & \text{for } i = k, \\ 0 & \text{for } i \neq k. \end{cases}$$

Then, using $\eta_t = B_0^{-1}w_t$ and denoting the columns of B_0^{-1} by b_i , $i = 1, \dots, r$, we obtain

$$\mathbb{E}(\eta_t\zeta_t) = B_0^{-1}\mathbb{E}(w_t\zeta_t) = b_k\rho. \quad (16.2.7)$$

Thus, a multiple of the k^{th} column of B_0^{-1} can be obtained as the covariance of the reduced form error η_t and the instrument ζ_t . If a suitable instrumental variable is available that is only correlated with the k^{th} structural shock and is uncorrelated with all other structural shocks, a natural estimator of $b_k \rho$ is

$$T^{-1} \sum_{t=1}^T \eta_t \zeta_t,$$

where η_t can be replaced by the estimator $\hat{\eta}_t = \hat{W}'\hat{\mathcal{E}}_t$, as discussed in Section 16.1.2. Stock and Watson (2012) call such instruments external instruments if they are not part of the database x_t used in the factor analysis (see Chapter 15).

Note that for identification it is sufficient to know a multiple of the k^{th} column of B_0^{-1} because the responses of the variables to the k^{th} shock are obtained as $\Phi(L)b_k$ and multiplying the shocks by some constant only changes the magnitude of the shock, but does not change the shape of the response function. Hence, the magnitude of the shock can be chosen freely. Stock and Watson (2012) propose to normalize the initial response of one of the variables to one. For example, a monetary policy shock may be chosen such that the policy instrument changes by one unit on impact.

So far we have just explained how to identify the columns of B_0^{-1} and, hence, the structural impulse responses. Stock and Watson (2012) suggest that the corresponding structural shocks can be determined by regressing ζ_t on η_t and using the predicted values of that regression as estimates of the structural shocks (see Section III.a of Stock and Watson (2012) for details). Of course, different sets of instruments will imply different estimates of the structural shocks. As observed by Sims (1998), there is no limit to how different the estimated structural shock time series might appear, when using alternative equally valid instruments.

In principle, identification using external instruments is straightforward if suitable instruments can be found. Stock and Watson (2012) use as instruments external estimates of exogenous shocks of the type discussed in Chapter 7. For example, as an instrument for the productivity shock in the factor model they consider an estimate of the productivity shock series from the DSGE model of Smets and Wouters (2007). As instruments for the monetary policy shocks they use the corresponding shocks identified by Sims and Zha (2006b) and the shock to the monetary policy reaction function of the Smets and Wouters (2007) model.

Other Identification Strategies. Earlier we discussed the identification of structural shocks based on exclusion restrictions. In principle it is also possible to use sign restrictions in DFMAs (see Yamamoto 2016). This approach does not appear to have been used in applied work so far.

It is also possible to identify structural shocks based on a penalty function. For example, Giannone, Reichlin, and Sala (2004) postulate the existence of two factors in a system of U.S. macroeconomic variables and, hence, of two structural shocks. They identify the first shock as a real shock that maximizes the share of the variance of the real variables explained by the factor. The other shock is taken to be the nominal shock.

Another proposal is to link the factors to economic variables in a DSGE model (see Bäurle 2013). Other authors have considered so-called multi-level or hierarchical factor models. In these models, the variables are partitioned into blocks and there are block-specific and global common factors (e.g., Moench and Ng 2011; Hallin and Liška 2011). For example, Hallin and Liška investigate industrial production in a multiple-country study in which the blocks refer to different countries. Such models allow the separate identification of block-specific and global shocks.

Inference on Structural Impulse Responses. Structural impulse responses are functions of the structural parameters of the factor model and are obtained from the parameter estimates. Thus, asymptotic properties can be derived for the impulse responses in the usual way from the asymptotic properties of the parameter estimators. Since in frequentist impulse response analysis bootstrap methods are typically used for constructing confidence intervals around impulse responses, the question arises of how to bootstrap a DFM or GDFM. Based on earlier work of Gonçalves and Perron (2014), Yamamoto (2016) proposes and compares two procedures for a DFM in static form.

For expository purposes, consider the simplified model

$$x_t = \Lambda^f f_t + v_t, \quad f_t = \Gamma_1 f_{t-1} + \cdots + \Gamma_s f_{t-s} + \eta_t, \quad (16.2.8)$$

where v_t and η_t are mutually uncorrelated white noise processes. Moreover, it is assumed that the reduced-form parameters are identified by suitable restrictions, and the structural impulse responses are just-identified. Then a residual-based bootstrap for the structural impulse responses can be executed in the following steps.

Step 1. Estimate the factors and parameters of model (16.2.8) by one of the procedures presented in Section 16.1.2 and denote the parameter estimates by $\widehat{\Lambda}^f$ and $\widehat{\Gamma}_i$, $i = 1, \dots, s$. Let the corresponding (mean-adjusted) residuals be \widehat{v}_t and $\widehat{\eta}_t$.

Step 2. Draw a bootstrap sample v_t^* and η_t^* , $t = 1, \dots, T$, with replacement from the sets of residuals \widehat{v}_t and $\widehat{\eta}_t$, respectively, and compute the bootstrap sample of factors and observed variables in the usual way as $f_t^* = \widehat{\Gamma}_1 f_{t-1}^* + \cdots + \widehat{\Gamma}_s f_{t-s}^* + \eta_t^*$ and $x_t^* = \widehat{\Lambda}^f f_t^* + v_t^*$.

Step 3. Use the bootstrap realizations x_t^* for $t = 1, \dots, T$ to reestimate the parameters and the factors of model (16.2.8). Then recompute the structural impulse responses.

These steps are repeated a large number of times to obtain the bootstrap distribution of the quantities of interest. Yamamoto (2016) makes the case that this procedure is asymptotically valid. He also suggests that other bootstrap procedures could be used to accommodate more general assumptions about the error term. Stock and Watson (2016, section 5.1.3) propose a related bootstrap procedure that allows for autoregressive dynamics in the idiosyncratic errors, v_t , and involves drawing the bootstrap innovations η_t^* as well as the bootstrap innovations v_t^* of the autoregressive process from normal distributions.

In Step 3 of the procedure the parameters and factors of the model are reestimated from the bootstrap sample x_t^* . Yamamoto (2016) also considers an alternative procedure in which the factors are not reestimated, but instead the bootstrap factors are used in estimating the model parameters. In other words, Step 3 of the procedure is replaced by

Step 3*. Use the bootstrap sample x_t^* and factors f_t^* to estimate the parameters of model (16.2.8) by regression methods and then compute the quantities of interest from these estimates.

He argues that this procedure is also asymptotically valid. Based on a simulation study he finds that the procedure that reestimates the factors as in Step 3, has higher coverage accuracy in some situations than the procedure based on Step 3*. There do not appear to be results in the literature on bootstrapping structural impulse responses from DDFMs.

A Critique of Structural Analysis with DFM_s. The identifying assumptions discussed earlier are, of course, critical for the structural analysis considered in this chapter. One could question the identification of structural shocks as a linear transformation of the errors η_t driving the factors in model (16.2.5). It is difficult to see why this assumption makes sense in approximate DFM_s that allow more structure in the part of the model not explained by the common factors. Such approximate DFM_s make explicit that the relationship between the variables is only partly captured by the common factors. Hence, the transmission of the shocks to x_t is only partly captured by $\Phi(L)B_0^{-1}$. Even if the DGP is an exact DFM, the Wold MA representation of the observed variables x_t is different from representation (16.2.5). Hence, building on the MA representation of x_t results in a different transmission of the shocks.

Stock and Watson (2005) rightly point out that if the DFM is taken as the DGP, the factors contain all the dynamic interaction between the variables. Conditioning on the factors, none of the variables is Granger-causal for any other variable. Still this does not mean that the shocks are best extracted

from the model driving the factors. It may well be that important shocks enter through the idiosyncratic component of x_t rather than through the η_t errors in the representation (16.2.6). For example, a discretionary change in monetary policy can be viewed as an idiosyncratic shock to the interest rate that affects both the factors and x_t .

Although dynamic factor models help address the missing information problem in low-dimensional VAR models by allowing us to include a large set of variables in the analysis, extending the data set also creates new problems. For example, not much is known about the sensitivity of structural impulse responses with respect to changes in the information set and/or the model structure. In particular, the choice of the number of factors and the question of which factors to include play an important role. A sensitivity analysis with respect to the set of variables, the sampling period, the model specification, and the type and number of factors is therefore recommended.

16.2.3 Empirical Examples of FAVAR Models and DFMs

As already discussed, one of the reasons why researchers have been attracted to structural FAVAR and dynamic factor models is the ability to mitigate informational deficiencies in conventional structural VAR models. Another reason is that these models may be used to characterize the response of many economic variables to a given shock. For example, we may describe the responses of the economy by sector or of the inflation rate by expenditure item. This feature helps alleviate concerns that conventional structural VAR models are silent on the response of many of the variables that users of macroeconomic models are interested in and that traditional DSEMs were able to capture (see Chapter 6).

There are many examples of structural analysis with factor models in the literature. Many of these studies deal with one of two applications: monetary policy analysis and international business cycle analysis. Next, a small number of examples is reviewed to convey the flavor of these studies.

The Effects of Monetary Policy. Structural FAVAR analysis was first proposed by Bernanke, Boivin, and Elias (2005). For their analysis of U.S. monetary policy they extracted several factors from a panel of 120 monthly time series. Their baseline FAVAR model includes industrial production, the consumer price index (CPI), the federal funds rate, and one of the factors. In another specification they add three factors to the federal funds rate. Identification of the monetary policy shock is achieved by assuming that none of the other variables or factors respond instantaneously to a monetary policy shock. This assumption is consistent with a recursive identification scheme in which the federal funds rate is ordered last. Bernanke et al. find that taking into account the additional information summarized in the factors makes a substantial difference for the structural impulse responses and paint a different picture of the

transmission of monetary policy shocks than a conventional small-scale VAR model.

Del Negro and Otron (2007) use a FAVAR model for quarterly U.S. variables to investigate the impact of monetary policy shocks on house prices. They include information on state-level house prices in a VAR model of monetary policy. They find that part of the house price changes in each state can be attributed to a national factor constructed from these price series. They include this factor in a FAVAR model consisting of six variables: the factor, total reserves, CPI inflation, the GDP growth rate, a 30-year mortgage rate, and the federal funds rate. They use sign restrictions for identifying the monetary policy shock (see Chapter 13). Specifically, they assume that the federal funds rate increases and the growth rate of total reserves, changes in CPI inflation, and changes in GDP growth do not increase for several quarters after a monetary policy shock. They find that monetary policy shocks have an effect on house prices and hence contribute to housing price booms, although the impact is small.

Favero, Marcellino, and Neglia (2005) use a FAVAR model to investigate the impact of monetary policy shocks in the United States and four large European economies (Germany, France, Italy, Spain). They compare the Stock-Watson principal-components approach with estimates of the dynamic factors constructed as in the work of Forni, Hallin, Lippi and Reichlin. The common factors are extracted from large monthly panels of variables from the U.S. and the four European countries. The country-specific FAVAR models include small sets of macroeconomic variables such as output, inflation, commodity price inflation, an exchange rate, and, in addition, the interest rate in the case of the United States and selected foreign variables for European countries. For example, U.S. inflation is included in the model for Germany. The study presents estimates for models augmented with alternative sets of factors. Identification is achieved by a recursive scheme with the interest rate ordered last. The authors conclude that including common factors can make a difference for impulse response analysis. For example, it may remove the ‘price puzzle’, which refers to an increase in inflation after a contractionary monetary policy shock. This phenomenon is often attributed to omitted variables bias (see Chapter 8). Including further information in the form of factors can be thought of as mitigating this omitted variables bias.

Boivin, Giannini, and Mihov (2009) investigate the impact of macroeconomic factors and monetary policy shocks on sectorally disaggregated consumer and producer prices. They construct a FAVAR model based on a large number of monthly U.S. time series for the period 1976m1-2005m6. The number of factors is five. The federal funds rate is viewed as the policy instrument. Identification of the monetary shocks is achieved by the assumption that none of the common factors reacts instantaneously to surprise changes in the policy rate, which amounts to a lower-triangular recursive scheme where the interest rate is ordered last. Boivin et al. find that the reaction of sector specific prices

to macroeconomic shocks and sector-specific shocks is very different. The response of disaggregated prices to a monetary shock is delayed and little evidence is found for a price puzzle.

International Business Cycle Analysis. The objective of business cycle analysis with dynamic factor models is typically to find a factor that describes the business cycle fluctuations globally or in a large region. For example, Kose, Otkrok, and Whiteman (2003) use a dynamic factor model and Bayesian estimation techniques to investigate the business cycle fluctuations in a set of 60 countries that covers seven regions of the world. They consider aggregate output, consumption, and investment variables and find a dynamic factor that explains some of the fluctuations in the aggregates in most countries and can thus be viewed as a world business cycle factor. They decompose the variance into components that can be attributed to the different factors and thereby determine how much of the variance in specific variables is determined by the business cycle factor and how much is accounted for by other factors. They find that a large part of the fluctuations in many aggregate variables can be attributed to the global business cycle factor, whereas region-specific factors are less important in determining fluctuations in economic activity. In a related study Kose, Otkrok, and Whiteman (2008) investigate possible differences in the business cycle dynamics over specific historic periods. Other FAVAR studies of the international business cycle include Eickmeier (2007), Mansour (2003), Helbling and Bayoumi (2003), and Bordo and Helbling (2010).

Transmission of Oil Supply Shocks. In a comprehensive study, Stock and Watson (2016) investigate the interaction between the global oil market and the U.S. economy using structural DFM and FAVAR models, building on the analysis in Kilian (2009). They find that global oil supply shocks explain only a fraction of the variation in the price of oil and a very small fraction of the variation in major U.S. macroeconomic aggregates.

16.3 Large Bayesian VAR Models

As mentioned earlier, instead of frequentist estimation methods, one may use Bayesian methods for estimating dynamic factor models or FAVAR models (see, e.g., Otkrok and Whiteman 1998; Kose, Otkrok, and Whiteman 2003; Amir Ahmadi and Uhlig 2009). If Bayesian methods are used, it is not obvious, however, that one would want to focus on factor models. Recall that the motivation for using factor models is that they allow us to investigate large panels of variables in a structural VAR analysis. In the context of Bayesian estimation, suitable priors on VAR models may serve the same purpose (see, e.g., De Mol, Giannone, and Reichlin 2008), even when not imposing a factor structure. In fact, as pointed out by Bańbura, Giannone, and Reichlin (2010), using Bayesian shrinkage methods to overcome the degrees-of-freedom

problem in structural VAR analysis has several advantages. For example, having no limits on the number of observed variables included in the VAR model allows one to include all variables considered by macroeconomists. Even sectoral information can be included and the impact of specific shocks such as monetary policy shocks on the disaggregated variables can be traced. Thereby large-scale international comparisons become possible without imposing ad hoc restrictions on the structure of the model.

Another advantage of starting from an unrestricted VAR model rather than summarizing some of the information in factors is that levels variables can be included easily. Recall that standard factor analysis tends to be based on stationary variables without stochastic trends. Thereby they may miss out on common trend structures except to the extent that known cointegration constraints are imposed by including cointegration relations as stationary variables. Although factor analysis can in principle also be applied to trending variables, the additional assumptions regarding the stochastic trends required for proper inference can be restrictive. From a practical point of view it may therefore be advantageous to work with the variables in levels, thereby potentially accommodating unit roots, long-range dependence, near unit root behaviour and the like.

A crucial problem with using large-scale BVAR models is the choice of the prior. This issue is reviewed in Section 16.3.1. Structural identification is discussed in Section 16.3.2.

16.3.1 Priors for Large Bayesian VARs

Baímbura, Giannone, and Reichlin (2010) use the so-called Minnesota or Litterman prior as their point of departure, which postulates a reduced-form Gaussian VAR model,

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

and imposes a normal prior with a random walk prior mean,

$$A^* = [0, I_K, 0, \dots, 0],$$

where $A = [v, A_1, A_2, \dots, A_p]$ and hence, the prior mean of A_1 is the identity matrix (see Chapter 5). This prior has to be modified if there are variables that are not persistent. For the latter variables the prior mean of the corresponding diagonal element of A_1 is set to zero instead of 1. The prior variance of the i^{th} element of A_l is

$$v_{ij,l} = \begin{cases} (\lambda/l)^2 & \text{if } i = j, \\ (\lambda\theta\sigma_i/l\sigma_j)^2 & \text{if } i \neq j, \end{cases}$$

where λ is the prior standard deviation of $a_{ii,1}$, $0 < \theta < 1$, and σ_i^2 is the i^{th} diagonal element of the reduced form residual covariance matrix Σ_u . If Σ_u is

known, the posterior is also normal and quite easy to deal with. Thus, if one is prepared to replace the covariance matrix by some known quantity such as a plausible estimator, the Bayesian estimation problem is basically solved. Of course, we cannot just replace Σ_u by its unrestricted LS estimator because this estimator is typically infeasible, given the degrees-of-freedom limitations. One alternative is to estimate the variances by fitting univariate AR models by LS and to assume that Σ_u is diagonal. This solution is sometimes used in practice (see, e.g., Koop 2013a).

The original Minnesota prior is regarded as unattractive by Ba  nbara, Giannone, and Reichlin (2010) because of the restrictive requirements for the reduced form residual covariance matrix. Instead they propose using a Gaussian-inverse Wishart prior, which is a natural conjugate prior if $\theta = 1$. Using that prior, the posterior mean is

$$\bar{A} = (A^*V(\lambda)^{-1} + YZ') (V(\lambda)^{-1} + ZZ')^{-1},$$

where $Y = [y_1, \dots, y_T]$, Z is the corresponding matrix of regressors, and the prior covariance matrix is $V(\lambda) \otimes \Sigma_u$. For given Σ_u and assuming $\theta = 1$, the prior covariance matrix depends only on the tightness parameter λ , which has to be provided by the user.

The posterior mean may be interpreted as a shrinkage estimator where the shrinkage is completely determined by λ . For large models the matrix ZZ' will not even be invertible and the posterior mean can only be determined by adding another matrix ($V(\lambda)^{-1}$) that makes the sum invertible and, hence, effectively determines the outcome of the estimation. In other words, the prior determines the estimation outcome to a large extent. Thus, the question is how to choose the shrinkage or tightness parameter. For example, if forecasting is the objective, one could choose λ such that the model tends to forecast accurately (see, e.g., Carriero, Kapetanios, and Marcellino 2009). Alternatively, λ may be chosen so as to maximize the marginal likelihood in a hierarchical modeling framework, as proposed by Giannone, Lenza, and Primiceri (2015) (see also Carriero, Kapetanios, and Marcellino 2012 for a similar approach). For models with hundreds of variables the latter procedure poses computational challenges, however.

In related work, De Mol, Giannone, and Reichlin (2008) and Ba  nbara, Giannone, and Reichlin (2010) propose decreasing the parameter λ , as the model grows larger, such that the estimated model has the same in-sample fit as a small VAR model estimated by LS that only includes the key economic variables of interest. This procedure works as follows.

Denote the posterior means of the parameters obtained from a model with tightness parameter λ and K variables by $v^{(\lambda,K)}$ and $A_i^{(\lambda,K)}$, $i = 1, \dots, p$, and the corresponding 1-step ahead predictions as

$$y_{t|t-1}^{(\lambda,K)} = v^{(\lambda,K)} + A_1^{(\lambda,K)}y_{t-1} + \dots + A_p^{(\lambda,K)}y_{t-p}.$$

Moreover, let $y_{k,t|t-1}^{(\lambda,K)}$ be the k^{th} component of $y_{t|t-1}^{(\lambda,K)}$, i.e., $y_{k,t|t-1}^{(\lambda,K)}$ is the 1-step ahead prediction of the k^{th} variable of a system with K variables and prior tightness parameter λ . The corresponding full-sample mean squared prediction error (MSPE) is

$$\text{MSPE}_k^{(\lambda,K)} = \frac{1}{T-p} \sum_{t=p+1}^T (y_{k,t|t-1}^{(\lambda,K)} - y_{k,t})^2.$$

Suppose that there is a small number, K^* , of variables of central interest with index set \mathcal{K} . Then the tightness parameter for the large model including all K variables, λ_K , is chosen such that

$$\lambda_K = \arg \min_{\lambda} \left| \text{FIT} - \frac{1}{K^*} \sum_{k \in \mathcal{K}} \frac{\text{MSPE}_k^{(\lambda,K)}}{\text{MSPE}_k^{(0,K)}} \right|, \quad (16.3.1)$$

where $\text{MSPE}_k^{(0,K)}$ corresponds to the MSPE of the large model evaluated at its prior mean and $\text{MSPE}_k^{(\lambda,K)}$ corresponds to the MSPE of the large model for a given choice of λ .

The benchmark fit is defined analogously as

$$\text{FIT} = \frac{1}{K^*} \sum_{k \in \mathcal{K}} \frac{\text{MSPE}_k^{(\infty,K^*)}}{\text{MSPE}_k^{(0,K^*)}},$$

where the small model only includes $K^* < K$ variables, $\text{MSPE}_k^{(\infty,K^*)}$ corresponds to the MSPE of the small model estimated by LS, and $\text{MSPE}_k^{(0,K^*)}$ corresponds to the MSPE of the small model evaluated at its prior mean. The actual minimization with respect to λ in (16.3.1) can be done by a grid search over λ because only one parameter is involved (see Koop 2013a).

In practice, after estimating the small model, λ is chosen for the large-scale model such that the in-sample fit for the equations corresponding to the central variables remains constant. This procedure worked well in a forecasting experiment reported in Bańbura, Giannone, and Reichlin (2010). Note that their choice of the tightness parameter amounts to specifying a tighter prior for larger models with more variables and lags.

Koop (2013a) explores an analogous procedure for the original Minnesota prior. He expresses concern about the very restrictive nature of the Minnesota prior, however, which uses only one or two parameters to determine the degree of shrinkage. Koop therefore considers another prior based on a proposal by George, Sun, and Ni (2008) that allows different parameters to be shrunk differently. In Koop's forecast comparison the effectiveness of this prior tends to deteriorate, when the number of variables in the VAR increases. Given that we

are primarily interested in dealing with large BVAR models, we do not consider this prior.

In related work, Bańbura, Giannone, and Reichlin (2010) report that a sum-of-coefficients prior which is a variant of the Minnesota prior worked best in their forecasting application. They recommend that prior also for structural analysis. As discussed in Chapter 5, the sum-of-coefficients prior accommodates long-run relations more easily and may therefore improve the accuracy of impulse response estimates when cointegration relations exist in the data.

Finally, Korobilis (2013) proposes to combine BVAR models with Bayesian variable selection. In his approach an indicator variable is specified for each parameter that tells us whether the parameter is included or its coefficient is set to zero. The prior for the indicator variables can be combined easily with the Minnesota prior, for example. Korobilis (2013) presents a modification of this approach intended to make it feasible to deal with large panels of variables. However, the largest model he considers contains only 13 variables, which is far from the dimensions we have in mind in this section.

16.3.2 Structural Identification in Large BVARs

In large-scale BVAR models, the identification of the structural shocks is most easily achieved by linking the properties of the shocks to their impact effects. For example, exclusion restrictions can be specified for the impact effects. Bańbura, Giannone, and Reichlin (2010) are interested in the effects of monetary policy shocks. As in Bernanke, Boivin, and Eliasz (2005), they split their set of variables into those that move slowly in response to the monetary policy shock and those that may respond to this shock within the impact period such that

$$y_t = (y_t^s, r_t, y_t^f)',$$

where y_t^f contains the fast moving variables such as financial variables, y_t^s is the vector of slow moving variables such as prices and real variables, and r_t is the policy interest rate. They identify the monetary policy shock by a lower-triangular Cholesky decomposition of the reduced-form covariance matrix Σ_u . Thereby the fast-moving variables are allowed to be affected instantaneously while the slow-moving variables are assumed to be known to the policymaker at the time when a decision is made.

In principle, one could also identify the shock of interest by sign restrictions. Although plausible sign restrictions for a large set of variables may be available, as argued in Amir Ahmadi and Uhlig (2009), such an approach is complicated by the large dimension of the structural-form covariance matrix

and the corresponding dimension of the possible rotation matrices that have to be considered in computing admissible shocks. Thus, such an approach may be computationally infeasible with current technology. A possible solution may be to reduce these computational problems by combining sign restrictions with exclusion restrictions.

In summary, the large-scale BVAR model has some advantages, but also some drawbacks. On the positive side, it dispenses with the additional structure imposed by FAVAR and dynamic factor models and it avoids some of the ambiguities of structural models based on factors. On the negative side, the priors required for the estimation of large-scale VAR models may distort the structural impulse response estimates without the user being able to assess the extent of this bias. Thus, the prior induces an element of arbitrariness.

16.4 Alternative Large-Dimensional VAR Models

So far we have discussed how to deal with large panels of time series data either by imposing a factor structure or by applying Bayesian shrinkage methods in estimation. Other proposals for dealing with this problem include panel VAR models, global VAR models, and spatial VAR models (see Canova and Ciccarelli 2013; Pesaran, Schuermann, and Weiner 2004; Chudik and Pesaran 2011; or Pesaran 2015). Each of these models imposes specific types of restrictions on the VAR parameters to allow larger VAR models to be estimated. These restrictions in some cases are strong and unrealistic, and they may distort the structural impulse responses in unknown ways. Of course, if some restrictions on the model can be defended on economic grounds, this approach will be preferable to imposing a factor structure or to fitting BVAR models, so it is important to think carefully about the types of restrictions that are best suited for a specific application.

16.4.1 Panel VARs

Large panels of variables often arise in studies of different countries or regions, but also when dealing with sectors, firms, plants, or households. Such a situation makes it convenient to assign an additional subscript to a variable. For example, we may denote the t^{th} observation for the i^{th} variable of country n by y_{int} , where $i = 1, \dots, M$ and $n = 1, \dots, N$. Thus, using our earlier notation, $K = M \cdot N$. Let $y_{nt} = (y_{1nt}, \dots, y_{Mnt})'$ be an M -dimensional vector and denote by $Y_{n,t-1}$ and Y_{t-1} vectors of lags of y_{nt} and all variables in the panel, respectively. Then the model for y_{nt} has the general form

$$y_{nt} = v_n + \mathbf{A}_n Y_{t-1} + u_{nt}, \quad (16.4.1)$$

with fully general error covariance matrix Σ_u for the system of all N units, where Σ_u is the covariance matrix of $u_t = (u'_{1t}, \dots, u'_{Nt})'$.

This panel structure has three characteristics. First, lags of all endogenous variables of all panel units are allowed to affect y_{nt} . Second, u_{nt} is in general correlated across n . Third, the model parameters may be specific to each n , allowing for cross-sectional heterogeneity. In this general form, the model is not different from large-dimensional BVAR models.

Clearly, it is difficult to estimate the panel VAR model in its general form (16.4.1) without imposing additional prior structure. Even when Bayesian estimation methods are used, it is common to impose additional restrictions in estimation. For example, we may impose the restriction that in small open economies global variables are exogenously given. Often researchers further restrict the panel VAR model such that every panel unit corresponds to a separate VAR model

$$y_{nt} = v_n + A_n Y_{n,t-1} + u_{nt},$$

where the coefficient matrix A_n is much smaller than \mathbf{A}_n in (16.4.1), with the interaction across units determined by the covariances $\mathbb{E}(u_{nt} u'_{ms})$. Unlike model (16.4.1), this restricted specification does not explicitly model the dynamic interactions across the panel units.

Even more restrictions can be imposed if the assumption of dynamic homogeneity can be justified. In that case, all units are assumed to have the same VAR coefficients such that

$$y_{nt} = v_n + A Y_{n,t-1} + u_{nt}.$$

Suitable estimation methods for panel VAR models are discussed in Canova and Ciccarelli (2013). In typical economic applications, of course, one would not expect the dynamics to be homogeneous across countries, sectors, or firms.

Panel VAR techniques based on (16.4.1) have become increasingly popular recently. As noted by Canova and Ciccarelli (2013), panel VAR models are particularly suited for analyzing the transmission of unit-specific shocks across panel units and time. For example, Canova, Ciccarelli, and Ortega (2012) study how shocks to U.S. interest rates are propagated to ten European countries, some in the Euro area and some not. Panel VAR models have also been used to construct the average effects across heterogeneous groups of countries. They may also be used to assess the small open economy assumption or to assess the existence of convergence clubs.

The estimation of panel VAR models in practice can be computationally challenging. An additional challenge in structural panel VAR models is the identification of the structural shocks. Typically such models are only partially identified. For these and other issues the reader is referred to the surveys by Breitung (2015) and Canova and Ciccarelli (2013).

16.4.2 Global VARs

Global VARs (GVARs) represent another class of models designed specifically to handle panels of time series data for many countries, which explains its name (e.g., Dees, Di Mauro, Pesaran, and Smith 2007; Chudik and Pesaran 2016; Pesaran 2015). The idea is to augment a VAR for each unit by global variables capturing the other countries. In that respect GVARs have some similarity with FAVAR models. More specifically, the model for the n^{th} country is set up as

$$y_{nt} = v_n + A_n Y_{n,t-1} + W(L)y_{nt}^* + u_{nt},$$

where $W(L)$ is a matrix polynomial in the lag operator and y_{nt}^* summarizes global variables such as the price of oil and possibly aggregated variables from other countries. Typically the latter variables are weighted averages of the variables from other countries corresponding to y_{nt} of the form

$$y_{nt}^* = \sum_{\substack{j=1 \\ j \neq n}}^N \omega_{nj} y_{jt},$$

where ω_{nj} is the weight attached to country j in the model for the n^{th} country. For example, Dees, Di Mauro, Pesaran, and Smith (2007) link the weights to the share of country j in world trade. These weights may also be matrices, if different variables are weighted differently. Generally y_{nt}^* can be thought of as common factors consisting of global variables and foreign variables computed from data for the other countries in the panel. Thus, the common factors are not necessarily determined by a statistical procedure, but typically are based on economic considerations. Apart from this important difference, the models can be viewed as factor-augmented VAR models.

The advantage of this approach is that several countries or regions can be modeled jointly as a global model (or GVAR model), which is easily recognized as a large restricted VAR model that may also include some additional unmodeled variables. The parameter restrictions are partly due to the choice of weights in aggregating the foreign variables and partly they are just exogeneity restrictions.

Since the global model has a VAR structure it is also possible to use standard tools such as impulse responses, provided shocks of interest can be identified. The GVAR literature has typically used non-structural generalized impulse responses to study the dynamics of the system, but occasionally structural identifying assumptions have been invoked. For example, Dees, Di Mauro, Pesaran, and Smith (2007) identify a U.S. monetary policy shock by ordering the U.S. block first and imposing a recursive ordering on the U.S. variables. The effects on the other countries are left open and are allowed to be instantaneous. Such an approach may be justified if a dominant country such as the United States is considered. The identification of shocks becomes more

difficult, however, if the effects of monetary policy shocks in other countries on the United States are of interest or, more generally, the effects of many shocks. It has also been common to trace out responses to variation in variables that are assumed exogenous, not unlike what one would do in traditional DSEMs (see Chapter 6).

The difficulties in identifying structural shocks is one drawback of GVAR analysis. Another drawback is that the transmission of the shocks is affected by the assumptions made when aggregating the variables from other countries. A review of the GVAR approach which addresses many related technical issues and includes a large number of further references is provided in Chudik and Pesaran (2016). This study also summarizes several applications of the GVAR approach.

16.4.3 Spatial Models

There are also other ideas about how to restrict the number of variables or the parameter space in large-dimensional VAR models. For example, spatial models assume that a region depends more strongly on its close neighbours than on more distant regions (see Anselin 2006; Chudik and Pesaran 2011; Canova and Ciccarelli 2013). In other words, the distance between units is used to impose restrictions on the dependence across units, thereby reducing the parameter space. Such assumptions can be problematic. In many applications it is unsatisfactory to link the relationship between panel units to their physical distance alone. For example, U.S. shocks may affect many other countries in the world more than would be expected based on their physical distance alone. Although there are more general norms of distance accounting for multiple attributes, the implicit weights used in constructing a measure of distance tend to be inherently ad hoc. Thus, spatial VAR models have not played an important role in applied work to date.

16.5 Discussion

The proposals for dealing with large panels of time series variables considered in this chapter all amount to imposing restrictions on the VAR model. This can be accomplished either by reducing the dimension of the set of model variables or by shrinking the parameter space. Both approaches deal with the curse of dimensionality. At one extreme, factor models rely entirely on a reduction in the number of model variables; at the other extreme, large-dimensional BVAR models only shrink the parameter space, while preserving the model dimension. Other approaches are somewhat in-between. For example, GVAR models to some extent reduce the space of variables by restricting the parameter space, but they also impose prior structure by choosing the aggregation weights for the model variables.

All the models considered in this chapter have their pros and cons. For example, factor models extract information from the variables first and aggregate it in factors. These factors are then used in a model together with selected observed variables to construct moderately-sized models that can be analyzed with standard frequentist methods. Aggregation of the data, however, causes distortions in the structural responses, the quantitative importance of which is usually unknown. Moreover, standard factor analysis is tailored to stationary variables with time-invariant moments. Estimates may be sensitive to the assumptions regarding the order of integration or to the long-range dependence properties of the model variables.

In contrast, large-scale BVAR models completely avoid this aggregation bias. Moreover, variables may be included in levels regardless of their unit root and trending properties. Recall that degrees-of-freedom deficiencies make frequentist estimation impossible, when VAR models become large. The Bayesian solution is to impose a prior on the parameter space. The priors typically imposed in large-scale BVAR analysis shrink the VAR parameters to zero or to values corresponding to multivariate random walks. The drawback of large-scale BVAR models is that the priors required for estimation do not account for the actual economic structures underlying a panel of time series variables. Even if they are not meant to be restrictive, they may lead to substantial distortions in the estimates of the structural impulse responses and related statistics.

In a forecast comparison based on large panels of variables, Bańbura, Giannone, and Reichlin (2010) and Koop (2013a) find that large BVARs forecast more accurately overall than factor models. In fact, De Mol, Giannone, and Reichlin (2008) provide conditions based on asymptotic theory which ensure that Bayesian shrinkage for large panels of time series that are driven by a limited number of factors, results in optimal forecasts asymptotically, if both the number of variables and the time series dimension go to infinity. Although such results can be used to make a case for BVARs, it is not clear that out-of-sample forecast accuracy is the best criterion for evaluating the transmission mechanism of shocks.

A potential alternative to Bayesian shrinkage is the use of LASSO-type estimators, as proposed in Song and Bickel (2011).² A key difference between Bayesian shrinkage and the LASSO approach is that the latter not only shrinks the parameter estimates towards zero but may also eliminate parameters from the model, resulting in a potentially sparse large-dimensional VAR structure. Like the BVAR approach, the estimator of Song and Bickel facilitates structural impulse response analysis directly on the model variables. However, the use of a penalty term involves an additional auxiliary parameter that

² LASSO stands for Least Absolute Shrinkage and Selection Operator. LASSO minimizes the residual sum of squares subject to a penalty term involving the absolute value of the regression coefficients (see Tibshirani 1996).

determines the degree of shrinkage. This additional structure may distort the structural impulse responses much like the prior in high-dimensional BVAR models. In addition, it is not clear how robust the LASSO approach is to the presence of unit roots and cointegration among the model variables.

The alternative of imposing a panel structure also involves imposing restrictions on the parameter space. Unlike the large-scale BVAR model, the panel VAR model often is restricted further by imposing additional exclusion restrictions. To the extent that the possibly restricted structural panel VAR model tends to be estimated by Bayesian methods, there is also an important element of shrinkage, however.

Global VAR models in turn may be viewed as a combination of imposing direct restrictions on the parameter space and imposing a specific factor structure. The factors are not chosen based on purely statistical criteria, but on the basis of economic considerations. For example, the use of trade weights in aggregating macroeconomic time series from different countries is common. Alternatively, exclusion restrictions on model parameters may be suggested by the fact that some countries are small open economies. The fact that the factor structure is not selected based on purely statistical criteria, of course, does not mean that the aggregation cannot distort estimates of the structural impulse responses. Another problem in the GVAR literature is the lack of convincing strategies for the identification of structural economic shocks. As mentioned in Section 16.4.2, the so-called generalized impulse responses often reported in the GVAR literature are based on shocks that are only statistically identified and lack a clear economic interpretation.

For applied work the distinction between reducing the set of model variables and shrinking the parameter space is not of prime importance. What is important is the question of whether the restrictions required for estimation can be defended in a particular application. Thus, knowing the pros and cons of these alternative models is important.

17 Nonfundamental Shocks

17.1 Introduction

In Chapter 2 we motivated VAR models as approximations to a DGP of the form

$$y_t = \Phi(L)u_t,$$

where u_t is the one-step ahead prediction error based on information $\{y_{t-1}, y_{t-2}, \dots\}$. Structural VAR analysis is based on the premise that the structural shocks can be obtained by transforming the prediction errors u_t . The MA representation based on the VAR prediction errors is known as a fundamental representation. It is conceivable, however, that the shocks constituting the MA representation of the DGP do not coincide with the prediction errors of a given VAR model. For example, the DGP may be

$$y_t = \Phi^*(L)u_t^*,$$

where u_t^* is not the prediction error based on $\{y_{t-1}, y_{t-2}, \dots\}$. In this case, there are two MA representations of the same time series, one of which involves the fundamental shock u_t , whereas the other involves the nonfundamental shock u_t^* . These two MA representations have the same first and second moments, but their MA operators $\Phi(L)$ and $\Phi^*(L)$ differ. If the coefficients of $\Phi^*(L)$ reflect the dynamic multipliers of shocks, the econometrician's effort to recover the true structural impulse responses from the prediction error u_t will be doomed. This chapter reviews how this problem of nonfundamental shocks u_t^* may arise, how it may be detected, and how it may be resolved or at least ameliorated.¹

Nonfundamental shocks have been the subject of a growing number of studies on structural macroeconomic modeling. Early theoretical work on nonfundamentalness dates back to Hansen and Sargent (1980, 1991). An application illustrating the potential importance of nonfundamental

¹ This chapter is based in part on Lütkepohl (2014).

representations is provided in Lippi and Reichlin (1993). Nonfundamental representations of stochastic processes are also discussed in Blanchard and Quah (1993) and Lippi and Reichlin (1994). More recent work on nonfundamental representations and shocks includes Giannone and Reichlin (2006), Fernández-Villaverde and Rubio-Ramírez (2006), Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007), Forni, Giannone, Lippi, and Reichlin (2009), and Canova and Hamidi Sahneh (2016). A review of that literature is provided by Alessi, Barigozzi, and Capasso (2011).

The main argument in favor of nonfundamental shocks being important in economic analysis is that the econometrician may not have all the information that economic agents have. In particular, expectations of households, firms, and policymakers may not be based on past information from the variables in the econometric model alone, but may also reflect additional information from outside this model. Hence, the prediction errors governing the DGP cannot be equated with the reduced-form innovations of the VAR model set up by the econometrician. As a result, the shocks of economic interest may be nonfundamental. For example, Leeper, Walker, and Yang (2013) consider a model of fiscal policy shocks in which agents are assumed to know about tax changes in advance. This feature gives rise to nonfundamental tax shocks that undermine conventional estimates of the tax multiplier.

In this chapter, we provide a review of this literature and make the case that some practical responses to possible nonfundamentalness may be more appealing than others. Two common responses to the nonfundamentalness problem are either to allow for noninvertible MA components or to view nonfundamentalness as an omitted-variables problem. In the latter case, additional variables must be added to the model to align the information set used by the econometrician with the agents' information set. For example, Ramey and Shapiro (1998), Chung and Leeper (2007), Romer and Romer (2010), and Ramey (2011) use this approach in the context of fiscal policy analysis (see Chapter 7). We make the case that the omitted variables problem encompasses the nonfundamentalness problem. Solving it will also solve the nonfundamentalness problem, whereas solving the nonfundamentalness problem without adding further variables will not solve all of the problems caused by omitted variables. We emphasize that using factor models or Bayesian techniques to allow for larger information sets, as has recently been proposed, may not suffice to resolve the omitted variables problem. In some cases an alternative solution is to augment the model by a judiciously chosen variable that encompasses the missing information (see, e.g., Leeper, Walker, and Yang 2013; Kilian and Murphy 2014).

In Section 17.2 we formally define fundamental and nonfundamental representations of stochastic processes and the corresponding shocks. In Section 17.3 we review the literature on fundamental and nonfundamental representations. We conclude in Section 17.4.

17.2 Fundamental and Nonfundamental Moving Average Representations

In this section, we focus on a purely nondeterministic process y_t because deterministic terms are not needed for the subsequent arguments. Moreover, we exclude processes with MA roots on the unit circle and cointegrated processes as well as other types of nonstationarities. Some of these features are important in practice, but they do not affect the main arguments in this chapter. Consider a K -dimensional stationary stochastic process y_t with MA representation

$$y_t = \Phi(L)u_t, \quad t \in \mathbb{Z}, \quad (17.2.1)$$

where $u_t \sim (0, \Sigma_u)$ is a zero-mean M -dimensional white noise process with time-invariant nonsingular covariance matrix Σ_u and uncorrelated u_t and u_s for $t \neq s$. By allowing the dimension of the white noise process to be different from the dimension of y_t we generalize the framework of previous chapters. The operator

$$\Phi(L) = \sum_{i=0}^{\infty} \Phi_i L^i \quad (17.2.2)$$

is a possibly infinite order $K \times M$ matrix power series in the lag operator L , with absolutely summable coefficient matrices Φ_i . This setup covers cases in which y_t admits VAR or VARMA representations.

What is Fundamentalness? The MA representation is fundamental if u_t is the one-step ahead prediction error associated with the optimal linear prediction of y_t based on lagged y_t . Our formal definition of fundamentalness and nonfundamentalness follows from Rozanov (1967) and Alessi, Barigozzi, and Capasso (2011). Suppose that all random variables of interest belong to the Hilbert space $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ based on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ as in Alessi, Barigozzi, and Capasso (2011), where Ω is the sample space, \mathcal{F} is the σ -algebra describing all possible events, and \mathbb{P} is the probability measure. The symbol \mathcal{H}_t^y denotes the subspace of $\mathcal{L}^2(\Omega, \mathcal{F}, \mathbb{P})$ spanned by $\{y_t, y_{t-1}, \dots\}$.

Definition. The process u_t is y_t -fundamental if $\mathcal{H}_t^y = \mathcal{H}_t^u$ for all $t \in \mathbb{Z}$. The process u_t is y_t -nonfundamental if $\mathcal{H}_t^y \subset \mathcal{H}_t^u$ and $\mathcal{H}_t^y \neq \mathcal{H}_t^u$ for any $t \in \mathbb{Z}$.

In other words, u_t is y_t -fundamental if it can be written as a function of y_t, y_{t-1}, \dots ; otherwise it is y_t -nonfundamental. Using results of Rozanov (1967), Forni, Giannone, Lippi, and Reichlin (2009) observe that u_t is y_t -fundamental if, for $M \leq K$, $\text{rank}(\Phi(z)) = M$ for all z in the complex unit circle. Thus, for $M = K$, if all roots of $\det(\Phi(z))$ are outside the complex unit circle, expression (17.2.1) is a fundamental representation that can be inverted to

obtain a VAR representation of possibly infinite order. The u_t in this case is the one-step ahead prediction error associated with the optimal linear prediction of y_t based on \mathcal{H}_{t-1}^y .

The Relationship Between the Fundamental and Nonfundamental Representations. If (17.2.1) is a fundamental MA representation, then there are also nonfundamental MA representations with the same first and second moments. For example, if u_t is a scalar zero-mean white noise process with variance σ_u^2 , the scalar MA(1) process $y_t = u_t + \phi u_{t-1}$ is y_t -fundamental if $|\phi| < 1$. There is an equivalent nonfundamental representation $y_t = u_t^* + \frac{1}{\phi} u_{t-1}^*$, where u_t^* is zero-mean white noise with variance $\sigma_{u^*}^2 = \phi^2 \sigma_u^2$. It is easy to verify that this representation implies the same autocovariance structure for y_t . Clearly, the MA polynomial $1 + \frac{1}{\phi}L$ has a root $-\phi$ inside the unit circle. Hence, u_t^* cannot be obtained from y_t and its lags by inverting the MA polynomial.

For more general K -dimensional fundamental processes y_t , we obtain

$$y_t = \Phi(L)\mathbb{B}(L)\mathbb{B}(L)^{-1}u_t = \Phi^*(L)u_t^*, \quad (17.2.3)$$

where $u_t^* = \mathbb{B}(L)^{-1}u_t$. The MA operator $\Phi^*(z)$ has roots inside the complex unit circle if $\mathbb{B}(L)$ is chosen to be a so-called Blaschke matrix (see Lippi and Reichlin 1994). A matrix operator $\mathbb{B}(z)$ is a Blaschke matrix if it has no poles in and on the complex unit circle and $\mathbb{B}(z)^{-1} = \mathbb{B}^*(\frac{1}{z})$, where the asterisk denotes the conjugate transpose. For example,

$$\begin{bmatrix} z - a & 0 \\ \bar{a}z & I_{K-1} \end{bmatrix}$$

with $|a| < 1$ is a Blaschke matrix. Here \bar{a} denotes the complex conjugate of a . An important property of Blaschke matrices is that $u_t^* = \mathbb{B}(L)^{-1}u_t$ is white noise if u_t is white noise.

In the nonfundamental representation (17.2.3) the white noise process u_t^* is not the optimal linear prediction error of y_t , given lagged y_t . Note, however, that nonfundamental MA representations are equivalent to the fundamental MA representation in that they have the same second-order properties and, hence, represent the same stochastic process y_t under Gaussianity. In a non-Gaussian setting, in contrast, it is possible to discriminate between fundamental and nonfundamental representations based on higher moments, as shown in Huang and Pawitan (2000), Chan and Ho (2004), and Gourieroux and Monfort (2014).

Nonfundamental representations are also related to noncausal VAR models. Noncausal VAR models include regressors y_{t+i} in the equation for y_t . Such models have nonfundamental shocks by construction (see, e.g., Brockwell and Davis (1987, Chapters 3 and 11) and Lanne and Saikkonen (2013) for discussions of noncausal processes). Lanne and Saikkonen (2013) propose statistical

procedures for discriminating between causal and noncausal processes in practice.²

17.3 Fundamental versus Nonfundamental Representations

MA representations are directly linked to impulse response analysis, which is the main tool used in economic analysis. The impulse responses based on nonfundamental MA representations may be quite different from those based on the fundamental MA representation. This fact explains why the question of fundamentalness is considered important in macroeconomics. It should also be noted that the nonfundamental MA representation may be quite different from the generalized impulse responses proposed by Koop, Pesaran, and Potter (1996) (see Chapter 18). The latter impulse responses are based on conditional expectations conditioning on past observations, whereas under nonfundamentalness such conditional expectations would be nonlinear functions that do not correspond to the coefficients of the MA lag polynomial.

17.3.1 Nonfundamental Shocks in Economic Models

The main argument in favor of nonfundamental representations is that the information set of economic agents may be larger than the information set available to the econometrician. In this case, agents' expectations differ from the conditional expectations obtained from the econometrician's model for y_t . Since the shocks underlying the fundamental MA representation are the econometrician's prediction errors, they will differ from the shocks of interest that would be obtained on the basis of the agents' larger information set.

Example: Fiscal Foresight. Leeper, Walker, and Yang (2013) consider a simple model for assessing the effects of fiscal policy shocks. The model implies that without fiscal foresight the deviation of the capital stock from its steady state value in quarter t , k_t , in equilibrium depends only on an iid technology shock, $w_t^{\text{technology}}$,

$$k_t = \rho k_{t-1} + w_t^{\text{technology}},$$

with capital accumulation being unaffected by iid shocks to lump-sum taxes w_t^{tax} . In contrast, with some degree of tax foresight, even serially uncorrelated tax hikes reduce capital accumulation. For example, if agents know the tax shock two quarters in advance,

$$k_t = \rho k_{t-1} + w_t^{\text{technology}} - \kappa(w_{t-1}^{\text{tax}} + \theta w_t^{\text{tax}}),$$

² The notion of causality in this literature is not the same as the standard notion of economic causality discussed in Chapter 7.

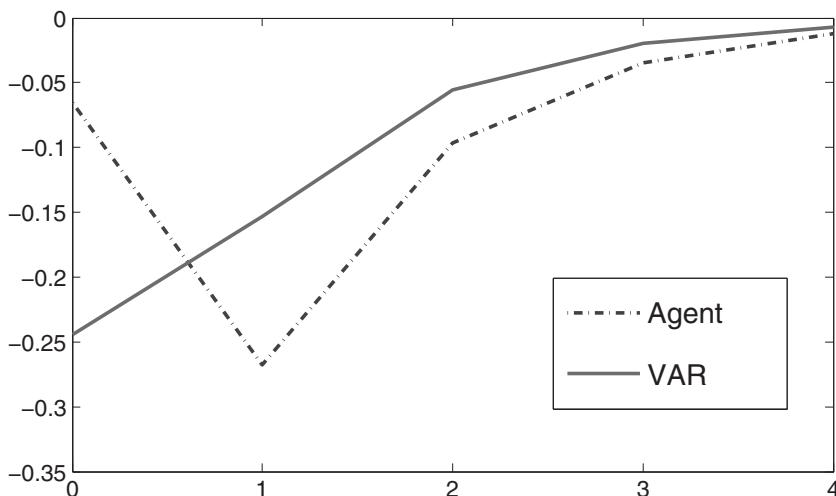


Figure 17.1. Response of the capital stock to an iid tax shock.

Source: Adapted from Leeper, Walker, and Yang (2013) with the help of MATLAB code kindly made available by Eric Leeper.

where $\kappa = (1 - \theta)(\tau/(1 - \tau))$, τ is the tax rate in steady state and $\theta < 1$ is a discount factor.

Leeper, Walker, and Yang (2013) prove that — abstracting from the technology shock — what the econometrician recovers as the tax innovation at time t by fitting an autoregression based on current and past observations for k_t is actually a discounted sum of the tax shocks observed by agents at date t or earlier. Thus, the tax shocks w_t^{tax} are not fundamental. Ignoring this fact results in distorted impulse responses. Figure 17.1 shows the difference between the response of the capital stock estimated by the econometrician and the true response implied by the model with fiscal foresight. Intuitively, this discrepancy arises because the econometrician attributes all of the dynamics associated with the anticipated component of the tax rate to the unanticipated component. The largest response to the tax shock occurs instantaneously in the econometrician's model. In contrast, in the model with tax foresight, the largest response of capital is seen to be one period after the arrival of the shock.

This simple example illustrates that, more generally, VAR models must account for forward-looking behavior to avoid spurious conclusions about structural impulse responses and forecast error variance decompositions. The discussion of fiscal policy in Leeper, Walker, and Yang (2013) is one prominent example of a situation in which economic foresight cannot be ignored. Economic foresight also plays an important role in modeling asset prices and commodity prices, for example. This realization has prompted concerns about the reliability of structural VAR analysis in many applications.

The central problem is that inverting the VAR lag operator by construction yields a fundamental MA representation, regardless of whether the DGP is fundamental or not. Two general approaches have been proposed to cope with the nonfundamentalness problem. One is to allow for MA representations in which the determinant has roots inside the complex unit circle when evaluating the VAR model, thereby explicitly accounting for the possibility of nonfundamental shocks. This approach requires suitable transformations of the model. The other approach is to broaden the information set by increasing the number of variables and, hence, the information included in the VAR model. The latter solution has been favoured by advocates of factor-augmented VAR (FAVAR) models in particular. As discussed in Chapter 16, FAVAR models can accommodate much larger sets of variables than conventional unrestricted VAR models, as can VAR models estimated by suitable Bayesian methods.

In Sections 17.3.2 and 17.3.3 we discuss these proposals in general, followed by specific comments in Section 17.3.4 on the use of FAVAR and large-scale BVAR models as a means of dealing with nonfundamentalness. In Section 17.3.5 we discuss other options for dealing with nonfundamental shocks that may be available in specific circumstances.

17.3.2 Nonfundamentalness Due to MA Roots in the Unit Circle

A fundamental MA representation can be turned into a nonfundamental one by applying the Blaschke matrices mentioned in Section 17.2 and, conversely, nonfundamental representations can be transformed into fundamental ones with equivalent second-order properties. In fact, if the process has a nonfundamental MA representation of the form (17.2.3) such that $\det(\Phi^*(z))$ has no roots on the unit circle, there always exists an equivalent fundamental MA representation. Thus, the question is not whether the true process is fundamental or nonfundamental (e.g., Alessi, Barigozzi, and Capasso 2011, p. 11). The process is always fundamental in the sense that it admits a fundamental representation as long as the process is Gaussian or only the first and second moments are of interest. Nevertheless, it may be the case that the shocks of interest for impulse response analysis are nonfundamental.

In a Gaussian setting, a VAR with fundamental errors can be fitted and used as the basis for determining the nonfundamental representation of interest as discussed in Section 17.2. For example, suppose that a univariate variable y_t admits an MA(1) representation,

$$y_t = u_t + \phi u_{t-1}, \quad (17.3.1)$$

where y_t -fundamentalness requires that $|\phi| < 1$. As mentioned in Section 17.2, an equivalent nonfundamental representation is $y_t = u_t^* + \frac{1}{\phi} u_{t-1}^*$. For the MA(1) process in (17.3.1) this is actually the only basic nonfundamental representation in the sense of Lippi and Reichlin (1994). Clearly, $|\frac{1}{\phi}| > 1$ and,

hence, the response of y_t to a nonfundamental shock after one period can be larger than 1, whereas the corresponding response to a fundamental shock is smaller than 1.

Whether it is useful or not to solve the nonfundamentalness problem by allowing for MA roots in the unit circle depends on whether the economic theory underlying the analysis requires nonfundamental representations. Obviously, allowing for nonfundamental representations in addition to fundamental ones potentially makes identification of the structural shocks of interest more difficult because it opens up more possibilities. If only fundamental shocks are considered, the structural shocks of interest are obtained by linearly transforming the residuals of the reduced-form VAR. If, however, in addition nonfundamental shocks are potentially of interest, all nonfundamental representations obtained by Blaschke matrix transformations plus all admissible instantaneous linear transformations must be considered when selecting a model representation, which may be a problem if economic theory does not deliver a full set of identifying restrictions. This is not a problem of the VAR or VARMA processes being the appropriate stochastic model for the DGP, but simply a problem of choosing the right representation for economic analysis.

It is worth mentioning that statistical tools can also help in determining whether a nonfundamental representation is required. As pointed out by Huang and Pawitan (2000), Davis and Song (2010), Lanne and Saikkonen (2013), and Gourieroux and Monfort (2014), nonfundamental and fundamental representations are equivalent for Gaussian processes whose distribution is determined completely by their first and second moments. For non-Gaussian processes fundamental and nonfundamental representations are not equivalent in that they give rise to processes that can be distinguished by their higher order moments. Hence, statistical procedures can help to discriminate between fundamental and nonfundamental representations. In particular, they may indicate that a nonfundamental representation is more suitable. Of course, they may also lead to the conclusion that a nonfundamental representation is not supported by the data.

17.3.3 Nonfundamentalness Due to Omitted Variables

More commonly, nonfundamental shocks are associated with an omitted-variables problem. If the econometrician considers a set of variables y_t , which does not contain all variables of importance to the economic agents, the VAR may still be a perfectly valid representation of the DGP of y_t , but due to omitted variables may not properly capture the structural impulse responses of economic interest. In other words, the system of interest for an economist may be $z_t = (y'_t, x'_t)'$, which contains the vector y_t of original VAR variables as well as a vector x_t of additional variables that are omitted from the original VAR

specification. Suppose

$$z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \Theta_{yy}(L) & \Theta_{yx}(L) \\ \Theta_{xy}(L) & \Theta_{xx}(L) \end{bmatrix} \begin{pmatrix} u_{yt} \\ u_{xt} \end{pmatrix} = \Theta(L)u_{zt} \quad (17.3.2)$$

is a fundamental MA representation of z_t . Here the partitioning of the matrix polynomial in the lag operator $\Theta(L)$ and the residual term u_{zt} corresponds to the partitioning of z_t . Clearly, $u_{zt} = (u'_{yt}, u'_{xt})'$ generally cannot be recovered from past and present y_t , i.e., \mathcal{H}_t^y is a strict subset of $\mathcal{H}_t^{u_z}$ in general and, hence, u_{zt} is not y_t -fundamental.

If a u_{yt} shock hits the system (17.3.2), the marginal responses of y_t are given by the appropriate elements of $\Theta_{yy}(L)$, which can be quite different from the impulse responses obtained from $\Phi(L)$, which denotes the MA operator in (17.2.1), for the simple reason that different shocks are considered. Of course, this insight is not new (see, e.g., Lütkepohl 1991, section 2.3.2; Giannone and Reichlin 2006). Note, however, that both MA representations (17.2.1) and (17.3.2) are fundamental, given their respective information sets: u_t in (17.2.1) is y_t -fundamental and u_{zt} in (17.3.2) (and hence, u_{yt}) is z_t -fundamental. Thus, the relevant MA operators have all roots outside the complex unit circle, which is a very different solution to the nonfundamentalness problem than the one discussed earlier.

Testing for Omitted Variables. Giannone and Reichlin (2006) make the point that the nonfundamentalness problem arises if any of the additional variables x_t is Granger causal for y_t (see Chapter 2). As is well known, if the MA representation is normalized such that $\Theta(0)$ is the identity matrix, x_t is Granger non-causal for y_t if and only if $\Theta_{yx}(L) = 0$ (see, e.g., Lütkepohl 1991, section 2.3.1, proposition 2.2). Thus, it is obvious that u_{yt} is y_t -fundamental, if $\Theta_{yx}(L) = 0$ and, hence, the MA operator in model (17.3.1) is lower block-triangular. In other words, if x_t is not Granger causal for y_t , u_{yt} can be recovered from present and past y_t . Conversely, if x_t is Granger causal for y_t , u_{yt} is not y_t -fundamental. A test along these lines has been conducted, for example, in Kilian and Murphy (2014) to address the question of whether the oil futures spread contains additional forward-looking information not already captured by the variables of a model of the physical market for crude oil. Canova and Hamidi Sahneh (2016) observe that Granger-noncausality tests may provide spurious evidence of nonfundamentalness under certain forms of model misspecification. They provide an alternative and more robust test of nonfundamentalness.

Why We Need More Information. To see that increasing the information set directly may be a better way of thinking about solving the fundamentalness problem than just allowing for MA roots in the unit circle, consider a simple illustrative example. Suppose y_t and x_t are both scalar variables that are jointly

generated by the bivariate MA(1)

$$\begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{pmatrix} u_{yt} \\ u_{xt} \end{pmatrix} + \begin{bmatrix} \theta_{yy} & \theta_{yx} \\ \theta_{xy} & \theta_{xx} \end{bmatrix} \begin{pmatrix} u_{yt-1} \\ u_{xt-1} \end{pmatrix}. \quad (17.3.3)$$

As usual, the MA coefficients represent the responses of the variables to u_{yt} and u_{xt} shocks. For instance, θ_{yy} is the response of y_t to a u_{yt} shock one period after it occurred.

The eigenvalues of the MA coefficient matrix in (17.3.3) are

$$\lambda_{1/2} = \frac{\theta_{yy} + \theta_{xx}}{2} \pm \sqrt{\left(\frac{\theta_{yy} + \theta_{xx}}{2}\right)^2 - (\theta_{yy}\theta_{xx} - \theta_{xy}\theta_{yx})}$$

and the process is fundamental if $|\lambda_1|, |\lambda_2| < 1$. Clearly, this condition is satisfied for a wide range of $\theta_{yy}, \theta_{yx}, \theta_{xy}$, and θ_{xx} values. This example illustrates the rich set of impulse responses that are possible when the additional information in x_t is taken into account. In particular, $\theta_{yy} > 1$ is possible and, hence, a u_{yt} shock can have an effect greater than 1 in the period following its occurrence. For example, for $\theta_{yy} = -\theta_{xx}$ it is easy to construct cases with $\theta_{yy} > 1$ and $|\lambda_1|, |\lambda_2| < 1$. Of course, the MA coefficients depend on the properties of x_t and y_t , but, given the large variety of potential x_t variables, it is difficult to exclude any feasible MA coefficient matrices a priori.

If instead of the bivariate joint process z_t only the first component y_t is considered, then the univariate marginal process is known to admit an MA(1) representation as in (17.3.1), with ϕ being a function of θ_{yy}, θ_{yx} , and the covariance matrix parameters of $(u_{yt}, u_{xt})'$ (see Sbrana and Silvestrini (2009) for the precise functional form). As mentioned earlier, in this case there is just one basic nonfundamental univariate MA(1) representation for the given marginal process y_t . In contrast, a continuum of alternative impulse responses is possible if further variables are added. As this discussion shows, a shock having an impact greater than one on y_t one period after its occurrence can also arise from adding further variables to the system under consideration.

Unfortunately, Granger noncausality of x_t for y_t does not solve the problem of distorted impulse responses due to omitted variables completely because there may also be an instantaneous causal relation, for example, if y_t and x_t are contemporaneously correlated. In that case, obtaining orthogonal shocks requires multiplying u_{zt} by some invertible matrix which may affect the impulse responses of y_t . Denoting by Σ_u the covariance matrix of u_{zt} and by I an identity matrix of suitable dimensions, let

$$\mathcal{Q} = \begin{bmatrix} Q_{yy} & Q_{yx} \\ Q_{xy} & Q_{xx} \end{bmatrix}$$

be some invertible matrix partitioned according to the partitioning of z_t such that $Q^{-1}\Sigma_{u_z}Q^{-1'} = I$. Then, if $\Theta(L)$ is lower block-triangular,

$$z_t = \begin{pmatrix} y_t \\ x_t \end{pmatrix} = \begin{bmatrix} \Theta_{yy}(L)Q_{yy} & \Theta_{yy}(L)Q_{yx} \\ \Theta_{xy}(L)Q_{yy} + \Theta_{xx}(L)Q_{xy} & \Theta_{xy}(L)Q_{yx} + \Theta_{xx}(L)Q_{xx} \end{bmatrix} w_t$$

is a possible structural fundamental MA representation with orthogonal shocks $w_t = Q^{-1}u_{zt}$. Hence, all shocks in w_t may have nonzero effects on y_t although x_t is not Granger causal for y_t . Thus, fundamentalness is necessary, but not sufficient for valid impulse response analysis, when important variables are omitted from the analysis. Even if x_t is not Granger causal for y_t , it may have to be considered in the impulse response analysis. This problem obviously cannot be solved by allowing for noninvertible MA representations. These considerations show that nonfundamentalness is only part of the problem caused by omitted variables.

Synopsis. The bottom line of this discussion is that there are two ways for overcoming the problem that the fundamental representation of the marginal process of y_t may not adequately reflect the responses to structural shocks. One possibility is to condition on the original set of variables y_t when deriving nonfundamental MA representations of y_t , while the second possibility is to increase the dimension of the process considered. The former possibility can be quite restrictive compared to the latter, because it does not account for instantaneous relations between included and omitted variables. When the basic argument for nonfundamentalness is that the econometrician has not accounted for all relevant information, it makes sense to add the missing information to the system to be considered. This is not to say that small-scale VAR models should never be used. Rather it highlights a potential drawback of fitting small-dimensional VAR models. Avoiding this drawback may call for greater care in how the VAR model is specified and identified. Sections 17.3.4 and 17.3.5 discuss several potential responses to this problem.

This discussion also has implications for the nature of the structural shocks identified by VAR models of the macroeconomy. In structural VAR analysis typical shocks of interest are demand, supply, monetary, fiscal, or technology shocks. Suppose the technology shock is of interest. Clearly in a small-scale VAR model real variables such as output and consumption are not broken down by sector, so an aggregate technology shock can only be thought of as an average technology shock and the implied impulse responses are the responses to such an average shock. In practice, it is rare for a technology shock to affect all sectors of the economy evenly, however; rather a technological innovation occurs in one sector and from there is transmitted to other sectors, eventually affecting the whole economy. Tracing the transmission from one sector to the

system may not only paint a different picture of the model dynamics than tracing an average shock, but it also requires sectoral variables to be contained in the system under consideration. Clearly, the correct transmission mechanism cannot be estimated by simply considering both fundamental and nonfundamental shocks in the original VAR model. It requires a larger-dimensional VAR model.

17.3.4 Avoiding Nonfundamentalness by Using Factor-Augmented or Large Bayesian VARs

Avoiding nonfundamental representations in structural impulse response analysis requires the use of VAR models that include a sufficient number of variables. Conventional VAR models include only a small number of variables because otherwise the estimation and specification uncertainty increases so much that the models become uninformative with respect to the relations between the variables involved. This does not mean that there is not a much larger number of variables that one could argue should be included in the model on economic grounds.

How Many Variables Do We Need? An important question is how many variables are required for a structural VAR model to include all the information required to answer the questions of interest. Forni and Gambetti (2014) suggest a formal statistical procedure that allows one to verify whether a given VAR model contains a suitable number of variables. Their proposal is to compute a possibly large number of principal components from the database of potentially relevant economic variables and to test whether these principal components are Granger causal for the smaller set of variables already contained in the VAR model. If none of the principal components is Granger causal for the variables in a given VAR, they conclude that the information in the VAR is sufficient for structural analysis. Otherwise further variables have to be added to the VAR. These variables can be observables or principal components, as would be the case in a FAVAR model. An important limitation of this approach is that it assumes that the factors fully describe the economic agents' information set.

An alternative test of the fundamentalness of a stochastic process has been proposed by Chen, Choi, and Escanciano (2017). Chen et al. focus on VAR processes with non-Gaussian iid structural errors. They prove, under some regularity conditions, that the innovations of the Wold MA representation of the VAR process are martingale difference sequences (mds), if and only if the structural VAR shocks are fundamental. Chen et al.'s test of the mds property of the Wold innovations has an asymptotic $\mathcal{N}(0, 1)$ distribution under the null

hypothesis. If the mds null hypothesis is rejected, the test may be repeated after augmenting the original VAR model by additional variables. It is shown by example that adding external estimates of fiscal news to a VAR of fiscal policy shocks may help overcome nonfundamentalness.

Why Large-Scale Models May Not Be the Answer. A common response to the problem of informational deficiencies has been the use of FAVAR models. The imposition of shrinkage restrictions on larger-dimensional VAR models has been another. A third proposal, associated with Forni, Giannone, Lippi, and Reichlin (2009), is to adapt the class of dynamic factor models to allow the construction of structural impulse responses (see Chapter 16).

These alternative models, however, have their own limitations. For example, the structural impulse responses generated by any of these three classes of models are only approximately valid, and the approximation may be poor. In particular, whether factors summarize the information in a large set of variables more accurately than simple aggregation of sectoral information, for example, is not at all clear. Neither form of information aggregation takes into account the economic relations between the disaggregates.

The same type of problem is encountered in large-scale Bayesian VARs (see Chapter 16). For example, Ba  bura, Giannone, and Reichlin (2010) use the classical Litterman or Minnesota prior that was derived with the model's forecast accuracy in mind (see also Chapter 5). Priors in large-scale Bayesian VARs serve to compensate for degrees-of-freedom limitations. Hence, they are bound to distort the parameter estimates and also the estimated impulse responses. Caution is called for in interpreting these estimates because unlike in small-dimensional VAR models we cannot evaluate the effect of the prior on the estimates against the benchmark of an unrestricted model.

An even bigger concern than the econometric challenges of working with large-dimensional VAR models is that the information that the econometrician is lacking more often than not is not included in standard macroeconomic databases. In this case, FAVAR models or large-scale BVAR models will not be able to overcome the informational deficiencies of the VAR model. For example, information about the credibility of promises of fiscal reform cannot be captured by simply adding a broad range of macroeconomic indicators to the VAR model. Moreover, to the extent that there are any data at all related to agents' expectations about future fiscal policies from opinion polls or prediction markets, these unconventional data sets tend to be too short for inclusion in a VAR model.

This point is even more apparent when considering the example of a structural VAR model of the global oil market. Expectations about future oil supply disruptions in the Middle East may depend on one's view of the stability of the Saudi monarchy, the resilience of Kurdish fighters to ISIS, the

willingness of the Iranian government to compromise on its nuclear ambitions, and numerous other determinants that are inherently unobservable, but nevertheless potentially important for the forward-looking component of the real price of oil. Such expectations cannot possibly be captured by including additional macroeconomic variables in the VAR model.

A more promising approach therefore may be to determine a small number of judiciously chosen additional variables that capture the most important missing information from an economic point of view. This alternative approach is discussed in the next section based on two concrete examples from the literature. We also consider a third example that incorporates forward-looking behavior directly into the structural VAR model. Finally, the reader is referred to Chapter 15 for alternative approaches to augmenting VAR models with external information obtained from survey data or other proxies for non-model-based expectations.

17.3.5 Other Approaches to Dealing with Anticipation

One option for dealing with the informational deficiencies of VAR models is to appeal to the ability of asset markets to aggregate all available information. If asset markets are efficient, asset prices will incorporate all the information available to agents and adding asset prices to the VAR model should help align the information set of the econometrician and the agent.

Example 1: A VAR Model of Anticipated Fiscal Policy. With respect to fiscal foresight, this approach was exploited by Leeper, Walker, and Yang (2013). In the United States, municipal bonds are exempt from federal taxes. The differential tax treatment of municipal and treasury bonds may be used to identify news about tax changes. All else equal, the municipal bond yield spreads allow one to infer the implicit tax rate at which financial investors are indifferent between tax-exempt bonds and taxable bonds. Because bond traders are forward looking, the implicit tax rate predicts future individual income tax rates. If there is news that individual tax rates are expected to increase, for example, investors will drive up the yield on taxable bonds until equilibrium is restored. By comparing the implicit tax rates at different maturities, we may infer shifts in expectations about future taxes. As stressed by Leeper, Walker, and Yang (2013), this approach has several advantages in practice. First, there is no need to specify a priori the period of foresight or for that period to be constant over time. Second, there is no need to specify the functional form of the information flow in the economy. Third, conditioning on the implicit tax rate resolves the nonuniqueness of the MA representation.

Leeper, Walker, and Yang (2013) report that the implicit tax rate variable Granger causes the variables in the fiscal policy VAR model of Blanchard and

Perotti (2002) that we already discussed in Chapter 8, indicating that the latter model is informationally deficient. How important the informational deficiency is, can be shown by revisiting the impulse response analysis. Leeper, Walker, and Yang (2013) propose augmenting Blanchard and Perotti's structural VAR model by the spread variable s_t such that

$$\begin{aligned} u_t^{ax} &= 2.08u_t^{gdp} + w_t^{tax} \\ u_t^{gov} &= c_{21}w_t^{tax} + w_t^{gov} \\ u_t^{gdp} &= -b_{31,0}u_t^{tax} - b_{32,0}u_t^{gov} + w_t^{gdp} \\ u_t^s &= -b_{41,0}u_t^{tax} - b_{42,0}u_t^{gov} - b_{43,0}u_t^{gdp} + w_t^s \end{aligned}$$

They add the identifying assumption that news contained in the interest rate spread, w_t^s , has no direct effect on current output (u_t^{gdp}), tax revenue (u_t^{ax}), and government spending (u_t^{gov}). The resulting structural VAR model can be used to construct responses both to unanticipated and anticipated tax revenue shocks. Leeper et al. show that their model produces markedly different impulse response estimates from Blanchard and Perotti's model in that anticipated tax increases raise output substantially for about three years, before output begins to decline. There is evidence that agents' foresight extends as far as five years.

As acknowledged by Leeper, Walker, and Yang (2013), there are obvious limitations to using municipal bonds as a measure of anticipated tax changes. First, this approach relies on other factors that affect municipal bonds such as default risk and liquidity risk being negligible. Second, the marginal investor in this market may not be representative of the tax payer. Third, municipal bonds respond to changes in individual income taxes only. Finally, this approach appears specific to the United States. Nevertheless, this example illustrates that there are creative approaches to overcoming informational deficiencies of standard VAR models.

Example 2: A VAR Model of Anticipated Oil Supply Shortfalls. Our second example is the model of the global market for crude oil since 1973 proposed in Kilian and Murphy (2014) (see Chapter 13). Essentially the same model has also been employed by Kilian and Lee (2014). The starting point of Kilian and Murphy's analysis is the observation that there clearly is an important forward-looking component in the real price of oil that cannot be captured by the variables traditionally included in structural oil market models. To resolve this informational deficiency they propose the inclusion of the change in global crude oil inventories as an additional variable in the VAR model.

Their approach is more subtle than that in Leeper, Walker, and Yang (2013) in that Kilian and Murphy (2014) make no attempt to measure expectations directly. Oil price expectations reflect the expected shortfall of future oil supply relative to future oil demand. Kilian and Murphy (2014) exploit the fact that

shifts in expectations not already captured by conventional oil market models affect the price of oil in the physical oil market by shifting the demand for crude oil stocks or inventories. To the extent that we can isolate the component of the change in crude oil inventories driven by such demand shifts from the remaining components associated with more conventional oil supply and oil demand shocks, we can infer shocks to expectations without having to quantify these expectations. Kilian and Murphy (2014) propose a set of identifying assumptions motivated by economic theory that is designed to disentangle these shocks and their effect on all model variables simultaneously. This approach explicitly deals with the problem that many of the determinants of oil price expectations are unobservable.

Thus, including the change in crude oil inventories in the model resolves the informational deficiency of the VAR model without having to model expectations. It should be noted that under perfect arbitrage exactly the same information conveyed by the change in inventories will also be contained in the spread of the oil futures price over the spot price of oil (see Alquist and Kilian 2010). Thus, including the oil futures spread in principle would be an alternative way of resolving the model's informational deficiency. In practice, the use of inventory data is preferred (1) because liquid oil futures markets only exist for a small part of the sample, (2) because traders in oil futures markets may not be representative for traders in the physical market for oil, and (3) because there is no need to take a stand on the extent of arbitrage between the physical and the financial market for crude oil when using oil inventory data. Moreover, possible concerns over the quality of the inventory data can be dealt with by comparing the fit of the model against extraneous evidence and by examining the robustness to alternative definitions of the inventory variable (see Kilian and Lee 2014).

It should be noted that including both the oil futures spread and the change in oil inventories in the structural VAR model is not an option. This would induce a singularity in the model to the extent that both variables capture the same information. Although it is conceivable that the oil futures spread may convey additional information not already captured by oil inventories, it can be shown that the oil futures spread has no additional predictive power for the variables included in the Kilian-Murphy model during the subsample for which the oil futures spread is available, suggesting that nothing is lost by excluding it from the model.

Kilian and Murphy (2014) demonstrate that explicitly modeling forward-looking behavior in oil markets makes a substantial difference for the interpretation of several episodes of rising or falling oil prices including 1979, 1986, 1991. Although their approach is specifically motivated by the economic structure of oil markets, it once again illustrates that there are creative solutions to the problem of nonfundamentalness that do not involve the use of large-scale VAR models.

Example 3: A VAR Model of Anticipated Technology Shocks. Not all structural VAR models dealing with forward-looking behavior rely on augmented information sets. A very different approach to modeling forward-looking behavior was proposed by Barsky and Sims (2011). This study focuses on expectations about aggregate technology specifically. They postulate that the log of aggregate technology, a_t , is characterized by a stochastic process driven by two structural shocks. The first shock is the traditional surprise technology shock, which impacts the level of productivity in the same period in which agents observe it. The second shock reflects information about future technology and is defined to be orthogonal to the first shock.³ The two structural shocks jointly account for all variation in a_t . They are identified as follows:

$$a_t = [B_{11}(L), B_{12}(L)] \begin{pmatrix} w_{1t} \\ w_{2t} \end{pmatrix},$$

where $B_{12}(0) = 0$ such that only w_{1t} affects current productivity instantaneously, making w_{2t} the future technology shock. Effectively, Barsky and Sims treat a_t as predetermined with respect to the rest of the economy. This identifying assumption leaves a wide range of possible choices for w_{2t} . In practice, w_{2t} is identified as the shock that best explains future movements in a_{t+1}, \dots, a_{t+H} , not accounted for by its own innovation, where H is some finite horizon.

The estimated VAR model includes a total factor productivity series as well as selected macroeconomic aggregates. The variable a_t is ordered first. The procedure is implemented by constructing candidate solutions of the form PQ , where P denotes the lower-triangular Cholesky decomposition of the reduced-form error covariance matrix Σ_u and Q denotes a conformable orthogonal matrix, as in the case of sign-identified VAR models. The ability of a shock to explain future movements of the data is measured in terms of the forecast-error variance decomposition. Because the contribution of the second shock to the forecast error variance of a_t depends only on the second column of B_0^{-1} , Barsky and Sims choose the second column, γ , to solve the optimization problem:

$$\gamma^* = \arg \max_{\gamma} \sum_{h=0}^H \omega_{12}(h),$$

subject to the first element of γ being zero and $\gamma'\gamma = 1$. Here $\omega_{ij}(h)$ denotes the share of the forecast error variance of variable i attributable to structural shock j at horizon h expressed in terms of the structural parameters of the

³ Barsky and Sims refer to this shock as a news shock, following a terminology common in the recent macroeconomic literature. This is somewhat misleading in that news shocks have traditionally been defined as unexpected changes to observed aggregates (see Chapter 7). Rather the second shock captures expected changes in future productivity.

model (see Chapter 4). The criterion function can be maximized, for example, by searching the space of possible subrotation matrices much like in the literature on sign-identified VAR models (see Chapter 13). One potential limitation of this solution method is that it generates the best possible case for expectations shocks rather than the most likely case. To the extent that the fraction of the forecast error variance of real GDP is small even in the best possible case, of course, this procedure will generate informative results.

Example 4: VAR Models with External Policy Shocks. A fourth example of how to deal with forward-looking expectations involves the use of external measures of policy shocks, as discussed in Chapter 15. For example, the Romer and Romer (2004) measure of monetary policy shocks is derived on the basis of internal forecasts of the Federal Reserve Board that reflect more information than can be captured by a VAR model. Likewise measures of exogenous policy shifts based on changes in interest rate futures prices around FOMC dates allow for richer information sets to be added to the VAR model.

17.4 Conclusions

One argument against using conventional structural VAR models for macroeconomic analysis is that they capture only fundamental shocks. These are shocks that can be recovered from past and present observed variables. In turn, shocks that are not recoverable from present and past observations are called nonfundamental. The fact that economic agents use additional information in decision-making that is not available to, or at least not used by the econometrician in specifying the VAR model, is a typical argument in favor of nonfundamental shocks. Two main approaches have been proposed to address the nonfundamentalness problem. First, in fundamental MA representations all roots of the determinantal polynomial obtained from the MA operator are outside the complex unit circle. Thus, allowing for roots inside the unit circle produces nonfundamental representations and shocks. The second response to the nonfundamentalness problem is to augment the information set by including more variables in the VAR model. Both approaches are plausible solutions to the nonfundamentalness problem, but they also have their limitations.

Allowing for MA roots inside the unit circle is plausible if the underlying economic model has this feature. This approach will not necessarily capture the true impulse responses, however, if the nonfundamentalness arises from omitted variables in the VAR model. Moreover, allowing for MA roots inside the unit circle complicates the identification of the structural shocks, unless economic theory provides a full set of identifying restrictions, which is rarely the case in practice.

Viewing nonfundamentalness as an omitted variables problem is attractive, given that nonfundamentalness typically is caused by the econometrician not using all relevant information in specifying the VAR model. Adding the missing information to the model is a natural response to the problem. There are two complications, however.

First, adding all information that may be important in agents' decision-making involves a large number of variables, resulting in the well-known curse of dimensionality. The latter problem can be tackled in different ways. One proposal is to use structural FAVAR models. We observed that, while these models are useful for integrating large panels of variables within the same VAR model, it is not straightforward to extract the structural impulse responses of interest from these models. Similar concerns apply to the use of large-scale Bayesian structural VAR models.

Second, the information relevant to the decisions of economic agents may not be captured by the variables contained in standard databases. Clearly, neither FAVAR models nor large-scale Bayesian VAR models are designed to address this second concern.

In short, each of the solutions to the problem of nonfundamentalness proposed in the literature is beset with its own limitations. In the end, any proposal for overcoming the curse of dimensionality, which results from including a large number of variables, can be criticized on the grounds that it may induce distortions in impulse responses. A more promising approach therefore may be to identify a small number of variables on economic grounds that reflect the relevant information for a particular problem at hand and including those in the model. Although it may not be easy to agree on the most important variables in general, in specific cases economic theory may provide some guidance. We discussed four examples in the empirical literature that propose creative solutions to the problem of nonfundamentalness.

One final point to keep in mind is made by Sims (2012). He shows by example that nonfundamentalness leads to distortions in standard VAR analysis, but that these distortions may be small in practice. For example, impulse response functions obtained from standard structural VAR models may be very similar to the true responses even if the true shocks are nonfundamental. A similar point is also made in Beaudry and Portier (2014). These authors provide a detailed discussion of the nonfundamentalness problem in the context of the news shocks defined in their work.

18 Nonlinear Structural VAR Models

18.1 Motivation

The standard VAR model, as discussed in previous chapters, is designed to capture the linear dependence of y_t on its own lags. This model is linear in the slope parameters as well as linear in the lagged model variables. More generally, however, the conditional mean may be nonlinear in the lagged variables and/or the model parameters.

Consider the class of VAR models where y_t depends nonlinearly on its lags. A K -dimensional nonlinear VAR process for y_t may be defined as

$$y_t = \mathbf{F}_t(y_{t-1}, \dots, y_{t-p}) + u_t, \quad (18.1.1)$$

where the white noise reduced-form innovations are additively separable and the nonlinear function $\mathbf{F}_t(\cdot)$ may depend on t . The appropriate form of the function $\mathbf{F}_t(\cdot)$ depends on the economic context.

If $\mathbf{F}_t(\cdot) = \mathbf{F}(\cdot)$, as long as the function $\mathbf{F}(\cdot)$ is well behaved, the nonlinear conditional mean specification may equivalently be expressed as

$$\begin{aligned} y_t = & v + \text{linear component} + \text{quadratic component} \\ & + \text{cubic component} + \dots + u_t, \end{aligned}$$

based on a Taylor-series expansion of the conditional mean about zero. For example, in a model including only one autoregressive lag, the quadratic component would involve all squares and pairwise products of the elements of y_{t-1} , and the cubic component would contain regressors of the form $y_{m,t-1}y_{n,t-1}y_{l,t-1}$, with m, n , and $l \in \{1, \dots, K\}$. In that case, the standard linear reduced-form VAR model

$$y_t = v + A_1 y_{t-1} + u_t,$$

may be viewed as a first-order linear approximation to this more general process.

Often the linear VAR approximation is quite accurate, but it is important to recognize that the fact that the approximating model is linear in the lagged variables and linear in the autoregressive slope parameters imposes several restrictions on the structural impulse response functions. First, the impulse responses increase proportionately with the magnitude of the structural shock. Second, responses are symmetric in positive and negative structural shocks. In other words, the response to a positive structural shock of given magnitude is the exact mirror image of the response to a negative structural shock of the same magnitude. Third, the responses to structural shocks are invariant to when the structural shock occurs and to the state of the economy, at the time when the structural shock occurs. There are situations in which these restrictions are clearly unrealistic, invalidating the use of linear approximations.

For example, in international trade transportation costs may prevent arbitrage between the prices of the same good in different countries expressed in the same currency. Arbitrage will occur only when the price differential exceeds the transportation cost. The existence of such thresholds implies that for small values of the price differential, there will be no response to price shocks, whereas for higher values there will be a disproportionate adjustment, invalidating the assumption of linearity. Similar zones of inaction arise in financial economics when trades involve a transaction cost. Other potential sources of thresholds include capacity constraints in production as well as ceilings and floors in modeling inventories. In addition, thresholds may arise from government regulation. Exchange rate target zones are a good example. It is possible to adapt the linear VAR model to allow for such threshold dynamics, resulting in a threshold VAR model.

Often the rationale for hard thresholds of this type is compelling at the microeconomic level, but not at the macroeconomic level. Aggregation across firms or households with different thresholds implies the existence of smooth thresholds (or smooth transitions) in the data. Thus, when working with aggregate data, it is common to postulate a smooth transition function that depends on how far the variable of interest is from its latent equilibrium value. Typical choices include an exponential or a logistic function.

Such smooth-transition models are also useful in modeling traders' uncertainty about the correct specification of their model as well as, more generally, heterogeneity in beliefs among traders. For example, even if a given trader were convinced that a dollar-euro exchange rate at parity is the equilibrium value, there is enough disagreement among traders and enough uncertainty about the equilibrium exchange rate that it would be unwise for that trader to take a position against an exchange rate that is only slightly higher than parity. As a result, arbitrage does not take place, and the exchange rate may evolve seemingly at random in the neighborhood of the latent equilibrium. A market consensus that the exchange rate is overvalued only emerges if the exchange rate reaches much higher values. Once there is a consensus that the exchange

rate is overvalued, there will be a disproportionate increase in traders anticipating a decline in the exchange rate and acting on these beliefs, ensuring that the exchange rate ultimately reverts back toward the latent equilibrium. These corrective forces, however, will weaken, as the exchange rate approaches the latent equilibrium. Smooth-transition VAR models are well suited to capturing such phenomena. It is also possible to adapt VAR models to allow for smooth one-time transitions following major discrete structural changes. For example, one may model the transition of inflation to a new steady state following the introduction of a new monetary system or a shift in the credibility of the central bank.

Another important type of nonlinearity involves the economy alternating between two or more states (or regimes), say the economy being in recession and being in expansion, with the VAR model parameters differing across regimes and with the transition between the regimes being governed by a stochastic process. Such regime-switching models allow the effect of structural shocks to differ across regimes. For example, fiscal policy shocks may have larger effects during a recession than during an expansion. This notion plays an important role in the debate about the magnitude of the fiscal multiplier in macroeconomics. More generally, even the sign of the response may be affected by the state of the system at the time of the shock.

It is also possible to allow measures of uncertainty to have a direct effect on the conditional mean of economic time series. Such a specification can be motivated by real options theory, which implies that uncertainty about the price of a commodity may delay investment decisions, if the cash flow of the investment depends on that commodity price. For example, it has been argued that shifts in the conditional variance of the percent change in the real price of oil lower U.S. real GDP growth. Such nonlinearities may be captured using GARCH-in-mean VAR models or VAR models with stochastic volatility in the error term.

Another potential source of nonlinearities is smooth structural change. Smooth structural change calls for the use of time-varying coefficient VAR (TVC-VAR) models that allow the VAR model coefficients to evolve continuously over time, according to a prespecified law of motion. Such models are designed for situations when the data are subject to many potential sources of nonlinearities of unknown form. For example, in modeling the global market for crude oil, changes in market structure, changes in energy efficiency and energy conservation, temporary capacity constraints, the emergence of unconventional oil production, delays in the response of new oil production to price incentives, gradual substitution among different forms of energy consumption, changes in refining and transportation infrastructure, and regulatory changes, all may render the VAR model coefficients unstable over time. If this instability is quantitatively important, a TVC-VAR model may be preferable to a linear VAR model.

Finally, the linear VAR model has been adapted to allow for responses that are asymmetric in positive and negative structural shocks. Such asymmetries are often modeled by censoring some of the model variables. This practice complicates the identification of the structural shocks and invalidates standard methods of estimation and inference based on linear VAR models, even when the model remains linear in the parameters. There are alternative model specifications, however, that remain valid, whether the responses in the DGP are symmetric or not.

As this overview shows, there are many types of nonlinear VAR models, each of which calls for a different specification of the function $F_t(\cdot)$. The common feature of all these models is that they allow current values of a set of variables to depend nonlinearly on lagged values of these variables. Referring to such models as nonlinear VAR models is conventional, but technically misleading because genuine VAR models are linear. We, nevertheless, follow this convention and broaden our definition of VAR models in this chapter to include nonlinear specifications.

The focus of the chapter is on the use of nonlinear VAR models for structural analysis. A detailed account of the specification and estimation of nonlinear time series models, including many of those relevant in the current context, is provided in Granger and Teräsvirta (1993) and Teräsvirta, Tjøstheim, and Granger (2010).

Section 18.2 outlines the general setup of nonlinear VAR models and reviews some of the challenges involved in conducting structural analysis in nonlinear models. In the subsequent sections a number of special cases of nonlinear structural VAR models are considered. We focus on threshold and smooth-transition VAR models (Section 18.3), Markov-switching VAR models (Section 18.4), time-varying coefficient VAR models (Section 18.5), and GARCH-in-mean VAR models (Section 18.6). A brief review of nonparametric and semiparametric VAR models is presented in Section 18.7. Some general comments on nonlinear VAR modeling are provided in Section 18.8. Finally, Section 18.9 focuses on models that are linear in the parameters, but are nonlinear in the variables.

18.2 Nonlinear VAR Analysis

18.2.1 General Setup

The standard linear VAR model has the reduced form

$$y_t = v + A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t,$$

where y_t is a K -dimensional vector of observed variables, v is a $K \times 1$ constant term, A_1, \dots, A_p are the $K \times K$ VAR slope coefficients, and u_t is an unobserved white noise error term. Thus, the current vector of observations, y_t ,

depends linearly on the past values of the variables under consideration. The corresponding structural form is

$$B_0 y_t = v^* + B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t,$$

where w_t is the structural error term with diagonal covariance matrix. The structural form allows us to explicitly model the instantaneous relations between the variables. The structural errors are linearly related to the variables. Their impact effects are given by the matrix B_0^{-1} .

This model may be generalized to allow for nonlinearities by considering structural processes of the form

$$\mathbf{G}_t(y_t, y_{t-1}, \dots, y_{t-p}) = w_t, \quad (18.2.1)$$

where the structural shocks w_t may have nonlinear impact effects on y_t , and $\mathbf{G}_t(\cdot)$ is a nonlinear function of current and lagged data that depends on t (see Mann and Wald 1943). The reduced form of model (18.2.1) may be obtained by expressing y_t as a nonlinear function of the lagged variables.

Much of the literature on nonlinear VAR models focuses on models with the somewhat simpler reduced-form representation (18.1.1),

$$y_t = \mathbf{F}_t(y_{t-1}, \dots, y_{t-p}) + u_t, \quad (18.2.2)$$

where the white noise reduced-form innovations are additively separable and the nonlinear function $\mathbf{F}_t(\cdot)$ may depend on t . Because the error term enters linearly, $u_t = B_0^{-1}w_t$ as in the linear model. There is, of course, no compelling economic reason why structural shocks in general should not be nonlinear functions of the reduced-form errors. In the latter case, one would have to estimate the structural model (18.2.1) directly rather than estimating the reduced form first and then recovering the structural model parameters from the reduced-form model.

It is important to recognize that the presence of nonlinearities in the DGP implies time variation in the coefficients of the linear VAR representation. Some of the nonlinear models considered in this chapter can be viewed as special cases of the time-varying coefficient model

$$y_t = v(t) + A_1(t)y_{t-1} + \cdots + A_p(t)y_{t-p} + u_t, \quad (18.2.3)$$

where u_t is white noise and where the dependence of the coefficients on t indicates that the slope coefficients $\alpha(t) = \text{vec}[A_1(t), \dots, A_p(t)]$ evolve according to a linear VAR(1) process

$$\alpha(t) = \Gamma\alpha(t-1) + \eta_t, \quad (18.2.4)$$

with Γ denoting a $pK^2 \times pK^2$ transition matrix consisting of time-invariant parameters and the error η_t following a pK^2 -dimensional white noise process that is independent of u_t . Such models are sometimes referred to as time-varying parameter VAR models in the literature. A more appropriate

terminology is time-varying coefficient VAR (TVC-VAR) models because parameters in frequentist analysis are time-invariant.

Time-varying coefficient VAR models can capture very general nonlinear dynamics. They can also accommodate discrete structural changes, provided the specification of model (18.1.1) is chosen appropriately. For example, if

$$\begin{aligned} & [v(t), A_1(t), \dots, A_p(t)] \\ &= \begin{cases} [v_1, A_{1,1}, \dots, A_{p,1}] & \text{for } t = 1, \dots, T_B, \\ [v_2, A_{1,2}, \dots, A_{p,2}] & \text{for } t = T_B + 1, \dots, T, \end{cases} \end{aligned}$$

the change in coefficients is best understood as a discrete structural shift in the model coefficients after time T_B . Clearly, this process is nonstationary. Likewise, time-varying coefficient VAR models can be used to capture smooth structural changes in the model coefficients by postulating that the elements of $\alpha(t)$ follow independent random walks, for example.

Alternatively, a time-varying coefficient VAR model of the form (18.1.1) may be used to capture genuine nonlinearities that arise in the absence of structural change. In that case, the DGP may in fact be stationary, allowing one to construct the MA representation of the data,

$$y_t = \mu_y + v_t + \sum_{i=1}^{\infty} \Phi_i v_{t-i}. \quad (18.2.5)$$

Here v_t is a white noise process, and v_t and v_s are uncorrelated, but not necessarily independent for $t \neq s$. The MA representation (18.2.5) is of limited use in practice, however, because it obscures the nonlinear relations between the observed variables. These features are hidden in the potentially complex dependence structure of v_t . As a result, the MA representation of stationary nonlinear models cannot be used for constructing statistics such as impulse responses and hence is irrelevant for the analysis in this chapter.

18.2.2 Structural Analysis

One of the drawbacks of working with nonlinear structural models is that in practice the analysis is restricted to impulse response analysis and to assessing the incremental effect of a sequence of structural shocks. In contrast, it is not obvious how to define forecast error variance decompositions in nonlinear structural VAR models nor is it clear how to construct historical decompositions. Our discussion in this section focuses on structural impulse response analysis. It is useful to first outline the general principles in the context of model (18.1.1). For expository purposes, it is assumed for now that B_0^{-1} is known, allowing us to infer the structural shocks from the reduced-form shocks.

The nonlinearity of the conditional mean specification complicates the construction of structural impulse responses. Because impulse responses depend on the sign and magnitude of the structural shock as well as on the history of the data, the standard approach to constructing structural impulse responses discussed in Chapter 4 cannot be used when working with nonlinear VAR models.

Structural impulse responses measure by how much future realizations of y_{t+h} for $h = 0, 1, \dots, H$ are expected to differ in response to a one-time structural shock at date t , compared with a baseline in which no such shock occurs. A natural definition of a nonlinear structural impulse response thus is as the difference between two conditional expectations of the realizations of y_{t+h} , $h = 0, 1, 2, \dots, H$, where the first expectation is conditional on the information set available at date $t - 1$, denoted Ω_{t-1} , as well as on the magnitude δ of the i^{th} structural shock, whereas the second expectation only conditions on Ω_{t-1} , but not on δ . This structural impulse response is a conditional impulse response in that its definition depends on the particular history Ω_{t-1} . More formally, we can define the conditional response function with respect to the i^{th} structural shock up to horizon H as the $K(H + 1)$ -dimensional vector

$$I_y(H, \delta, \Omega_{t-1}) \equiv \begin{pmatrix} \mathbb{E}(y_{t+0}|w_{it} = \delta, \Omega_{t-1}) - \mathbb{E}(y_{t+0}|\Omega_{t-1}) \\ \vdots \\ \mathbb{E}(y_{t+H}|w_{it} = \delta, \Omega_{t-1}) - \mathbb{E}(y_{t+H}|\Omega_{t-1}) \end{pmatrix}, \quad (18.2.6)$$

where Ω_{t-1} consists of the history of the model data up to time $t - 1$. How many lagged values of y_t are included in this history depends on the lag structure of the underlying model.

In practice, the expectations in question must be evaluated by Monte Carlo integration. Having estimated the K -dimensional nonlinear reduced-form model (18.1.1) and having recovered the sequence of structural shocks in the structural representation (18.2.1), the structural impulse responses may be computed by the following bootstrap procedure, if we are willing to impose the additional assumption that the structural shocks are mutually independent rather than merely mutually uncorrelated. Under the added assumption of Gaussian errors, these conditions would be equivalent.

Algorithm *Conditional impulse response function at date t*

1. The sequence of lagged data up to period $t - 1$ defines the history Ω_{t-1} at date t .
2. Consider a structural shock of magnitude δ at date t in one of the elements of w_t . Given Ω_{t-1} , we simulate two time paths of the realizations of y_{t+h} , $h = 0, \dots, H$. When generating the first time path,

the value of the structural shock of interest is set equal to the pre-specified value δ , denoting the magnitude of this shock. For example, we might set δ equal to 1. Subsequent realizations of this shock for period $t + h$, $h = 1, \dots, H$, are drawn from the marginal empirical distribution of the structural shock of interest. The realizations of the other $K - 1$ structural shocks for $h = 0, 1, \dots, H$ are all drawn independently from their respective marginal distributions. Given this K -variate sequence of structural shocks, we simulate the path of y_{t+h} , $h = 0, \dots, H$, by recursively updating the fitted model (18.1.1), conditional on Ω_{t-1} .

This simulated path of the data must be compared to the baseline path in which no structural shock of magnitude δ occurs in period t . When generating this second time path, all K structural shocks for $h = 0, \dots, H$ are drawn independently from their respective estimated marginal distributions. Given this alternative K -variate sequence of structural shocks, we again simulate the path of y_{t+h} , $h = 0, \dots, H$, by recursively updating the fitted model (18.1.1), conditional on Ω_{t-1} .

3. We now subtract the second time path for y_{t+h} , $h = 0, 1, \dots, H$, from the first time path. This difference is an estimate of the structural impulse response function conditional on Ω_{t-1} , but this estimate is noisy because it depends on random draws for the structural shocks, and hence is of little use in practice.
4. To eliminate the random variation in the impulse response estimate we repeat steps 2 and 3 many times and average the resulting impulse response estimates. By the law of large numbers, this average converges to the conditional response of y_{t+h} at horizon $h = 0, 1, \dots, H$ to a shock of magnitude δ conditional on Ω_{t-1} :

$$I_y(H, \delta, \Omega_{t-1}).$$

This conditional response function is of interest, if we are concerned with the response of the model variables at a particular point in time. An economist studying the performance of the economy over a longer time span based on a stationary nonlinear model, in contrast, may be interested in the unconditional impulse response function instead. The unconditional response of the variable y is defined as

$$I_y(H, \delta) = \int I_y(H, \delta, \Omega^r) d\Omega^r,$$

where Ω^r is a randomly selected history. In practice, $I_y(H, \delta)$ is simply obtained by averaging the value of the conditional response function $I_y(H, \delta, \Omega^r)$ over many histories, Ω^r , each of which is randomly drawn with replacement from the original data.

Alternatively, one could also condition on the state of the economy. For example, one might condition on the average of the subset of all histories in which the economy is in recession.

Regardless of whether we condition on a particular history or not, an important question is how to choose δ . One possibility is to report the range of nonlinear structural responses for a grid of δ values, keeping in mind that the magnitude of δ must not exceed the support of the structural residuals. It is also common to focus on one standard deviation and two standard deviations of the structural shock of interest, representing a shock of typical magnitude and an unusually large (but not unrealistically large) shock, respectively. Note that in comparing or averaging conditional impulse response functions across time it is important to keep constant the value of δ .

Impulse responses constructed in this manner are sometimes referred to as generalized impulse responses. Generalized impulse response functions were first proposed by Koop, Pesaran, and Potter (1996) who laid the foundations for nonlinear VAR impulse response analysis.¹ It is important to note, however, that Koop et al.'s analysis was not concerned with structural impulse response analysis, but with reduced-form impulse response analysis. Their focus was on variable-specific or system-wide reduced-form shocks drawn from the joint distribution of u_t . In contrast, our focus in this chapter is on characterizing nonlinear responses to innovations in w_t . Hence, the algorithm described above differs from Koop et al.'s original proposal.

There is some confusion in the literature about the precise sense in which Koop et al.'s analysis generalizes the traditional approach to impulse response analysis. As a result, the term generalized impulse response has been used to denote very different types of impulse responses in applied work.

Without doubt, the definition of the impulse response function as the difference between two conditional expectations is more general than the definition in Chapter 4 in that it can be applied to both linear and nonlinear VAR models. Moreover, it can be shown that this definition reduces to the standard definition of structural impulse responses when the VAR model is linear.

This is not the only sense in which definition (18.2.6) has been considered a generalization of the traditional approach, however. For example, Pesaran and Shin (1998) explicitly advocate the use of the generalized impulse responses proposed in Koop, Pesaran, and Potter (1996) as being more general than traditional structural VAR impulse responses even in linear models. They suggest that studying the response of the model variables to reduced-form shocks is more general because it avoids the use of economic identifying assumptions that may be controversial. This argument is not persuasive because

¹ A related approach is the conditional moment profile of Gallant, Rossi, and Tauchen (1993) who consider not only the conditional mean as in (18.2.6) but also higher conditional moments such as the conditional variance.

reduced-form shocks have no economic interpretation and violate the *ceteris paribus* assumption required for quantifying causal relationships in the data. At best, these responses may serve as a description of the properties of the data. They are not informative about the structural model. Moreover, if we were actually interested in responses to reduced-form shocks in linear VAR models, such responses could easily have been generated using the standard tools already discussed in Chapter 4.

In light of these conflicting interpretations, the terminology of generalized impulse responses should perhaps be avoided. Indeed, it is not uncommon in applied work for researchers not to be explicit about how their generalized responses are constructed or to give the misleading impression that structural responses may be obtained from reduced-form models without further identifying assumptions. When applying definition (18.2.6) or its generalizations to structural nonlinear VAR models in this chapter, we therefore simply refer to nonlinear structural impulse responses.

So far we have maintained the simplifying assumption that B_0^{-1} is known, allowing us to map the reduced-form innovations into structural shocks. In practice, B_0^{-1} is not known, of course, but has to be estimated from the data. Like in linear VAR models the challenge is how to identify the structural shocks. As long as we restrict attention to models of the form (18.1.1), this question may be addressed in much the same way as in linear models because the structural shocks are linear transformations of the reduced-form residuals. For example, we may impose short-run identifying restrictions as discussed in Chapter 8.

In contrast, the use of long-run exclusion restrictions as in Chapter 10 is problematic in the current context because there is no closed-form solution for the long-run impact of structural shocks in nonlinear VAR models. Sometimes in applied work based on TVC-VAR models or other nonlinear VAR models long-run identifying restrictions are imposed on the long-run response of the instantaneously linear model observed at date t . This practice is problematic because it ignores the expected change in the model coefficients.

Likewise, identification by sign restrictions as discussed in Chapter 13 is not straightforward in the presence of nonlinearities. Typically, researchers using nonlinear sign-identified VAR models report the posterior median response function, but this practice is problematic for the reasons discussed in Chapter 13. Alternative procedures for summarizing the posterior of the structural models, as proposed in Inoue and Kilian (2013), rely on the existence of a one-to-one closed-form mapping between the VAR model parameters and the set of structural impulse responses and hence do not apply in the nonlinear setting. This problem remains unresolved thus far.

Finally, in previous chapters we have seen that the statistical properties of the model can also be helpful in structural modeling (see Chapter 16). This approach may also be feasible in nonlinear models. Although such purely

statistical approaches cannot substitute for economic identifying assumptions, they may be helpful in assessing the consistency of conventional economic identifying restrictions with the data.

The following sections discuss several specific structural nonlinear VAR models and their economic applications.

18.3 Threshold and Smooth-Transition VAR Models

Threshold models allow the model coefficients to evolve from one regime to another when some model variable exceeds a prespecified threshold value. For example, the central bank may tighten monetary policy only if the inflation rate exceeds a certain level. In a smooth-transition model this transition from one regime to another is allowed to be gradual. The framework discussed in this section encompasses both of these models as special cases. Threshold and smooth-transition VAR models have been employed by Balke and Fomby (1997), Teräsvirta, Tjøstheim, and Granger (2010), Rothman, van Dijk, and Franses (2001), Camacho (2004), and Galvao and Marcellino (2014), among others. A recent survey of this literature is Hubrich and Teräsvirta (2013).

18.3.1 Model Setup

Consider the reduced-form model

$$y_t = \nu + \sum_{j=1}^p A_j y_{t-j} + G(x_t, \theta) \left(\nu^+ + \sum_{j=1}^p A_j^+ y_{t-j} \right) + u_t, \quad (18.3.1)$$

where ν^+ is a constant. The $K \times K$ matrix function $G(x_t, \theta)$ depends on the variable x_t and the parameter θ and determines when and to what extent there is a change in the model coefficients. If $G(x_t, \theta) = 0$, the model (18.3.1) is a standard linear reduced-form VAR model with intercept term ν , slope parameter matrices A_j , $j = 1, \dots, p$, and white noise error term u_t . If $G(x_t, \theta)$ is nonzero, the parameters ν^+ and A_j^+ , $j = 1, \dots, p$, affect the determination of y_t . The specific form of $G(x_t, \theta)$ determines the nature of the nonlinear behavior. The variable x_t may consist of lagged values of y_t, y_{t-d} , where d is known as the delay, or of additional variables not included in y_t . The precise choice depends on the economic context. A common choice in applied work is the squared deviation of some model variable from its long-run equilibrium value, as a measure of the deviation from equilibrium. Often this variable is lagged by more than one period and/or averaged over several periods. Another common choice is the cumulative change in a model variable over the d most recent periods.

For example, in an exponential smooth-transition VAR (EST-VAR) model the function $G(x_t, \theta)$ is a diagonal matrix with exponential transition functions

on the diagonal such that

$$G(x_t, \theta) = \begin{bmatrix} 1 - \exp[-\gamma(x_{1t} - c_1)^2] & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & 1 - \exp[-\gamma(x_{Kt} - c_K)^2] \end{bmatrix}, \quad (18.3.2)$$

where $\gamma > 0$, $x_t = (x_{1t}, \dots, x_{Kt})'$ and $\theta = (\gamma, c_1, \dots, c_K)'$. As long as the transition variable x_{kt} is equal to the constant c_k , $1 - \exp[-\gamma(x_{kt} - c_k)^2] = 0$. If the difference between x_{kt} and c_k is very large, $1 - \exp[-\gamma(x_{kt} - c_k)^2] \approx 1$. Thus, if all transition variables x_{kt} are equal to c_k until some period T_B and then increasingly deviate from c_k , the process (18.3.1) gradually transforms itself from a VAR with coefficients v, A_1, \dots, A_p to a VAR with coefficients $v + v^+, A_1 + A_1^+, \dots, A_p + A_p^+$, which is why this process is referred to as a smooth-transition model.

A possible shift variable is, for example,

$$x_{kt} = \begin{cases} c_k & \text{for } t < T_B, \\ t & \text{for } t \geq T_B, \end{cases}$$

where $1 < T_B < T$. The parameter γ determines the speed of the transition from one regime to the other. A very large value of γ makes for very fast changes from one regime to the other. In that case, the model is similar to the fixed threshold model discussed later.

Perhaps the most common specification involves an exponential transition function, $G(x_t, \theta) = (1 - \exp[-\gamma(x_t - c)^2])I_K$ (see, e.g., Rothman, van Dijk, and Franses 2001). More general versions of the model (18.3.1) may involve different γ parameters in different equations or additional transition functions. Moreover, the innovation variance may be allowed to change as discussed in Chapter 14. If integrated and cointegrated variables are involved, one may want to specify the ST-VAR model as a vector error correction model rather than a VAR in levels or in differences.

In contrast, if the function $G(x_t, \theta)$ has the form

$$G(x_t, \theta) = \mathbb{I}(x_t > c)I_K,$$

where x_t is a scalar variable, $\mathbb{I}(\cdot)$ is an indicator function, and c is a constant, then model (18.3.1) is called a threshold VAR (TVAR) model. The parameters change if the threshold variable x_t exceeds the value c . The value of the threshold may be known or unknown, depending on the economic context. Such models were proposed by Balke and Fomby (1997) and used by Galvao and Marcellino (2014). In threshold models it may make sense to consider a number of separate threshold values because different transmission mechanisms may be valid for very low, very large, and intermediate values of a threshold

variable. More generally, one could consider a TVAR model with M regimes

$$y_t = \begin{cases} v^{(1)} + \sum_{j=1}^p A_j^{(1)} y_{t-j} + u_t^{(1)} & \text{if } x_t \leq c_1, \\ v^{(2)} + \sum_{j=1}^p A_j^{(2)} y_{t-j} + u_t^{(2)} & \text{if } c_1 < x_t \leq c_2, \\ \vdots \\ v^{(M)} + \sum_{j=1}^p A_j^{(M)} y_{t-j} + u_t^{(M)} & \text{if } x_t > c_{M-1}, \end{cases}$$

where c_1, \dots, c_{M-1} are the (possibly unknown) threshold values.

The parameters of threshold VAR and smooth-transition VAR models can be estimated by nonlinear least squares (NLS) or by ML methods under standard assumptions. Estimation can be difficult if the models are large or if a specific regime does not appear very often in the sample. In addition, the ML estimator of the parameters in the transition function may not have a standard asymptotic distribution. A case in point is the ML estimator of the threshold parameter in a TVAR model (see Chan (1993) and Hansen (1997b, 2000) for discussions of asymptotic properties of ML estimators of univariate threshold AR models). Smooth-transition VAR models may also be estimated by Bayesian methods (see Gefang and Strachan 2010; Gefang 2012).

18.3.2 Example: A TVAR Model of U.S. Monetary Policy

Galvao and Marcellino (2014) construct a TVAR model for a three-dimensional system of U.S. economic variables based on quarterly data for 1960 to 2008. Let $y_t = (gdp_t, p_t, i_t)'$, where gdp_t is the log of real GDP, p_t is the log price index, and i_t is the federal funds rate. The model is partially identified. It relies on a recursive identification scheme in which the first shock has an instantaneous impact on all model variables, whereas the last shock has a nonzero impact effect only on the interest rate i_t and, hence, is interpreted as a monetary policy shock (see Chapter 8).

Galvao and Marcellino consider a TVAR model with three regimes of the form

$$y_t = \begin{cases} v^{(1)} + \sum_{j=1}^p A_j^{(1)} y_{t-j} + u_t^{(1)} & \text{if } x_t \leq c_1, \\ v^{(2)} + \sum_{j=1}^p A_j^{(2)} y_{t-j} + u_t^{(2)} & \text{if } c_1 < x_t \leq c_2, \\ v^{(3)} + \sum_{j=1}^p A_j^{(3)} y_{t-j} + u_t^{(3)} & \text{if } x_t > c_2, \end{cases}$$

where c_1 and c_2 are unknown threshold values. They postulate that $u_t^{(i)} \sim \mathcal{N}(0, \Sigma_u^{(i)})$, $i = 1, 2, 3$, is a normally distributed reduced-form error term whose variance-covariance matrix may depend on the regime.

The transition variable x_t is based on the deviation from a Taylor rule in period $t - 1$. More precisely, the transition variable is defined as

$$x_t = 1 + 1.5(p_{t-1} - p_{t-5}) + 0.5(gdp_{t-1} - gdp_{t-5}) - i_{t-1}$$

such that x_t depends on lagged year-on-year inflation ($p_t - p_{t-4}$) and output growth ($gdp_t - gdp_{t-4}$). The Taylor rule involves annual output growth and inflation rather than the level of output and prices because the transition variable has to be stationary. A trending transition variable eventually would drive the system permanently into the first or last regime, ruling out future regime changes. Note that the transition variable is positive when the actual interest rate is below the value predicted by the Taylor rule, but is negative when i_t is larger than the rate implied by the Taylor rule such that large values of the transition variable indicate a loose monetary policy, whereas small values of x_t indicate a tight policy.²

Galvao and Marcellino (2014) estimate the model parameters including the threshold values c_1 and c_2 by Gaussian ML. They use information criteria for choosing between different model specifications. The structural shocks are obtained by applying a lower-triangular Cholesky decomposition to the reduced-form residual covariance matrix in each regime. The nonlinear structural impulse responses are constructed as discussed in Section 18.2.2.

Confidence bands for the nonlinear structural impulse responses are constructed by applying a residual-based bootstrap method to the estimated nonlinear model. This bootstrap method is intended to take into account the sampling uncertainty about the threshold parameters. It is not clear, however, what the theoretical justification for this bootstrap method is. In general, little is known about the asymptotic validity of bootstrapping nonlinear VAR models and the bootstrap methods employed in applied work tend to be ad hoc.

The study of Galvao and Marcellino (2014) is one example of the use of threshold VAR models. Other applications of smooth-transition and threshold models include Weise (1999), Balke (2000), Atanasova (2003), and Dios Tena and Tremayne (2009).

18.4 Markov-Switching VAR Models

Changes in regimes may also be modeled by making them dependent on a discrete Markov process with, say, M states. This approach has several advantages over the nonlinear models considered so far. In each period the state of the process is determined endogenously and the specific state can change from period to period. The process can even be in-between two states in a particular period. Thereby very flexible changes in the VAR model coefficients can be accommodated.

The theory and practice of Markov-switching VAR (MS-VAR) models was laid out by Krolzig (1997) who generalized the univariate model proposed

² Galvao and Marcellino (2014) also consider a generalization of this baseline TVAR model that allows for additional structural changes during the sample period. Given that structural changes arising from shifts in the monetary policy regime are already accounted for in the baseline TVAR specification, these additional structural changes must arise from unrelated economic events.

by Hamilton (1989) for business cycle analysis. Subsequent refinements were developed by Sims, Waggoner, and Zha (2008), Sims and Zha (2006b), Rubio-Ramírez, Waggoner, and Zha (2005), and Hubrich, Waggoner, and Zha (2016). In this section we first present the general model framework and then discuss estimation, structural analysis, and applications.

18.4.1 Model Setup

Consider a reduced-form $\text{VAR}(p)$ model

$$y_t = v(s_t) + A_1(s_t)y_{t-1} + \cdots + A_p(s_t)y_{t-p} + u_t, \quad (18.4.1)$$

where $v(s_t)$ is a $K \times 1$ intercept vector, $A_i(s_t)$ is $K \times K$ for $i = 1, \dots, p$, and $u_t|s_t \sim \mathcal{N}(0, \Sigma_u(s_t))$ is white noise with the covariance matrix depending on the state of a Markov chain s_t . The conditional distribution of u_t , given s_t , is assumed to be Gaussian. This framework encompasses a very general class of nonnormal unconditional distributions for u_t and, hence, for y_t . More details on the properties of s_t are provided later in this chapter.

This framework includes as special cases processes in which some of the model coefficients do not vary with the state of the Markov process. For example, for a conditionally homoskedastic process the covariance matrices are time invariant and $\Sigma_u(1) = \cdots = \Sigma_u(M)$. Similarly, in Chapter 14 we considered the special case of an MS-VAR model in which only the reduced-form error covariance depends on the Markov chain. Krolzig (1997) provides a full classification of these and other special cases of this framework and discusses their properties in detail.

For our purposes it is worth noting that the reduced-form setup in (18.4.1) can be transformed to the structural form

$$B_0(s_t)y_t = v^*(s_t) + B_1(s_t)y_{t-1} + \cdots + B_p(s_t)y_{t-p} + w_t, \quad (18.4.2)$$

where $B_i(s_t) = B_0(s_t)A_i(s_t)$ for $i = 1, \dots, p$, are the structural-form parameters. The $K \times K$ matrix $B_0(s_t)$ may be normalized to have a unit diagonal such that each equation can be written with one of the variables on the left-hand side. If $B_0(s_t)$ is unrestricted, in contrast, the error process w_t may be standardized to have unit variances. In that case, we may postulate that $w_t \sim \mathcal{N}(0, I_K)$ without loss of generality, regardless of the state of s_t . In short, w_t may be specified as invariant to the Markov chain. Alternatively, $w_t|s_t \sim \mathcal{N}(0, \Sigma_w(s_t))$ may have a diagonal covariance matrix $\Sigma_w(s_t)$ that depends on the Markov process. Since it is not uncommon in structural VAR analysis to consider just-identified structural forms where the structural- and reduced-form parameters have a one-to-one relation, we focus the discussion on a reduced-form MS-VAR analysis from which the structural-form can then be obtained by a simple transformation. This class of models does not allow for overidentifying restrictions. The

latter case is discussed in detail in Sims, Waggoner, and Zha (2008) and Rubio-Ramírez, Waggoner, and Zha (2005).

Markov Chains. The model coefficients depend on the Markov chain process s_t . Generally, a process s_t that may take values $1, \dots, M$ is called a Markov chain, if the transition probabilities from one state to another depend on the most recent state only. In other words,

$$\begin{aligned} p_{ij} &\equiv \mathbb{P}(s_t = j | s_{t-1} = i) \\ &= \mathbb{P}(s_t = j | s_{t-1} = i, s_{t-2} = l, \dots), \quad i, j, l = 1, \dots, M. \end{aligned}$$

The transition probabilities are the parameters of the process and are summarized in the transition probability matrix

$$P \equiv \begin{bmatrix} p_{11} & \dots & p_{1M} \\ \vdots & \ddots & \vdots \\ p_{M1} & \dots & p_{MM} \end{bmatrix}.$$

Note that the transition probability matrix is sometimes written in transposed form (e.g., Hamilton 1994, chapter 22). All elements of P are probabilities and, hence, are between zero and one, $0 \leq p_{ij} \leq 1$. Boundary values of 0 and 1 imply additional restrictions. For example, if $p_{ij} = 0$, the j^{th} state cannot be reached from state i . More generally, if P is lower block-triangular,

$$P = \begin{bmatrix} P_{11} & 0 \\ P_{21} & P_{22} \end{bmatrix},$$

where P_{11} contains the transition probabilities of the first M_1 states and P_{22} is the corresponding matrix for the last $M - M_1$ states, then, once one of the first M_1 states is reached, there is no possibility to return to states $m = M_1 + 1, \dots, M$. In that case the Markov chain is called reducible. Moreover, $p_{ij} = 1$ implies that state j is an absorbing state to which the process moves with probability one from state i .

Clearly, $\sum_{j=1}^M p_{ij} = 1$. In other words, the rows of P sum to 1 such that

$$P\mathbf{j}_M = \mathbf{j}_M,$$

where \mathbf{j}_M denotes an $M \times 1$ vector of ones. Thus, P has an eigenvalue of 1 and \mathbf{j}_M is the corresponding eigenvector.

In general, the transition probability matrix P contains M^2 parameters that depend on $M^2 - M$ parameters due to the restriction that $\sum_{j=1}^M p_{ij} = 1$. It may depend on fewer parameters, for example, if some of the transition probabilities are restricted to zero.

As mentioned earlier, MS-VAR processes are very flexible models. The description of the process in expressions (18.4.1) and (18.4.2) may seem to suggest that the process in each period has to be in one of the M states. This

is not necessarily the case, however. More precisely, this DGP implies that in any given period the Markov process is in any one of the M states with a certain probability. In other words, the DGP allows for mixtures of states, which allows the process to describe far more general unconditional distributions. Even under the conditional normality assumption, for example, the unconditional distribution of y_t can have heavy tails.

18.4.2 Identification

Structural identification in MS-VAR models is discussed in detail in Rubio-Ramírez, Waggoner, and Zha (2005). Their study considers identification via short-run and long-run restrictions as well as sign restrictions. We have already seen in Chapter 14 how identification may be achieved if there is MS in the residual covariance only. Although exact identification restrictions can be stated in very general terms, in practice it is common to impose these restrictions on the coefficients of each regime separately. For example, Sims, Waggoner, and Zha (2008) consider a model for $y_t = (gdp_t, \pi_t, i_t)'$, where gdp_t is log real GDP, π_t is GDP deflator inflation, and i_t denotes the federal funds rate. They identify the shocks by imposing that $B_0(s_t)$ (and hence also $B_0(s_t)^{-1}$) are lower triangular in each state. The model is partially identified in that only the monetary policy shock is identified. The monetary policy shock is identified as an economic shock which affects the federal funds rate in the impact period, whereas it affects gdp_t and π_t only with a lag.

Rubio-Ramírez, Waggoner, and Zha (2005) also discuss the use of long-run identifying restrictions (see Chapter 10). They even provide conditions for identification if both short-run restrictions on the impact effects and long-run restrictions are considered jointly. These results, however, only pertain to imposing the long-run restrictions on the linear model for each state and do not account for the fact that the regime may change over time. Although conditioning on the economy remaining in the recession regime, for example, is common in applied work, conditioning on one regime in constructing the impulse responses is problematic because it amounts to changing the structure of the model, making the impulse response analysis subject to the Lucas Critique. Nor is it clear that such responses are relevant for understanding the response of the actual economy because agents in the real world will correctly perceive that the regime may change in the future. Once we allow for endogenous switches in the regime by computing nonlinear structural impulse responses, as discussed earlier, imposing long-run restrictions becomes impossible, because there is no longer a closed-form solution for the long-run structural responses.

Rubio-Ramírez et al. also discuss the use of sign restrictions on the impact multiplier matrix. Assuming that sign restrictions are imposed in each state, this approach poses no special difficulties. Let us assume that the covariance matrix of w_t is normalized to be an identity matrix. Then, for a given

identified set of structural parameters $[B_0(m), B_1(m), \dots, B_p(m)]$ we have to find all orthogonal matrices Q such that the impulse responses obtained from $Q[B_0(m), B_1(m), \dots, B_p(m)]$ satisfy the desired sign restrictions for $m = 1, \dots, M$ (see Chapter 13 for details). In contrast, the imposition of dynamic sign restrictions can only be done by simulation, once we allow for the regime to evolve over time. Either way, however, the challenge remains of how to properly evaluate the posterior distribution (or the sampling distribution) of the set of structural impulse responses.

18.4.3 Estimation

Maximum Likelihood Estimation. Using the conditional normality of u_t and w_t , the likelihood function corresponding to models (18.4.1) and (18.4.2) is

$$l = \prod_{t=1}^T \left(\sum_{m=1}^M \mathbb{P}(s_t = m | y_{t-1}, \dots, y_1) f(y_t | s_t = m, y_{t-1}, \dots, y_1) \right), \quad (18.4.3)$$

where

$$\begin{aligned} f(y_t | s_t = m, y_{t-1}, \dots, y_1) &= (2\pi)^{-K/2} [\det(\Sigma_u(m))]^{-1/2} \exp \left\{ -\frac{1}{2} u_t' \Sigma_u(m)^{-1} u_t \right\} \\ &= (2\pi)^{-K/2} |\det(B_0(m))| [\det(\Sigma_w(m))]^{-1/2} \\ &\quad \times \exp \left\{ -\frac{1}{2} w_t' \Sigma_w(m)^{-1} w_t \right\}. \end{aligned}$$

Clearly, in this case even the log-likelihood is nonlinear and the estimator must be solved for numerically. Krolzig (1997) discusses EM algorithms that simplify this optimization task for a range of important special cases. In general, ML estimation of MS-VAR models may be unreliable in small samples and may become infeasible for larger models. Moreover, inference is complicated by some of the parameters having nonstandard asymptotic distributions. This problem also undermines the reliability of frequentist model selection procedures. Moreover, little is known about the asymptotic validity of the bootstrap for MS-VAR models.

A common response in applied work to the convergence problems of the ML estimator has been the use of Bayesian estimation methods as developed in Sims, Waggoner, and Zha (2008), Sims and Zha (2006b), and Rubio-Ramírez, Waggoner, and Zha (2005). These methods are outlined in the next subsection.

Bayesian Estimation. Bayesian estimation involves setting up prior distributions for all the parameters of the model and deriving the posterior distributions

of interest by simulation. For a given state of the Markov process the prior of the VAR parameters is usually set up as in the linear model. A Minnesota prior or a sum-of-coefficients prior combined with an inverse Wishart prior are common choices. Although the same prior distribution may be used in each state, this approach would be difficult to reconcile with the maintained heterogeneity of the states in the MS-VAR model.

The transition probabilities are the parameters of the Markov process. Typically, one imposes independent Dirichlet priors for the rows of P . The Dirichlet distribution is a multivariate generalization of the Beta distribution. Its density is given by

$$f(x_1, \dots, x_M; b_1, \dots, b_M) = \frac{1}{B(b)} \prod_{i=1}^M x_i^{b_i-1}, \quad (18.4.4)$$

where $b_1, \dots, b_M > 0$ are the parameters of the distribution. The density is defined for all $0 \leq x_1, \dots, x_M \leq 1$ such that $\sum_{m=1}^M x_m = 1$ and hence satisfies the restrictions on the row elements of the transition probability matrix P . The quantity $B(b)$ is a normalizing constant that depends on the parameter vector $b = (b_1, \dots, b_M)'$. $B(b)$ is a multivariate version of the Beta function that can be represented in terms of Gamma functions as

$$B(b) = \frac{\prod_{i=1}^M \Gamma(b_i)}{\Gamma(\sum_{i=1}^M b_i)},$$

with $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$ denoting the usual Gamma function.

If all parameters of the Dirichlet distribution are 1, i.e., $b_1 = \dots = b_M = 1$, the Dirichlet distribution reduces to a multivariate uniform distribution on the unit interval $[0, 1]$ for each of the components. If little is known a priori about the actual transition probabilities, this specification may be useful in setting up the prior. In related work, Sims, Waggoner, and Zha (2008) propose setting $b_j = 1$ for the off-diagonal elements of P and using a prior parameter that reflects the persistence of the regime under consideration for the diagonal elements of P . Note that the diagonal elements of P are the probabilities of remaining in a given state. These probabilities are often rather close to 1, if the appearance of the states is not very erratic. Thus, one may want to choose the prior parameter corresponding to the i^{th} diagonal element, p_{ii} , of P such that the expected value of the marginal prior of p_{ii} is 0.9, say. Since for $b_j = 1$, $j \neq i$,

$$\mathbb{E}(x_i) = \frac{b_i}{\sum_{k=1}^K b_k} = \frac{b_i}{b_i + (M - 1)},$$

a value of 0.9 is obtained by choosing $b_i = 18$ for a system with $M = 3$ states.

More generally, for the i^{th} row of P the prior may be chosen as

$$f_{Pi}(p_{i1}, \dots, p_{iM}; 1, \dots, 1, b_{ii}, 1, \dots, 1) = \frac{1}{B(b)} p_{ii}^{b_{ii}-1}.$$

Thus, if the priors of the different rows are independent, the joint prior of all the rows of P is

$$\prod_{i=1}^M f_{Pi}(p_{i1}, \dots, p_{iM}; 1, \dots, 1, b_{ii}, 1, \dots, 1).$$

Assuming independence between the elements of P and the other model parameters, the overall prior is obtained by simply multiplying this function with the prior density of the other parameters.

Combining the prior density with the likelihood yields the posterior. Unfortunately the posterior will typically not reduce to a known distribution. Sims, Waggoner, and Zha (2008) break down the posterior into conditional posteriors, at least some of which may be of a known form. For the remaining distributions they recommend the use of a Metropolis-Hastings algorithm. They do point out some problems that need to be taken into account in simulating the posterior distribution. In addition to the usual problem of choosing the sign of the shocks properly, there is also the problem of labelling the states. In other words, it is important, which state is labelled as State 1, State 2, etc. The same labelling must be used for each draw of the posterior. Sims, Waggoner, and Zha (2008) propose a complete algorithm for simulating the posterior that is designed to work even when estimating larger structural MS-VAR models. For the case of a just-identified model, draws from the reduced-form posterior can be turned into draws for the structural parameters, resulting in computational simplifications.

18.4.4 Model Selection

For MS-VAR models the number of lags and the number of Markov states have to be chosen at the specification stage. In a frequentist setting, these two quantities can be determined with penalized likelihood criteria such as the AIC subject to the caveats discussed in Chapter 2. For the MS-VAR class of models this criterion can be written as

$$AIC = -2 \log l + 2n,$$

where $\log l$ denotes the maximum of the log-likelihood function and n is the number of free parameters. The free parameters include the reduced-form MS-VAR parameters for the different regimes as well as the unconstrained elements of the transition probability matrix.³ Similarly, other model selection criteria such as the SIC and the HQC may be used.

³ Note that the transition probabilities add to one so that in a 2-state model we only have two unrestricted transition probabilities, in a 3-state model we have six unrestricted transition probabilities, etc.

In this context it is important to note that understating the number of Markov states may result in a change in the lag order required for the proper representation of the DGP. Krolzig (1997, chapter 3) shows that certain MS-VAR processes have a VARMA representation without Markov switching. Thus, if VAR order selection is based on a VAR model without MS, misleading results are to be expected. In other words, the choice of the number of Markov states and VAR lag-order selection cannot be seen as separate problems. Ideally a search over the full range of combinations of all plausible VAR lag orders and MS states should be performed. Unfortunately, given the computational problems in optimizing the log-likelihood function, there are limitations to the number of models that can be compared.

Psaradakis and Spagnolo (2006) report simulation results for the performance of model selection criteria for jointly choosing the number of MS states and the number of autoregressive lags for univariate AR models with MS coefficients. They conclude that penalized likelihood criteria such as the AIC are useful for determining the quantities of interest if the sample sizes are relatively large. Their study examines sample sizes of 200 and 400. A closer look at their simulation results, however, reveals that even with 400 observations and fairly simple models, the correct combination of the number of Markov states and AR order is often not detected. A reasonable conjecture is that the selection problem becomes even more difficult in multivariate models.

Of course, one could also compare models based on their relative likelihood values. Unfortunately, such tests are problematic as well because there may be unidentified parameters under the null hypothesis if a model with a smaller number of MS states is tested against a model with more states. In that situation LR tests have nonstandard distributions (for the present case, see Garcia 1998; Carrasco, Hu, and Ploberger 2014).

Finally, from a Bayesian perspective, if all models are equally likely a priori, a natural criterion for model comparison would be the marginal likelihood, as discussed in Chapter 5. The marginal likelihoods can be computed based on the algorithm discussed in the previous subsection. It should be understood, however, that the computational challenges involved make it unattractive to compare large numbers of alternative models. Thus, in practice, the choice will tend to be between a very small set of competitors.

18.4.5 Example: An MS-VAR Model of U.S. Monetary Policy

Sims, Waggoner, and Zha (2008) study a three-dimensional model of the U.S. economy comprised of the log of real GDP (gdp_t), GDP deflator inflation (π_t), and the federal funds rate (i_t). The data are quarterly and cover the period 1959q1–2005q4. The shocks are identified by assuming $B_0(s_t)$ to be a lower-triangular matrix. Sims et al. are interested in comparing models with different

types of coefficient variation. Their models allow for possible variation in $B_0(s_t)$ as well as in the shock variances. The reduced-form error covariance matrix is decomposed as $B_0(s_t)^{-1} \Sigma_w(s_t) B_0'(s_t)^{-1}$, where $\Sigma_w(s_t)$ is a diagonal matrix and $B_0(s_t)$ has unit diagonal. Thus, both the impact effects of the shocks and the variances are allowed to vary across the regimes of the Markov process.

In addition, Sims, Waggoner, and Zha (2008) consider models in which the coefficient variation is restricted to subsets of the coefficients. In particular, they consider models in which only the variances vary, but the VAR slope coefficients are regime invariant, i.e., $\Sigma_w(s_t)$ varies with the Markov regimes, but $B_i(s_t) = B_i$, $i = 1, \dots, p$, and $B_0(s_t) = B_0$. They also consider models in which only the coefficients of the interest rate equation are varying, whereas the coefficients in the other equations remain regime invariant.

The priors are chosen along the lines discussed in Section 18.4.3. The specific choices of the prior parameters are not justified much. The models are compared via their marginal data density or the marginal likelihood. Sims et al. find that a model with four regimes and only the variances changing fits best under this criterion. This result suggests that the transmission of shocks has not changed, but the magnitude of the shocks has. Thus, changes in the conduct of monetary policy appear to have played a limited role during the sample period.

This result, of course, is conditional on the Markov-switching structure being an adequate representation of the data. Whether the behavior of the U.S. economy in general (and of U.S. monetary policy in particular) is as regular as implied by a Markov-switching model remains an open question, as does the sensitivity of these results to the choice of the prior. An alternative framework is the TVC-VAR model discussed in the next subsection.

Given the computational difficulties in estimating MS-VAR models, this class of models has not been widely used to date. Another example is Sims and Zha (2006b) who investigate the effects of changes in U.S. monetary policy based on a larger model including six variables. A generalization of the MS-VAR model that allows the transition probabilities to be time-varying has recently been proposed by Hubrich, Waggoner, and Zha (2016).

18.5 Time-Varying Coefficient VAR Models

There is a rapidly growing literature on time-varying coefficient VAR (TVC-VAR) models that allow for very flexible variation in the VAR coefficients, possibly including the coefficients of the error covariance matrix (e.g., Canova 1993; Canova and Gambetti 2009; Canova and Ciccarelli 2009; Cogley and Sargent 2001, 2005; Primiceri 2005; Koop and Korobilis 2009; Koop, Leon-Gonzalez, and Strachan 2009). In the next subsection, the framework developed by Primiceri (2005) is used to illustrate the problems and the potential advantages of this class of models. This framework has also served as the point of departure for many subsequent studies.

18.5.1 Model Setup

Consider the following general reduced-form process

$$y_t = v(t) + A_1(t)y_{t-1} + \cdots + A_p(t)y_{t-p} + u_t. \quad (18.5.1)$$

The model coefficients depend on t because they are allowed to vary over time. The error term u_t is assumed to be a zero-mean white noise process with time-varying covariance matrix, $u_t \sim (0, \Sigma_u(t))$. To facilitate structural analysis, the error covariance matrix is parameterized as

$$B_0(t)\Sigma_u(t)B_0'(t) = \Sigma_w(t) \quad \text{or} \quad \Sigma_u(t) = B_0(t)^{-1}\Sigma_w(t)B_0'(t)^{-1}, \quad (18.5.2)$$

where $\Sigma_w(t) = \text{diag}[\sigma_1^2(t), \dots, \sigma_K^2(t)]$ is a diagonal matrix with the variances of the structural errors on the main diagonal and $B_0(t)$ is restricted such that the structural coefficients in $B_0(t)$ and $\Sigma_w(t)$ can be recovered from equation (18.5.2). For example, $B_0(t)$ may be a lower-triangular matrix with unit main diagonal,

$$B_0(t) = \begin{bmatrix} 1 & 0 & \cdots & 0 \\ b_{21,0}(t) & 1 & & 0 \\ \vdots & \ddots & \ddots & \vdots \\ b_{K1,0}(t) & \cdots & b_{K,K-1,0}(t) & 1 \end{bmatrix}. \quad (18.5.3)$$

Restrictions on $B_0(t)$ may be used to uniquely identify the structural shocks $w_t = B_0(t)u_t$.

The structural form of the model can be written as

$$y_t = v(t) + A_1(t)y_{t-1} + \cdots + A_p(t)y_{t-p} + B_0(t)^{-1}w_t.$$

Suppose that the reduced-form VAR slope coefficients are collected in the vector $\alpha(t) = \text{vec}[v(t), A_1(t), \dots, A_p(t)]$, and the unrestricted elements of $B_0(t)$ are summarized in $\mathbf{b}(t)$, where $\mathbf{b}(t) = [b_{21,0}(t), b_{31,0}(t), b_{32,0}(t), \dots, b_{K1,0}(t), \dots, b_{K,K-1,0}(t)]'$ if $B_0(t)$ is lower triangular as in expression (18.5.3). In that case $\mathbf{b}(t)$ is the $\frac{1}{2}K(K-1)$ -dimensional vector of elements below the main diagonal of $B_0(t)$. This vector is arranged row-wise such that the parameters for the individual equations are grouped together. Finally, let $\sigma(t) = [\sigma_1(t), \dots, \sigma_K(t)]'$ be the vector of the standard deviations of the components of w_t .

The law of motion of the vectors of coefficients is assumed to be

$$\alpha(t) = \alpha(t-1) + \eta_t^\alpha, \quad (18.5.4)$$

$$\mathbf{b}(t) = \mathbf{b}(t-1) + \eta_t^b, \quad (18.5.5)$$

$$\log \sigma(t) = \log \sigma(t-1) + \eta_t^\sigma, \quad (18.5.6)$$

where the error terms η_t^α , η_t^b , and η_t^σ are white noise processes. In other words, the VAR coefficients change randomly in every period, but there is considerable persistence in their movement. The first two vectors, $\alpha(t)$ and $b(t)$, follow random walks, while the process driving the standard deviations, $\sigma(t)$, is a geometric random walk. In other words, this model allows for stochastic volatility. The error terms of the model equations are typically assumed to have a block-diagonal covariance matrix

$$\text{Cov} \begin{bmatrix} w_t \\ \eta_t^\alpha \\ \eta_t^b \\ \eta_t^\sigma \end{bmatrix} = \begin{bmatrix} \Sigma_w(t) & 0 & 0 & 0 \\ 0 & \Sigma_\alpha & 0 & 0 \\ 0 & 0 & \Sigma_b & 0 \\ 0 & 0 & 0 & \Sigma_\sigma \end{bmatrix}, \quad (18.5.7)$$

where Σ_α , Σ_b , and Σ_σ are the covariance matrices of η_t^α , η_t^b , and η_t^σ , respectively. Under the additional assumption of Gaussian errors, this structure implies independence of the individual error terms.

Defining $Z_{t-1} \equiv (1, y'_{t-1}, \dots, y'_{t-p})'$, equation (18.5.1) can be written as

$$y_t = (Z'_{t-1} \otimes I_K)\alpha(t) + u_t. \quad (18.5.8)$$

The two equations (18.5.8) and (18.5.4) are easily recognized as a special case of a state-space model with measurement equation (18.5.8) and transition or state equation (18.5.4). Such models are well known in the time series literature. Many results exist that simplify their analysis (e.g., Lütkepohl 2005, chapter 18; Anderson and Moore 1979; Aoki 1987; Hannan and Deistler 1988; Harvey 1989; Durbin and Koopman 2012). An important tool for estimating state-space models is the Kalman filter or Kalman smoother, which refers to a very efficient algorithm for computing conditional moments such as

$$y_{t|t-1} = \mathbb{E}(y_t | y_{t-1}, y_{t-2}, \dots, y_1) \quad \text{and} \\ \Sigma_y(t|t-1) = \text{Cov}(y_t | y_{t-1}, y_{t-2}, \dots, y_1)$$

recursively. These moments are useful for constructing estimation algorithms. We will return to the issue of estimation shortly.

The identification of structural shocks in TVC-VAR models is nontrivial in general. As Primiceri (2005) points out, there are many ways for shocks to enter the model (18.5.1). They may enter through the structural disturbances $w_t = B_0(t)u_t$ in the measurement equation or they may enter through any of the error terms in expressions (18.5.4)–(18.5.6). In the latter case it is difficult to identify the structural shocks. If w_t can be identified, conditional impulse responses can be compared over time. Obviously, in comparing the effects of structural shocks across time, the magnitude of these shocks may have to be standardized. The computation of the structural impulse responses can be based on the algorithm described in Section 18.2.2.

18.5.2 Estimation

Maximum Likelihood Estimation. If y_t is modeled as a Gaussian process, the log-likelihood function is

$$\begin{aligned} \log l = & -\frac{KT}{2} \log(2\pi) - \frac{1}{2} \sum_{t=1}^T \log(\det(\Sigma_y(t|t-1))) \\ & - \frac{1}{2} \sum_{t=1}^T (y_t - y_{t|t-1})' \Sigma_y(t|t-1)^{-1} (y_t - y_{t|t-1}). \end{aligned} \quad (18.5.9)$$

This likelihood function is easy to evaluate if the conditional moments $y_{t|t-1}$ and $\Sigma_y(t|t-1)$ are available. The Kalman filter provides a way of computing these expressions. Thus, for given parameter values, the numerical evaluation of the likelihood function is not difficult. Computing the maximum of the likelihood may still be a challenge, however, because the likelihood function is very nonlinear and the parameter vector can be high-dimensional. As a result, unrestricted ML estimation of TVC-VAR models tends to be problematic and typically Bayesian methods are used.

Bayesian Estimation. Primiceri (2005) describes an MCMC algorithm for the Bayesian estimation of this model. He imposes priors that are convenient for setting up such an algorithm. He also makes the simplifying assumption that Σ_b is block-diagonal with blocks corresponding to the equations of the system. In other words, for the lower-triangular $B_0(t)$ matrix in (18.5.3) the blocks have size $1 \times 1, 2 \times 2, \dots, (K-1) \times (K-1)$. Priors are required for the initial values of the coefficients and the covariance matrices of the error terms in (18.5.4)–(18.5.6). Primiceri (2005) postulates Gaussian priors for the initial conditions of $\alpha(t)$, $\mathbf{b}(t)$, and $\log \sigma(t)$ and inverse Wishart priors for the corresponding residual covariance matrices. Specifically, his priors are

$$\alpha(0) \sim \mathcal{N}(\alpha^*, V_\alpha), \quad (18.5.10)$$

$$\mathbf{b}(0) \sim \mathcal{N}(\mathbf{b}^*, V_b), \quad (18.5.11)$$

$$\log \sigma(0) \sim \mathcal{N}(\log \sigma^*, V_\sigma), \quad (18.5.12)$$

$$\Sigma_\alpha \sim \mathcal{IW}_{(pK^2+K)}(S_\alpha^\alpha, n_\alpha), \quad (18.5.13)$$

$$\Sigma_b \sim \mathcal{IW}_{K(K-1)/2}(S_b^b, n_b), \quad (18.5.14)$$

$$\Sigma_\sigma \sim \mathcal{IW}_K(S_\sigma^\sigma, n_\sigma). \quad (18.5.15)$$

The numerical values of the prior parameters depend on the application at hand. An example is discussed in the next subsection. As noted by Primiceri (2005), a drawback of this approach is that the prior depends on the ordering of the variables. If $B_0(t)$ is lower triangular, then the elements in $\mathbf{b}(t)$ are determined by the lower triangularity of $B_0(t)$.

The details of the simulation algorithm for the joint posterior of

$$\boldsymbol{\alpha}(1), \dots, \boldsymbol{\alpha}(T), \mathbf{b}(1), \dots, \mathbf{b}(T), \boldsymbol{\sigma}(1), \dots, \boldsymbol{\sigma}(T), \Sigma_{\alpha}, \Sigma_b, \Sigma_{\sigma}$$

are given in Primiceri (2005). His MCMC algorithm is based on a Gibbs sampler and proceeds in four steps:

1. Draw from the joint distribution of $\boldsymbol{\alpha}(1), \dots, \boldsymbol{\alpha}(T)$, given $\mathbf{b}(1), \dots, \mathbf{b}(T), \boldsymbol{\sigma}(1), \dots, \boldsymbol{\sigma}(T)$ and values of the hyperparameters $\Sigma_{\alpha}, \Sigma_b, \Sigma_{\sigma}$. This distribution is a product of normal distributions that can be sampled with standard algorithms.
2. Draw from the joint distribution of $\mathbf{b}(1), \dots, \mathbf{b}(T)$, given the other coefficients and hyperparameters. A similar argument as in step 1 implies that this distribution is also a product of normals.
3. Draw from the joint distribution of $\boldsymbol{\sigma}(1), \dots, \boldsymbol{\sigma}(T)$, given $\boldsymbol{\alpha}(1), \dots, \boldsymbol{\alpha}(T), \mathbf{b}(1), \dots, \mathbf{b}(T)$ and the hyperparameters. For this step sampling algorithms from the stochastic volatility literature may be used (e.g., Kim, Shephard, and Chib 1998).
4. Finally, draws from the conditional posteriors of $\Sigma_{\alpha}, \Sigma_b$, and Σ_{σ} may be obtained by standard methods for sampling from inverse Wishart distributions, utilizing the fact that the distributions are mutually independent.

Bayesian estimation of the TVC-VAR model considered in this subsection is also discussed in detail in Koop and Korobilis (2009) who propose a generalization of this model that allows for a slightly more general transition equation (18.5.4) (see, e.g., Canova 2007). This involves replacing expression (18.5.4) by

$$\boldsymbol{\alpha}(t) = \Pi \boldsymbol{\alpha}(t-1) + (I - \Pi) \bar{\boldsymbol{\alpha}} + \eta_t^{\alpha},$$

where $\bar{\boldsymbol{\alpha}}$ is a constant vector. The additional parameter matrix Π can be treated as unknown, it can be parametrized in some way, or it simply can be treated as known. For example, one could impose restrictions that reflect the Minnesota prior. Canova (2007) discusses the case in which $\Pi = cI$ for a scalar constant c . Obviously, when $c = 1$ we are back in the previously considered case with random walk dynamics for $\boldsymbol{\alpha}(t)$. When $c = 0$, $\boldsymbol{\alpha}(t)$ is white noise, fluctuating around a constant mean.

Clearly, for TVC-VAR models with a large number of variables and/or a large number of autoregressive lags, the posterior becomes high-dimensional, and these simulation methods quickly become tedious or computationally infeasible. Improved Bayesian estimation methods continue to be developed (see, e.g., Koop 2013b). Nevertheless, Bayesian estimation of TVC-VAR models at this point is only feasible for small-dimensional models with a small number of autoregressive lags, restricting the use of these models in applied work. For example, it is not feasible to estimate monthly TVC-VAR models with large K and/or p .

18.5.3 Example: A TVC-VAR Model of U.S. Monetary Policy

Primiceri's (2005) study provides an example of how to apply the TVC-VAR methodology. Primiceri specifies a model with only three variables: the inflation rate (π_t), the unemployment rate (ur_t), and a short-term interest rate (r_t). The first two variables represent the non-policy sector of the economy. The interest rate equation is interpreted as a monetary policy rule. The model is semistructural. Identification is based on a recursive ordering. The model is estimated using quarterly data for 1953q1–2001q3. The first 40 observations of the sample (1953q1–1962q4) are used as a training sample to determine the numerical values for some of the parameters in the prior based on the estimate of a constant coefficient VAR model. The actual TVC-VAR analysis is conducted only on the remaining observations. The VAR order is set to $p = 2$ to keep the model tractable.

The prior is specified as follows:

$$\begin{aligned}\boldsymbol{\alpha}(0) &\sim \mathcal{N}(\widehat{\boldsymbol{\alpha}}_{OLS}, 4 \cdot V(\widehat{\boldsymbol{\alpha}}_{OLS})), \\ \mathbf{b}(0) &\sim \mathcal{N}(\widehat{\mathbf{b}}_{OLS}, 4 \cdot V(\widehat{\mathbf{b}}_{OLS})), \\ \log \sigma(0) &\sim \mathcal{N}(\log \widehat{\sigma}_{OLS}, I_K), \\ \Sigma_\alpha &\sim \mathcal{IW}_{(pK^2+K)}(0.01^2 \cdot 40 \cdot V(\widehat{\boldsymbol{\alpha}}_{OLS}), 40), \\ \Sigma_\sigma &\sim \mathcal{IW}_K(0.01^2 \cdot 4 \cdot I_K, 4), \\ \Sigma_{b,11} &\sim \mathcal{IW}_1(0.02 \cdot V[\widehat{b}_{21,0}], 2), \\ \Sigma_{b,22} &\sim \mathcal{IW}_2(0.03 \cdot V[\widehat{b}_{31,0}, \widehat{b}_{32,0}], 3),\end{aligned}\tag{18.5.16}$$

where $\widehat{\boldsymbol{\alpha}}_{OLS}$, $\widehat{\mathbf{b}}_{OLS}$, and $\widehat{\sigma}_{OLS}$ are the LS estimates obtained from the training sample. $V(\widehat{\boldsymbol{\alpha}}_{OLS})$ is the corresponding estimator of the covariance matrix of $\widehat{\boldsymbol{\alpha}}_{OLS}$. $V[\widehat{b}_{21,0}]$ is the estimated variance of the LS estimator $\widehat{b}_{21,0}$ and $V[\widehat{b}_{31,0}, \widehat{b}_{32,0}]$ is the estimated covariance matrix of the LS estimators $(\widehat{b}_{31,0}, \widehat{b}_{32,0})$. Recall that

$$\Sigma_b = \begin{bmatrix} \Sigma_{b,11} & 0 \\ 0 & \Sigma_{b,22} \end{bmatrix}$$

is assumed to be block-diagonal. Hence, the joint distribution of Σ_b follows from that of the two diagonal blocks.

Clearly, the prior specification is arbitrary. Notably the multiplicative constants in the expressions in (18.5.16) are chosen subjectively. The choices are made for tractability and reflect the lack of prior knowledge by the analyst. As usual, the impact of the prior assumptions can be investigated by a sensitivity analysis that varies the prior parameters. Primiceri (2005) reports a range of such checks and concludes that the main results are robust to the choice of the prior.

Once the posterior simulations are available, they can be used to address a range of questions. For example, one may focus on the variability of the VAR coefficients or one may ask which model coefficients were particularly unstable over time. The evolution of $\sigma(t)$ may also help identify periods of high volatility in shocks. As in previous subsections, the identification of the structural shocks is complicated by the absence of closed-form solutions for long-run responses. Conditional and unconditional nonlinear structural responses may be computed, as discussed in Section 18.2.2.

Primiceri (2005) concludes that the differences in how monetary policy was conducted under different Fed chairmen (Burns, Volcker, and Greenspan) “were not large enough to have any relevant effect on the dynamics of inflation and unemployment” (p. 841). Of course, that conclusion hinges on the TVC-VAR model being correctly specified and, in particular, on the specification of the monetary policy rule being appropriate for the entire estimation sample. One concern is that this model does not explicitly model other important sources of time series variation in the U.S. economy; another concern is how shocks to monetary policy can be identified from this model during time periods when the interest rate was not the policy instrument used by the Federal Reserve (see Barsky and Kilian 2001; Kozicki and Tinsley 2009); a third concern is that this semistructural model of monetary policy suffers from all the shortcomings of such models already discussed in Chapter 8.

Readers interested in other examples of the TVC-VAR methodology can find a wide range of studies that employ alternative identification strategies. For example, Benati and Mumtaz (2007) analyze a small U.S. macro model with four variables using sign restrictions and find evidence that the improved macroeconomic development after the Volcker period is due to good luck as opposed to good policy. Another example is Baumeister and Peersman (2013) who use sign restrictions to identify oil demand and oil supply shocks in a TVC-VAR analysis for the crude oil market. A common problem in such studies is how to differentiate between genuine time variation in the parameters and overfitting. Another (not unrelated) challenge is how to interpret apparent evidence of time variation from an economic point of view.

18.6 VAR Models with GARCH-in-Mean

18.6.1 Model Setup

VAR models with GARCH-in-mean have been used in a number of economic contexts. For example, Elder (2004) uses this class of models to study the effects of inflation uncertainty on output growth in the U.S. The structural form of this model is

$$B_0 y_t = v^* + B_1 y_{t-1} + \cdots + B_p y_{t-p} + \Lambda \sigma_t + w_t, \quad (18.6.1)$$

where w_t is a vector of conditionally heteroskedastic structural errors such that $w_t | \Omega_{t-1} \sim \mathcal{N}(0, \Sigma_w(t|t-1))$ with $\Sigma_w(t|t-1)$ being a conditional covariance matrix with multivariate GARCH structure and σ_t denoting the vector of square roots of the diagonal elements of $\Sigma_w(t|t-1)$. The term $\Lambda\sigma_t$ is included in the conditional mean specification to capture the impact of shifts in uncertainty on the dynamics of the model observations y_t . Note that Λ may be replaced by a $K \times K$ lag polynomial $\Lambda(L)$ so that lags of conditional standard deviations of all the shocks can in principle appear in any of the VAR model equations.

Since the errors in structural model (18.6.1) are mutually uncorrelated, the multivariate GARCH structure reduces to a system of individual GARCH processes for each of the error terms. This fact allows us to impose a diagonal GARCH structure such that $\Sigma_w(t|t-1)$ is a diagonal matrix with diagonal elements

$$\sigma_{kt}^2 = c_k + \gamma_k w_{k,t-1}^2 + g_k \sigma_{k,t-1}^2, \quad k = 1, \dots, K. \quad (18.6.2)$$

Of course, higher-order GARCH models could also be considered, although this is rarely done in practice.

Whereas conditional heteroskedasticity in the form of GARCH errors poses no fundamental problems in otherwise linear VAR models, the presence of the conditional standard deviation in the conditional mean specification renders the conditional mean nonlinear and conventional structural impulse response analysis invalid. Conditional and unconditional nonlinear structural impulse responses may be computed by simulation, as discussed in Section 18.2.2. In related work, Elder (2003) proposes a closed-form solution of the conditional nonlinear structural response for the VAR-GARCH-in-mean model that allows structural shocks to affect the conditional expectation of y_{t+h} through the conditional variance (see also Elder 2004; Elder and Serletis 2010).

18.6.2 Estimation

VAR models with GARCH-in-mean are typically estimated by Gaussian ML methods. Apart from a constant term, the log-likelihood function is

$$\begin{aligned} \log l = & -T \log |\det(B_0)| \\ & -\frac{1}{2} \sum_{t=1}^T [\log(\det \Sigma_w(t|t-1)) + w_t' \Sigma_w(t|t-1)^{-1} w_t]. \end{aligned} \quad (18.6.3)$$

Because $w_t = B_0 y_t - v^* - B_1 y_{t-1} - \dots - B_p y_{t-p} - \Lambda\sigma_t$, the log likelihood is a highly nonlinear function of the parameters. As a result, ML estimation may be problematic for models of larger dimension, models with larger orders, and/or models with more complicated forms of conditional heteroskedasticity. VAR Models with GARCH-in-mean are usually assumed to satisfy the

conditions required for ML estimators to have standard asymptotic properties. In addition to the usual identification restrictions for specifying the structural shocks, w_t , this means that we also need to ensure the identification of the GARCH parameters which in turn may require further identification conditions if more complicated GARCH structures are allowed for (see, e.g., Lütkepohl 2005, chapter 16).

18.6.3 Example: The Effect of Oil Price Uncertainty on U.S. Real Output

This class of models is best illustrated in the context of a study by Elder and Serletis (2010) that investigated the impact of oil price volatility on U.S. output growth based on a VAR model with GARCH-in-mean. This study was a major improvement on earlier studies that incorrectly treated changes in the price of oil as strictly exogenous with respect to the U.S. economy.

Elder and Serletis' baseline model was bivariate and included real output growth and the percent change in the real price of oil. They used quarterly data for 1974q2–2008q1, fit a model with diagonal GARCH specification, and identified the shocks by letting B_0 be lower triangular. In other words, they treated the price of oil as predetermined with respect to U.S. real output growth, consistent with evidence in Kilian and Vega (2011). They found that increased oil price volatility has a negative effect on U.S. real output, as measured by the coefficient of the conditional standard deviation in the real output equation of the VAR model. Negative effects were also found for various other measures of output.

This evidence is economically important because it implies that output contracts sharply in response to positive oil price shocks, but fails to expand in response to negative oil price shocks. This asymmetry result is surprising because, as we will see later in this chapter, other less parametric studies looking for asymmetric responses to positive and negative oil price shocks have failed to find such evidence.

Elder and Serletis interpret their evidence as being supportive of the real options theory of Bernanke (1983), which their model was designed to capture. Real options theory states that firms will delay investment in response to increased uncertainty about the price of crude oil to the extent that the cash flow of investment projects depends on the price of crude oil. Reductions in investment in turn have a recessionary effect.

Elder and Serletis' interpretation has been challenged by a number of studies (see Kilian 2014). One concern has been that what matters for real options theory is the uncertainty about the price of oil at the horizons that matter for investment decisions. Most investment projects take at least one year to become operational. For a typical investment project (such as an auto maker deciding whether to build a new plant producing sports utility vehicles) the

relevant horizon for the cash flow will typically extend from one year after the investment project is approved to several years after the project has started operating.

For concreteness, suppose that the decision to invest is based on the cash flow within the first five years of the plant becoming operational. Given the fact that the conditional variance from monthly or quarterly GARCH models quickly converges to the time-invariant unconditional variance at longer prediction horizons, one would not expect the conditional variance between years 1 and 6 to change very much at all. Thus, there is little reason for such an investment project to be delayed in response to an increase in the GARCH estimate of the monthly conditional variance.

Moreover, the share of investments whose cash flow depends heavily on the price of crude oil is quite small outside the oil industry, making it even less likely that the effects on aggregate U.S. real output would be large. For these two reasons, there is every reason to believe that whatever effect Elder and Serletis estimated is unrelated to real options theory and may reflect model misspecification.

A second concern is that GARCH models are inherently backward-looking, which makes them potentially poor measures of uncertainty. For example, GARCH models imply a sharp increase in uncertainty after the sharp fall in the price of oil in 1986. This fall in the price of oil was caused by Saudi Arabia's decision to no longer restrict its oil production, reversing an earlier policy decision to curb Saudi oil production in an effort to prop up the price of crude oil (see Alquist, Kilian, and Vigfusson 2013). Most observers would have interpreted this policy reversal as a reduction in uncertainty rather than an increase in uncertainty, because this event revealed Saudi Arabia's and hence OPEC's inability to control the price of oil, which in turn alleviated fears about what OPEC might do in the future to raise the price of oil.

A third and related concern is that the GARCH-in-mean specification imposes the assumption that the conditional mean and the conditional variance of y_t are driven by one and the same structural shock. As observed by Jo (2014), a more credible specification would be a VAR model with stochastic volatility that allows the conditional variance to evolve independently from the conditional mean. This modification, for example, would allow oil price uncertainty to fall, as the price of oil dropped in 1986.

The challenge in implementing this alternative modeling approach in practice is to find an independent measure of oil price uncertainty. Based on a novel measure of short-term oil price uncertainty, Jo finds that the recessionary effects of increases in uncertainty are considerably smaller than reported by Elder and Serletis. Even her evidence does not decisively resolve this question, however. The fundamental challenge for the stochastic volatility approach is that the economically relevant measure of uncertainty remains medium-term uncertainty (defined as uncertainty at horizons between one year and five years,

for example), but measures of medium-term stochastic volatility are not likely to be exogenous, complicating the identification of the responses to uncertainty shifts.

A final concern is that the price of oil responds to actual and expected fluctuations in global real economic activity. As a result, oil price uncertainty also tends to reflect macroeconomic uncertainty which directly affects the cash flow of many investment projects. This fact may help explain the unexpectedly large effects of oil price uncertainty reported in this literature.

18.7 Other Nonlinear Models

In this section we review several alternative nonlinear models that may become useful for structural analysis in the future, although their use has been limited so far. We focus on two classes of models. The nonparametric models considered in Section 18.7.1 allow flexible approximations to very general nonlinear functions. They are designed for situations in which the specific nonlinear form of the model is unknown. Estimates of nonparametric models may also provide empirical support for the use of a specific nonlinear functional form in structural modeling. In contrast, the noncausal VAR models discussed in Section 18.7.2 are intended for situations in which economic agents have more information than the econometrician. If the predictions of the economic agents differ from those implied by the econometric model, this fact has to be taken into account in setting up the model (see also Chapter 17).

18.7.1 Nonparametric VAR Analysis

Prediction involves capturing the conditional means, conditional variances, or the conditional densities of a vector of time series, given its past realizations. These characteristics are also of interest in reduced-form modeling. If little is known about the DGP, it makes sense to explore very general potentially nonlinear functional forms. This situation calls for the use of nonparametric estimation methods. For example, a nonparametric analysis of the conditional mean would start from the model

$$y_t = \mu(y_{t-1}, y_{t-2}, \dots) + u_t, \quad (18.7.1)$$

where u_t is a martingale difference series with respect to $\{y_{t-1}, y_{t-2}, \dots\}$. In that case, $\mu(\cdot)$ represents the conditional expectation in period t , given past observations y_{t-1}, y_{t-2}, \dots . The conditional expectation is the minimum mean squared error (MSE) one-step ahead predictor for y_t . For a finite-order VAR(p) process

$$\mu(y_{t-1}, y_{t-2}, \dots) = v + A_1 y_{t-1} + \dots + A_p y_{t-p}.$$

In parametric time series analysis, the function $\mu(\cdot)$ is chosen from some parametric class. A specific model within that class is then obtained by specifying the finite-dimensional set of model parameters associated with that function. In contrast, nonparametric time series analysis allows $\mu(\cdot)$ to be from a more flexible class of functions. We approximate $\mu(\cdot)$ such that the approximation error declines with the sample size.

There are several techniques that can be used for this purpose. For example, we may construct local approximations of $\mu(\cdot)$ in the neighborhood of any given argument by decreasing the neighborhood (and thereby improving the approximation), as the sample size increases. In this approach the number of lagged values of y_t is usually fixed. In other words, $\mu(y_{t-1}, y_{t-2}, \dots)$ is replaced by $\mu(y_{t-1}, \dots, y_{t-p})$ for given p . Alternatively, we may construct global approximations that involve parametric functions $\mu_T(\cdot)$. The number of parameters in that function (and hence the flexibility of the function) increases with the sample size T . The sequence of functions $\mu_T(\cdot)$ is chosen such that it approaches $\mu(\cdot)$ asymptotically, given the metric chosen by the researcher. This fact allows one to increase the number of lagged y_t with the sample size T without having to assume the existence of a particular finite-order parametric VAR model.

If the conditional volatility is of interest in addition to the conditional mean, the framework in (18.7.1) must be replaced by

$$y_t = \mu(y_{t-1}, y_{t-2}, \dots) + \Sigma^{1/2}(y_{t-1}, y_{t-2}, \dots)w_t \quad (18.7.2)$$

where $\Sigma^{1/2}(\cdot)$ is defined by $\Sigma(\cdot) = \Sigma^{1/2}(\cdot)\Sigma^{1/2}(\cdot)$, $\Sigma(\cdot)$ denotes the conditional covariance matrix of the process in period t , given the information from previous periods, and w_t is a white noise process with mean zero and identity covariance matrix I_K . As for the conditional mean model, several nonparametric approaches exist for jointly estimating $\mu(\cdot)$ and $\Sigma^{1/2}(\cdot)$. Of course, it is also possible to specify a parametric form of one of the two objects of interest and to model the other one nonparametrically.

More generally the complete predictive (conditional) density $f(y_t | y_{t-1}, y_{t-2}, \dots)$, may be of interest, in particular, if higher order moments are relevant for the analysis. For this case as well a number of nonparametric approaches have been proposed. We first deal with local smoothing methods, before presenting the so-called semi-nonparametric (SNP) approach as one possible example of a global approximation method. Our analysis merely conveys the general ideas. A more detailed review of nonparametric modeling techniques for univariate time series with additional references can be found in Härdle, Lütkepohl, and Chen (1997). Many of these techniques generalize to the multivariate case, but their applicability may be limited by the curse of dimensionality.

Local Methods.

Estimator of a joint density. Suppose we are interested in the joint density of the Kp -dimensional vector $Y_t = (y'_t, \dots, y'_{t-p+1})'$,

$$f(Y_t) = f(y_t, \dots, y_{t-p+1}).$$

Given a sample of time series variables y_1, \dots, y_T plus all required presample values, one way of estimating $f(\cdot)$ is the kernel estimator

$$\hat{f}_T(Y) = \frac{1}{Th^{Kp}} \sum_{t=1}^T \mathbf{K}\left(\frac{Y - Y_t}{h}\right), \quad (18.7.3)$$

where Y is a Kp -dimensional vector, $\mathbf{K}(\cdot)$ is a multivariate kernel function, and the scalar $h > 0$ in this subsection is used to denote the so-called bandwidth. For example, $\mathbf{K}(\cdot)$ may be a multivariate standard normal density or, more generally, it may be expressed as a product

$$\mathbf{K}(\cdot) = \prod_{j=1}^{Kp} \mathbf{K}_j(\cdot),$$

where the $\mathbf{K}_j(\cdot)$ are univariate kernel functions. The bandwidth h is chosen as a decreasing function of the sample size T such that the neighborhood from which we estimate the functional value shrinks, as the sample size increases. Clearly, for the estimate to become increasingly accurate with increasing sample size, we require that more and more observations lie in the neighborhood of the function value under consideration. Often $\mathbf{K}(\cdot)$ will have compact support. Under suitable mixing conditions, and appropriate kernel functions and bandwidth choice, the estimator $\hat{f}_T(Y)$ is consistent. For this result and related asymptotic properties see Robinson (1983). For an expository discussion of the univariate case see Tschernig (2004).

Once we have obtained the joint density of the Y_t , the conditional densities and their conditional moments can be derived.

Conditional moments. The conditional mean

$$\mu(y_{t-1}, \dots, y_{t-p}) = \mathbb{E}(y_t | y_{t-1}, \dots, y_{t-p}) = \int y f(y | y_{t-1}, \dots, y_{t-p}) dy$$

may serve as the one-step ahead predictor in period $t - 1$. For a given $Kp \times 1$ vector Y , the conditional expectation function may be estimated as

$$\hat{\mu}(Y_{t-1}) = \frac{\sum_{t=1}^T \mathbf{K}\{(Y - Y_{t-1})/h\} y_t}{\sum_{t=1}^T \mathbf{K}\{(Y - Y_{t-1})/h\}}, \quad (18.7.4)$$

where $\mathbf{K}(\cdot)$ is again a kernel function and h is the bandwidth.

Predictive density. Similarly, the conditional density

$$f(y_t|Y_{t-1}) = f(y_t, y_{t-1}, \dots, y_{t-p})/f(y_{t-1}, \dots, y_{t-p})$$

may be considered. This predictive density may be estimated analogously to the joint density as

$$\hat{f}(y_t|Y_{t-1}) = \frac{h^{-K} \sum_{t=1}^T \mathbf{K}\{(Y^* - Y_{t-1}^*)/h\}}{\sum_{t=1}^T \mathbf{K}\{(Y - Y_{t-1})/h\}} = \frac{\hat{f}_T(Y^*)}{\hat{f}_T(Y)}, \quad (18.7.5)$$

where Y^* and $Y_t^* = (y'_t, \dots, y'_{t-p})'$ are $K(p+1)$ -dimensional vectors.

Instead of a kernel estimator, local polynomial approximations could also be used, as proposed by Härdle and Tsybakov (1997) and Härdle, Tsybakov, and Yang (1998), among others. All these approaches require larger samples than may be available in applied work. In VAR modeling, some simplifications therefore may be helpful, as discussed next.

An additive nonparametric VAR model. Because reliably estimating fully flexible functions by nonparametric local methods requires large samples, and because such samples are often not available in macroeconomic analysis, Jeliazkov (2013) proposes a more parsimonious nonlinear VAR model with additive errors of the form

$$\begin{pmatrix} y_{1t} \\ \vdots \\ y_{Kt} \end{pmatrix} = \begin{pmatrix} a_{11,1}(y_{1,t-1}) + \dots + a_{1K,1}(y_{K,t-1}) \\ \vdots \\ a_{K1,1}(y_{1,t-1}) + \dots + a_{KK,1}(y_{K,t-1}) \end{pmatrix} + \dots + \begin{pmatrix} a_{11,p}(y_{1,t-p}) + \dots + a_{1K,p}(y_{K,t-p}) \\ \vdots \\ a_{K1,p}(y_{1,t-p}) + \dots + a_{KK,p}(y_{K,t-p}) \end{pmatrix} + u_t, \quad (18.7.6)$$

where the error process u_t is a white noise process and the $a_{ij,k}(\cdot)$ are fully flexible functions of the lagged variables. The advantage of this setup is that all these functions are only one-dimensional. Jeliazkov (2013) proposes a Bayesian approach to estimating these functions nonparametrically and for choosing a suitable lag order. He also proposes a number of extensions of the model. For example, changes in volatility can be accommodated. His algorithm relies on smoothness priors that have been used also by a number of other authors (see Jeliazkov (2013) for further references).

As an example consider a quarterly model of the U.S. economy. Let $y_t = (\Delta gdp_t, ur_t, r_t, \pi_t)'$, where Δgdp_t is real GDP growth, ur_t is the unemployment rate, π_t is consumer price inflation, and r_t is the three-month Treasury bill rate. Jeliazkov fits a nonparametric VAR(1) model to data for 1948q1–2005q1. He shows that a number of the functions $a_{ij,k}(\cdot)$ are not very different from linear functions. Such results may be used to simplify the model. He also

finds, however, evidence of nonlinearities that must be accounted for to avoid misspecification bias. Notably some of the functions involving lags of inflation, $a_{i4,1}(\pi_{t-1})$, appear to be nonlinear in the Δgdp_t and ur_t equations.

Analysis along the lines of this example may in some cases suggest a more tightly parameterized nonlinear specification that can be used for structural analysis, as discussed in earlier sections of this chapter. It may also be used to assess whether a given parametric nonlinear specification is consistent with the data.

Global Approximation – the SNP Approach. In related work, Gallant and Tauchen (1989) use Hermite expansions to approximate the one-step ahead conditional density of the DGP of a multiple time series variable, y_t , given its past. This approach is based on the fact that a large class of density functions, $f(z)$, is proportional to $[\psi(z)]^2\phi(z)$, where $z = \Sigma_y^{-1/2}(y - \mu_y)$, with μ_y and $\Sigma_y^{-1/2}$ denoting the location and scale parameters, respectively, of the distribution, with $\psi(z)$ denoting a multivariate polynomial of possibly infinite degree r , and with $\phi(z)$ representing the multivariate standard normal density. Dividing $[\psi(z)]^2\phi(z)$ by a normalizing constant, this expression is just the Hermite expansion of $f(z)$. Hence, the density may be written as the product of a standard normal density and the square of a polynomial.

Suppose we are interested in the conditional density $f(y_t | y_{t-1}, y_{t-2}, \dots)$. Then

$$f(y_t | y_{t-1}, y_{t-2}, \dots) \propto [\psi(z_t)]^2\phi(z_t), \quad (18.7.7)$$

where $z_t = \Sigma_t^{-1/2}(y_t - \mu_t)$ with μ_t and $\Sigma_t^{-1/2}$ being location and scale parameters, respectively, of the conditional distribution. The former is assumed to be a linear function of the past, $\mu_t = v + A_1 y_{t-1} + \dots + A_p y_{t-p}$, while the latter may be modeled as an ARCH-type function (see Gallant, Hsieh, and Tauchen 1991; Gallant and Tauchen 1994). Regardless of the specification, the location and scale parameters μ_t and $\Sigma_t^{1/2}$ are modeled parametrically, whereas higher-order moments are captured by the polynomial. Letting the polynomial degree increase with the sample size makes this approach nonparametric. The approach is also known as semi-nonparametric (SNP) in the literature because it combines parametric with nonparametric elements.

To achieve a flexible adjustment of the model to higher-order dynamics the coefficients of the polynomial $\psi(\cdot)$ may be made dependent on lagged y_t . Of course, for polynomial degree $r = 0$ we have

$$f(y_t | y_{t-1}, y_{t-2}, \dots) \propto \phi\left(\Sigma_t^{-1/2}(y_t - v - A_1 y_{t-1} - \dots - A_p y_{t-p})\right)$$

so that a linear VAR(p) process with conditionally heteroskedastic error term emerges as a special case of the SNP model.

For given values of the integer parameters specifying the lag lengths and the polynomial degrees, this model can be estimated by maximizing the normalized log-likelihood,

$$\log l(\theta) = \frac{1}{T} \sum_{t=1}^T \log f(y_t | y_{t-1}, \dots, y_{t-p}; \theta).$$

Asymptotic properties of this estimation procedure are provided in Gallant and Nychka (1987) who allow the order of the Hermite expansion to increase with the sample size. Gallant and Tauchen (1994) propose model selection criteria for choosing the integer parameters of the model.

Ideally the SNP model reduces to a model that can be handled within a standard linear or nonlinear parametric VAR framework. Even if a more general model is required for an adequate description of the DGP, it may still be possible to derive the structural form of such a model. At this point, however, there are few empirical examples of such studies.

18.7.2 Noncausal VAR Models

An important characteristic of stationary linear VAR processes is that the variables y_t depend on lagged values y_{t-1}, y_{t-2}, \dots , plus a white noise error term such that

$$y_t = A_1 y_{t-1} + \dots + A_p y_{t-p} + u_t.$$

If the autoregressive operator $A(z) = I_K - A_1 z - \dots - A_p z^p$ has no roots in and on the complex unit disk, i.e.,

$$\det(A(z)) \neq 0 \quad \forall z \in \mathbb{C}, |z| \leq 1,$$

the process has a one-sided MA representation,

$$y_t = A(L)^{-1} u_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i},$$

that depends only on past and present error terms u_{t-i} , $i = 0, 1, \dots$. Such processes are called causal because y_t is determined by its past.⁴ In this case, if u_t is a martingale difference process, the conditional expectation

$$\mathbb{E}(y_t | y_{t-1}, y_{t-2}, \dots)$$

is the best prediction of y_t in period $t - 1$ and u_t is the one-step ahead prediction error,

$$u_t = y_t - \mathbb{E}(y_t | y_{t-1}, y_{t-2}, \dots).$$

⁴ This terminology is not to be confused with the notion of causality discussed in Chapter 7.

As discussed in Chapter 17, in some economic models current values of y_t also depend on future values of y_t (see Hansen and Sargent 1980, 1991). Such a situation arises when economic agents have more information than the econometrician. In that case, predictions based on the econometrician's VAR model will not be as accurate as the expectations of the economic agents who anticipate future values of some of the model variables and take this information into account in their decision-making. In this situation, the class of causal VAR models is too restrictive. Future values of y_t have to be allowed for in the DGP, resulting in a noncausal model.

Noncausal univariate AR models have been considered by Brockwell and Davis (1987, chapter 3), Breidt, Davis, Lii, and Rosenblatt (1991), and Lanne and Saikkonen (2011), for example. There are several proposals for the construction of multivariate noncausal models. For example, a noncausal VAR model suggested by Lanne and Saikkonen (2013) takes the form

$$A(L)C(L^{-1})y_t = u_t, \quad (18.7.8)$$

where $A(L) \equiv I_K - A_1L - \dots - A_pL^p$ is a matrix polynomial in the lag operator L and $C(L^{-1}) \equiv I_K - C_1L^{-1} - \dots - C_qL^{-q}$ is a matrix polynomial in the inverse lag operator L^{-1} that is defined such that $L^{-1}y_t = y_{t+1}$. For $p = q = 1$, $A(L)C(L^{-1}) = I_K + A_1C_1 - A_1L - C_1L^{-1}$, and the process (18.7.8) simplifies to

$$(I_K + A_1C_1)y_t = A_1y_{t-1} + C_1y_{t+1} + u_t.$$

Thus, y_t depends on lagged and future values. Clearly, if $q = 0$ and, hence, $C(L^{-1}) = I_K$, y_t reduces to a standard causal VAR process.

Lanne and Saikkonen (2013) assume that both operators $A(L)$ and $C(L)$ are invertible with all roots outside the complex unit circle such that

$$\det(A(z)) \neq 0 \quad \text{and} \quad \det(C(z)) \neq 0 \quad \forall z \in \mathbb{C}, |z| \leq 1.$$

Hence, $x_t = C(L^{-1})y_t$ is a stable, stationary process,

$$A(L)x_t = u_t$$

with MA representation

$$x_t = \sum_{i=0}^{\infty} \Phi_i u_{t-i}.$$

The process y_t is also stationary in this case. However, in general y_t is easily seen to have a two-sided MA representation in terms of the u_t errors,

$$y_t = \sum_{i=-\infty}^{\infty} \Psi_i u_{t-i}.$$

Moreover, unlike in the standard causal VAR case,

$$y_t - \mathbb{E}(y_t | y_{t-1}, y_{t-2}, \dots) \neq u_t.$$

Hence, the VAR model errors u_t are not one-step ahead prediction errors in general. In other words, u_t is y_t -nonfundamental in the terminology used in Chapter 17. In fact, there is in general no closed-form expression for the conditional expectation $\mathbb{E}(y_t | y_{t-1}, y_{t-2}, \dots)$. This conditional expectation is a highly nonlinear function of lagged y_t and, hence, predictions may have to be computed by simulation. This is also the reason why noncausal models are discussed in this chapter on nonlinear VAR models. Since impulse responses can also be viewed as conditional expectations based on past information, they are nonlinear functions of the shocks as well.

Lanne and Saikkonen (2013) develop an ML procedure for estimating the parameters of their model. They also provide advice on how to specify noncausal VAR models. An important point to note in this context is that stationarity of y_t implies that there exists a one-sided Wold MA representation that generates the same autocovariance structure as that of y_t . Thus, if y_t is Gaussian and, hence, also u_t is Gaussian, the DGP is indistinguishable from a causal process with the same first and second moments. Since Gaussian processes are fully determined by the first and second moments, noncausal VAR models are of interest only if the distribution of y_t is nonnormal.

As mentioned earlier, there are other ways of specifying noncausal VAR models. So far we have focused on the framework of Lanne and Saikkonen (2013) who use the multiplicative VAR operator in (18.7.8). Alternatively, one may consider a general additive VAR operator with leads and lags and write $D(L)y_t = u_t$ with

$$D(L) = I_K - D_1L - \dots - D_pL^p - D_{-1}L^{-1} - \dots - D_{-r}L^{-r}.$$

Yet another approach is to consider the process

$$A^*(L)y_t = u_t \tag{18.7.9}$$

with one-sided operator $A^*(L) = I_K - A_1^*L - \dots - A_p^*L^p$. This process is allowed to have roots inside (but not on) the unit circle such that

$$\det(A^*(z)) \neq 0 \quad \forall z \in \mathbb{C}, |z| = 1,$$

but there may be some z_0 with modulus less than 1 such that $\det(A^*(z_0)) = 0$. In other words, only roots on the unit circle are excluded. If there are roots inside the unit disk, then y_t is noncausal. This framework has been used by Davis and Song (2012), for example. Their framework encompasses processes that are not covered by the multiplicative formulation used by Lanne and Saikkonen (2013). The reverse, however, is also true in that there are

multiplicative models that cannot be embedded in (18.7.9), so it may matter which framework is used.

Applications of this class of models can be found in Lanne and Saikkonen (2013) and Davis and Song (2012). These studies consider term structure models of interest rates with two factors, one defined as the difference between a long-term and a short-term interest rate and the other defined as the change in the short-term interest rate. Such systems have been used in the past to test the expectations hypothesis of the term structure within the framework of causal models. Lanne and Saikkonen and Davis and Song both find evidence for a noncausal model. Their result casts doubt on previous studies of the expectations hypothesis of the term structure based on causal VAR models.

18.8 Discussion of Nonlinear VAR Modeling

This chapter has shown that there are many types of nonlinear VAR models, raising the question of how to choose among these specifications. It is generally good advice to use nonlinear models only when there is a compelling economic rationale for nonlinearities. Often the economic context dictates the type of nonlinearity that is called for. For example, institutional constraints may suggest the existence of thresholds, a recurring pattern of expansions and contractions may suggest a Markov-switching model, or theoretical economic models may suggest a GARCH-in-mean VAR specification. Knowing what type of nonlinearity one is interested in also makes it easier to determine whether there is enough variation in the sample to accurately estimate the nonlinear phenomenon of interest. One would not expect to be able to estimate accurately any type of nonlinearity that occurs only rarely in the data or that occurs for the first time at the very end of the sample such as the zero lower bound on U.S. interest rates after 2007.

There are situations, however, when the choice of the nonlinear specification is not as clear-cut, because there may be more than one economic rationale to consider. A case in point is the literature on asymmetric responses to positive and negative oil price shocks. Such asymmetries may be motivated based on any number of mechanisms including real options theory, precautionary savings models, reallocation models with economic frictions, or simply behavioral models (see Kilian 2014). In such a situation it can be difficult to separate the empirical content of alternative economic models, and any nonlinear model focusing on one economic mechanism only, while excluding the other explanations, is likely to be misspecified. For example, a VAR model with GARCH-in-mean errors would be inherently misspecified in the presence of other sources of asymmetric responses. Moreover, the DGP may be subject to additional nonlinearities not considered by standard economic models,

which also would undermine the validity of the GARCH-in-mean VAR model. Examples are changes in the share of oil in real output or the development of alternative sources and uses of energy.

In this case, a less parametric approach that encompasses many competing economic explanations such as the TVC-VAR model may be preferable, provided that approach is computationally feasible. The approach of fitting TVC-VAR models is not without risk, however, because nonlinear models tend to overfit in small samples, especially in the presence of large economic events. It is, of course, possible to guard against overfitting by tightening the priors used in estimating TVC-VAR models, but the use of such prior information becomes problematic if the prior is chosen for convenience rather than on economic grounds, as is typically the case in practice. Nor is the standard TVC-VAR model with slope parameters evolving according to unrestricted random walks designed to approximate stationary nonlinear processes. Its use only makes sense if there is some credible source of nonstationarity. Another caveat is that the TVC-VAR model is not designed for situations where the model is time-invariant for part of the sample and subject to structural change for the remainder of the sample.

More generally, it is important to keep in mind that the greater flexibility of nonlinear models in fitting the data need not translate into meaningful estimates. Of course, the reverse is also true in that nonlinearities may not be apparent if there is not enough variation in the data. Only large economic events such as a major recession or a stock market crash may reveal nonlinear behavior in the data. Thus, to the extent possible, estimates of such models should be checked against extraneous evidence. Users of these models need to make the case that the timing, direction, and magnitude of the variation in the structural model parameters is economically plausible.

It is also worth pointing out that the distinction between a nonlinearity and a structural break may not be easy to make on empirical grounds. Some of the models presented in this chapter can in fact be viewed as models for structural change. For example, TVC-VAR models designed to capture smooth structural change in practice often are able to detect discrete structural changes as well. Likewise, an MS-VAR model may be suitable for detecting structural breaks. The MS-VAR model assumes that the economy can be in a finite number of alternative states. It assigns, on the basis of the sample information, probabilities to these states in each period of the sample. Although typically the system will alternate between states, such a model can also capture a single structural break.

Whether such a break should be modeled as a deterministic structural break or not, is mainly a question of its economic interpretation. Deterministic breaks tend to be associated with discrete changes in institutions that are effectively permanent such as a currency union or the liberalization of a market. Most

structural changes that economists are concerned with are likely to be stochastic rather than deterministic and continuous (gradual and smooth) rather than discrete (sudden and abrupt). The key difference is that deterministic breaks are assumed not to occur again in the future, whereas stochastic breaks would be expected to occur according to some possibly exogenous stochastic process. This fact makes a difference for the construction of structural impulse responses as well as for forecasting.

Finally, it should be noted that many events commonly associated with structural breaks in the literature such as World War II or the Great Depression are not likely sources of permanent structural changes. Standard economic theory, for example, implies that a temporary surge in government spending triggered by a war would have a temporary effect on the level of real output. Likewise, to the extent that the Great Depression was caused by monetary policy, it should represent a temporary contraction rather than a permanent structural shift. Such large transitory dynamics may be difficult to distinguish from nonlinearities or structural change even in moderately large samples (see Kilian and Ohanian 2002).

18.9 Linear Structural Models with Nonlinear Transformations of the Variables

Generalizing VAR models to incorporate nonlinearities does not necessarily imply specifications that are nonlinear in the model parameters. Instead models may be linear in the parameters, but nonlinear in the variables. This point is best illustrated by the literature on modeling the transmission of oil price shocks. It is common in applied work to rely on nonlinear transformations of the price of oil in modeling the relationship between the price of oil and the domestic economy. The objective of such transformations is to capture asymmetries and other potential nonlinearities in the transmission of oil price shocks to the economy. Often the transformations used involve some censoring of the data. For example, Mork (1989) considered the percent increase in the real price of oil, defined as $\Delta r_t^+ = \max(0, \Delta r_t)$, where Δr_t is the percent change in the real price of oil and r_t is the log level of the real price of oil. Hamilton (1996) proposed replacing the Mork measure by the 1-year net oil price increase measure, defined as $\Delta r_t^{net,+,\text{1yr}} = \max(0, r_t - r_t^*)$, where r_t^* denotes the highest log real price of oil over the preceding year. Finally, Hamilton (2003) recommended the 3-year net oil price increase measure based on the maximum price of oil over the preceding three years. The latter specification is the oil price transformation used most commonly in applied work.

Our discussion, without loss of generality, focuses on the 3-year net oil price increase measure. Both nominal and real oil price transformations have been used in applied work, although only the latter specification is consistent with

economic theory. We follow the recent literature in focusing on the real price of crude oil and its relationship with U.S. real GDP.

18.9.1 The Censored Oil Price VAR Model

Traditionally, researchers interested in the relation between the real price of oil and U.S. economic growth have relied on reduced-form VAR models of the form

$$\begin{aligned}\Delta r_t^{net,+3yr} &= v_1 + \sum_{i=1}^p a_{11,i} \Delta r_{t-i}^{net,+3yr} + \sum_{i=1}^p a_{12,i} \Delta gdp_{t-i} + u_{1t}, \\ \Delta gdp_t &= v_2 + \sum_{i=1}^p a_{21,i} \Delta r_{t-i}^{net,+3yr} + \sum_{i=1}^p a_{22,i} \Delta gdp_{t-i} + u_{2t},\end{aligned}\tag{18.9.1}$$

where Δgdp_t refers to the growth rate of U.S. real GDP and $\Delta r_t^{net,+3yr} = \max(0, r_t - r_t^*)$, where r_t^* denotes the highest log real price of oil over the preceding three years, as defined earlier. Identification is achieved by imposing that the real price of oil is predetermined with respect to U.S. real GDP (see Kilian and Vega 2011). In practice, the structural model is inferred by applying a lower-triangular Cholesky decomposition to the reduced-form white noise covariance matrix Σ_u under the assumption that the real price of oil is predetermined. Impulse responses are constructed exactly as in linear structural VAR models. Examples of this approach include Hamilton (1996), Bernanke, Gertler, and Watson (1997), and Ramey and Vine (2011), among many others.

These models became popular among macroeconomists because of their ability to generate “better looking impulse responses”, referring to the model’s ability to generate larger economic declines in response to positive oil price shocks than linear VAR models (see Bernanke, Gertler, and Watson 1997). This feature is an artifact of the model specification, however, rather than driven by the data.

Kilian and Vigfusson (2011a) demonstrate that asymmetric models of the transmission of oil price shocks cannot be represented as censored oil price VAR models of the form (18.9.1). Whether the DGP is symmetric or asymmetric in oil price increases and oil price decreases, the censored oil price model is inherently misspecified. This misspecification renders the parameter estimates inconsistent and inference invalid. Moreover, standard approaches to generating structural impulse responses from model (18.9.1) are invalid because of the presence of censored variables. As a result, the censored oil price VAR model tends to substantially overstate the recessionary impact of positive oil price shocks by construction.

18.9.2 A Nonlinear Structural Model Allowing for Asymmetric Responses

As an alternative, Kilian and Vigfusson (2011a, 2011b) propose the structural model

$$\begin{aligned}\Delta r_t &= v_1 + \sum_{i=1}^p b_{11,i} \Delta r_{t-i} + \sum_{i=1}^p b_{12,i} \Delta gdp_{t-i} + w_{1t}, \\ \Delta gdp_t &= v_2 + \sum_{i=0}^p b_{21,i} \Delta r_{t-i} + \sum_{i=1}^p b_{22,i} \Delta gdp_{t-i} \\ &\quad + \sum_{i=0}^p \gamma_i \Delta r_{t-i}^{net,+3yr} + w_{2t}, \\ \Delta r_t^{net,+3yr} &\equiv \max(0, r_t - r_t^*),\end{aligned}\tag{18.9.2}$$

where the structural shocks w_{1t} and w_{2t} are mutually uncorrelated and serially uncorrelated. The third equation defines the nonlinear regressors in the second equation. This model allows for all features that the censored oil price VAR model was intended to capture, while preserving the linearity of the equation of the real price of oil. It also directly imposes the conventional identifying assumption that the real price of oil is predetermined with respect to U.S. real GDP. Model (18.9.2) is not a conventional structural vector autoregression because the set of regressors differs across equations and because the net oil price increase affects Δgdp_t even in the impact period. As a result, it would not be possible to rewrite this model as a reduced-form model to be identified by a Cholesky decomposition. The model structure is nonlinear in that the effect of a positive oil price shock, w_{1t} , differs from the effect of a negative oil price shock of the same magnitude. Moreover, the response of real GDP changes disproportionately with the magnitude of the oil price shock and depends on the information set, Ω_{t-1} , at the time of the shock.

Model (18.9.2) encompasses as a special case the linear VAR(p) model

$$\begin{aligned}\Delta r_t &= v_1 + \sum_{i=1}^p b_{11,i} \Delta r_{t-i} + \sum_{i=1}^p b_{12,i} \Delta gdp_{t-i} + w_{1t} \\ \Delta gdp_t &= v_2 + \sum_{i=0}^p b_{21,i} \Delta r_{t-i} + \sum_{i=1}^p b_{22,i} \Delta gdp_{t-i} + w_{2t},\end{aligned}\tag{18.9.3}$$

for $\gamma_i = 0$, $i = 0, 1, \dots, p$, in model (18.9.2).

In estimating model (18.9.2) we can take advantage of the fact that the model is linear in the parameters. The model may be estimated consistently using equation-by-equation least squares, and asymptotic inference on the parameters is standard. The inclusion of contemporaneous regressors in the

second equation means that the structural shocks may be estimated directly. The construction of the structural impulse responses, in contrast, is complicated by the model being nonlinear in the variables which renders the impulse responses dependent on the state of the system at the time of the shock and the sign and size of the shock. It requires Monte Carlo integration, building on the algorithm outlined in Section 18.2.2.

18.9.3 Quantifying Nonlinear Responses to Oil Price Shocks

Having estimated the encompassing model (18.9.2) on the full sample, under the assumption of independent structural shocks, the structural impulse responses may be computed as follows:

Algorithm *Conditional response function at date t*

1. Consider a block of p consecutive values of Δr_τ , $\Delta r_\tau^{net,+3yr}$, and Δgdp_τ for $\tau = t - p, \dots, t - 1$. This sequence defines a history Ω_{t-1} . We are interested in quantifying the response of real GDP growth over the next $H + 1$ quarters to an oil price innovation of magnitude δ occurring at date t , conditional on Ω_{t-1} .
2. We simulate two alternative time paths for Δr_{t+h} and Δgdp_{t+h} for $h = 0, \dots, H$ from model (18.9.2) by iterating the model forward, given Ω_{t-1} and given the estimated coefficients of model (18.9.2). In simulating future realizations of the data, we need to take a stand on the structural shocks, w_{t+h} , for $h = 0, \dots, H$. When generating the first time path, the value of w_{1t} is set equal to a prespecified value δ , denoting the magnitude of the oil price shock of interest. The realizations of the subsequent structural innovations, $w_{1,t+h}$ for $h = 1, \dots, H$, are drawn from the marginal empirical distribution of w_{1t} . The realizations of $w_{2,t+h}$ for $h = 0, 1, \dots, H$ are drawn independently from the marginal empirical distribution of w_{2t} . When generating the second time path, all $w_{1,t+h}$ and $w_{2,t+h}$ realizations for $h = 0, \dots, H$ are drawn independently from their respective marginal distributions.
3. We then calculate the difference between these two time paths for Δgdp_{t+h} , $h = 0, 1, \dots, H$.
4. We average this difference across m repetitions of steps 2 and 3, where m has to be large enough to invoke the law of large numbers.

This average represents the conditional response of Δgdp_{t+h} at horizon $h = 0, 1, \dots, H$ to a shock of magnitude δ given Ω_{t-1} , denoted by

$$I_{\Delta gdp}(H, \delta, \Omega_{t-1}).$$

The corresponding unconditional response function $I_{\Delta gdp}(H, \delta)$ for horizons $h = 1, \dots, H$ is defined as the value of the conditional response function $I_{\Delta gdp}(H, \delta, \Omega^r)$ averaged over many randomly selected histories Ω^r , each of which is generated by randomly drawing with replacement a block of p consecutive observations from the observed data:

$$I_{\Delta gdp}(H, \delta) = \int I_{\Delta gdp}(H, \delta, \Omega^r) d\Omega^r.$$

The unconditional response, as implemented in Kilian and Vigfusson (2011a), is a measure of the overall effect of oil price shocks on U.S. real GDP. When studying the effects of oil price shocks during specific historical episodes commencing at date t , the appropriate impulse response measure instead is the conditional response to oil price shocks of magnitude δ , given the observed sequence of data preceding date t . The latter approach has been utilized in Kilian and Vigfusson (2017).

Bootstrap confidence intervals for the structural impulse responses may be constructed by bootstrapping the encompassing model. The asymptotic validity of this bootstrap approach has not yet been formally established, however.

18.9.4 Testing the Null of Unconditionally Symmetric Response Functions

One of the reasons researchers have investigated models based on censored oil price variables is the perception that the responses triggered by a positive and by a negative oil price shock of the same magnitude are asymmetric. In other words, the resulting impulse response functions are not perfect mirror images of one another. Typically, the expectation is that positive oil price shocks are more recessionary than negative oil price shocks are stimulating for the economy.

Symmetry in the impulse response functions requires that $\gamma_i = 0$ for $i = 0, 1, \dots, p$. It may be tempting therefore to test for symmetry in the response functions by testing whether the γ_i parameters are jointly equal to zero. Such a test would not be informative about the degree of asymmetry in the impulse response functions, however, because the impulse responses are highly nonlinear functions of the slope parameters of model (18.9.2). It can be shown that, even when statistical tests reject the null of symmetric slope parameters, the impulse response functions may be nearly symmetric. As usual, a failure to reject the null of symmetric slopes may reflect the low power of the test and does not guarantee symmetric response functions.

Moreover, we know that the degree of asymmetry of the impulse response functions depends on δ . Slope parameter based tests are invariant to the magnitude of δ by construction and hence ill-suited for assessing the degree of

asymmetry in the response functions. A more direct test of the symmetry of the response functions for a given δ involves three steps.

1. Estimate the encompassing model (18.9.2) on the full sample.
2. Compute $I_{\Delta gdp}(H, \delta)$ and $I_{\Delta gdp}(H, -\delta)$.
3. Test $H_0 : I_{\Delta gdp}(H, \delta) = -I_{\Delta gdp}(H, -\delta)$ against the alternative $H_1 : I_{\Delta gdp}(H, \delta) \neq -I_{\Delta gdp}(H, -\delta)$. In other words, under the null hypothesis the $(H+1) \times 1$ vector $I_{\Delta gdp}(H, \delta) + I_{\Delta gdp}(H, -\delta)$ should be equal to a vector of zeros. Standardizing this vector based on a bootstrap estimator of the variance covariance matrix of the vector $I_{\Delta gdp}(H, \delta) + I_{\Delta gdp}(H, -\delta)$ results in the Wald test statistic

$$\begin{aligned} & \left(\widehat{I}_{\Delta gdp}(H, \delta) + \widehat{I}_{\Delta gdp}(H, -\delta) \right)' \widehat{\Sigma}_{\widehat{I}_{\Delta gdp}(H, \delta) + \widehat{I}_{\Delta gdp}(H, -\delta)}^{*-1} \\ & \quad \times \left(\widehat{I}_{\Delta gdp}(H, \delta) + \widehat{I}_{\Delta gdp}(H, -\delta) \right), \end{aligned}$$

where $*$ denotes bootstrap estimators based on the unrestricted model (18.9.2). The null hypothesis of symmetric response functions is rejected if the Wald statistic exceeds the critical value from the $\chi^2(H+1)$ distribution.

This procedure can be easily adapted to testing the symmetry of cumulative impulse response functions.

In practice, it is common to consider results for oil price shocks of size $\delta = \sigma$ and $\delta = 2\sigma$, where σ is one standard deviation of the estimated w_{1t} . When the test rejects the null of a symmetric response function at conventional significance levels, it is recommended to check the plots of the response functions to ascertain whether the rejection is also economically significant. In choosing H , it has to be kept in mind that for large H the $\chi^2(H+1)$ approximation will become poor. In practice, a maximum horizon of one year is a common choice. Using one fixed H also avoids the problems associated with data mining across H .

18.9.5 Testing the Null of Conditionally Symmetric Response Functions

One concern with relying on tests of the null of unconditionally symmetric response functions is that the degree of asymmetry may evolve over time. In this case, unconditional tests may have low power against departures from symmetry. Kilian and Vigfusson (2017) propose a complementary test of the null of conditionally symmetric response functions on a specific date t . The test is conducted in three steps:

1. Estimate the encompassing model (18.9.2) on the full sample.
2. Compute $\widehat{I}_{\Delta gdp}(H, \delta, \Omega_{t-1})$ and $\widehat{I}_{\Delta gdp}(H, -\delta, \Omega_{t-1})$.

3. Test $\mathbb{H}_0 : I_{\Delta gdp}(H, \delta, \Omega_{t-1}) = -I_{\Delta gdp}(H, -\delta, \Omega_{t-1})$ against the alternative $\mathbb{H}_1 : I_{\Delta gdp}(H, \delta, \Omega_{t-1}) \neq -I_{\Delta gdp}(H, -\delta, \Omega_{t-1})$ using the Wald test statistic

$$\begin{aligned} & \left(\widehat{I}_{\Delta gdp}(H, \delta, \Omega_{t-1}) + \widehat{I}_{\Delta gdp}(H, -\delta, \Omega_{t-1}) \right)' \\ & \times \widehat{\Sigma}_{\widehat{I}_{\Delta gdp}^*(H, \delta, \Omega_{t-1}) + \widehat{I}_{\Delta gdp}^*(H, -\delta, \Omega_{t-1})}^{*-1} \\ & \times \left(\widehat{I}_{\Delta gdp}(H, \delta, \Omega_{t-1}) + \widehat{I}_{\Delta gdp}(H, -\delta, \Omega_{t-1}) \right), \end{aligned}$$

where * denotes bootstrap estimators based on the unrestricted model (18.9.2). The null hypothesis of conditionally symmetric response functions is rejected if the test statistic exceeds the critical value from the $\chi^2(H+1)$ distribution.

18.9.6 Testing the Null of No Time Dependence

A common question is whether the conditional response is time dependent or whether it is indistinguishable from a linear VAR model response. Kilian and Vigfusson (2017) propose a test of the null of conditionally linear response functions on a specific date t . The test compares the conditional responses of real GDP growth to a real oil price shock from the nonlinear model, $\widehat{I}_{\Delta gdp}(H, \delta, \Omega_{t-1})$, to the corresponding quantities from a linear VAR model, denoted by $\widehat{I}_{\Delta gdp}^{VAR}(H, \delta)$. Note that the latter quantities do not depend on the specific date t for which the test is performed. There are four steps.

1. Estimate the encompassing model (18.9.2) and the linear VAR model (18.9.3) on the full sample.
2. Compute the conditional nonlinear response function $\widehat{I}_{\Delta gdp}(H, \delta, \Omega_{t-1})$ from the estimate of the encompassing model (18.9.2).
3. Compute the linear VAR response function $\widehat{I}_{\Delta gdp}^{VAR}(H, \delta)$ from the estimate of the linear VAR model (18.9.3).
4. Test $\mathbb{H}_0 : I_{\Delta gdp}(H, \delta, \Omega_{t-1}) = I_{\Delta gdp}^{VAR}(H, \delta)$ against the alternative $\mathbb{H}_1 : I_{\Delta gdp}(H, \delta, \Omega_{t-1}) \neq I_{\Delta gdp}^{VAR}(H, \delta)$ using the Wald test statistic

$$\begin{aligned} & \left(\widehat{I}_{\Delta gdp}(H, \delta, \Omega_{t-1}) - \widehat{I}_{\Delta gdp}^{VAR}(H, \delta) \right)' \\ & \times \widehat{\Sigma}_{\widehat{I}_{\Delta gdp}^*(H, \delta, \Omega_{t-1}) - \widehat{I}_{\Delta gdp}^{VAR*}(H, \delta)}^{*-1} \\ & \times \left(\widehat{I}_{\Delta gdp}(H, \delta, \Omega_{t-1}) - \widehat{I}_{\Delta gdp}^{VAR}(H, \delta) \right), \end{aligned}$$

where * denotes bootstrap estimators. The null hypothesis of conditionally linear response functions is rejected if the test statistic exceeds the critical value from the $\chi^2(H+1)$ distribution. The

$\widehat{I}_{\Delta gdp}^*(H, \delta, \Omega_{t-1})$ and $\widehat{I}_{\Delta gdp}^{VAR*}(H, \delta)$ estimators are constructed by fitting models (18.9.2) and (18.9.3), respectively, to each of the bootstrap draws of the model data generated from the fitted encompassing model. The construction of $\widehat{\Sigma}^*$ takes account of the variances of the linear and the nonlinear bootstrap impulse response estimators as well as their covariances.

18.9.7 Conditional Prediction Error Decompositions

Impulse responses do not convey the cumulative effects of an entire sequence of oil price innovations on the log-level of U.S. real GDP over the course of a particular episode, starting at date t . Kilian and Vigfusson (2017) propose a conditional structural prediction error decomposition designed to address this problem. The total prediction error of model (18.9.2) at horizons h is $gdp_{t+h} - \mathbb{E}(gdp_{t+h}|\Omega_{t-1})$. A negative prediction error, for example, means that the model over-predicts U.S. real GDP. In other words, real GDP turned out lower than predicted.

One can decompose these prediction errors into the component explained by unforeseen oil price shocks, $w_{1,t+h}$, and the component explained by unforeseen residual shocks, $w_{2,t+h}$, in the second equation of the structural model. The conditional structural real GDP prediction error decomposition $D_{i,gdp}$, $i \in \{1, 2\}$, describes how real GDP would have evolved in the absence of further oil price innovations conditional on Ω_{t-1} :

$$D_{1,gdp}(h, \Omega_{t-1}) = \mathbb{E}\left(gdp_{t+h} | \{w_{1,t+i}\}_{i=0}^h, \Omega_{t-1}\right) - \mathbb{E}(gdp_{t+h}|\Omega_{t-1})$$

and, in the absence of further innovations in the second equation, conditional on Ω_{t-1} ,

$$D_{2,gdp}(h, \Omega_{t-1}) = \mathbb{E}\left(gdp_{t+h} | \{w_{2,t+i}\}_{i=0}^h, \Omega_{t-1}\right) - \mathbb{E}(gdp_{t+h}|\Omega_{t-1})$$

where $h = 0, \dots, H$. The expectations must be evaluated by Monte Carlo simulation.

A similar analysis could also be conducted with the linear VAR model (18.9.3). Whereas in the linear model by construction

$$D_{1,gdp}(h, \Omega_{t-1}) + D_{2,gdp}(h, \Omega_{t-1}) = gdp_{t+h} - \mathbb{E}(gdp_{t+h}|\Omega_{t-1}), \quad (18.9.4)$$

in the nonlinear model (18.9.2) there may be interactions between the two structural innovations that prevent the decomposition from adding up. Such interactions often are negligible in practice, however, and can be checked based on equation (18.9.4).

18.9.8 Extensions

Applications of tests for symmetry to disaggregate data inevitably invite data mining as these tests are applied repeatedly (see Inoue and Kilian 2004). Herrera, Lagalo, and Wada (2011) extend the Kilian-Vigfusson methodology by using a bootstrap approach to generate data-mining robust critical values under the null of symmetric response functions. Kilian and Vigfusson (2017) discuss extensions to nonlinear structural models with additional variables. An extension of the Kilian-Vigfusson methodology to structural vector error correction models is developed in Venditti (2013). Although the example of oil price shocks is well suited to illustrating the econometric issues in dealing with censored variables, the applicability of this methodology is not limited to studying the effects of oil price shocks. For example, Hussain and Malik (2016) use this methodology to investigate asymmetries in the response of U.S. real GDP to fiscal policy shocks.

19 Practical Issues Related to Trends, Seasonality, and Structural Change

This chapter reviews selected topics that are relevant to the specification of reduced-form VAR models, but are not covered in Chapters 2 and 3. Our discussion selectively draws on insights from earlier chapters including Chapter 18.

19.1 Alternative Trend Models

It is common to incorporate time trends into the specification of the VAR model. For example, we may include a deterministic trend polynomial in the model or we may allow for a stochastic trend. In Chapter 18, we saw that by relaxing the constraint of linearity, VAR models also can accommodate much richer stochastic trend models. There are situations, however, when it is more convenient to detrend the data upfront before fitting the VAR model. An example is the construction of data for the U.S. output gap by computing the deviation of U.S. real GDP from the level of potential U.S. real GDP estimated by the Congressional Budget Office. In addition, there are alternative trend models that do not allow for the joint estimation of the trend and the stochastic component of the VAR model. Examples include the trend filters commonly used in evaluating business cycle models in macroeconomics. In the latter case, the data must be detrended prior to fitting the VAR model.

Unlike a deterministic linear trend model, these trend filters allow the trend rate of growth to evolve over time. Perhaps the most popular trend filter in modern macroeconomics is the HP filter named after Hodrick and Prescott (1997). Other examples include the band-pass filter of Baxter and King (1999) and the random walk filter of Christiano and Fitzgerald (2003).

19.1.1 Hodrick-Prescott (HP) Filter

The HP filter was originally developed to determine the cyclical (or business cycle) component of fluctuations in U.S. postwar real GDP. It is based on the

premise that the rate of economic growth accelerates and decelerates in a fairly regular pattern of several years duration that can be captured by a smooth line. The HP trend is extracted from a scalar time series x_t using a two-sided symmetric moving average filter. Given a time series x_1, \dots, x_T , the trend component τ_1, \dots, τ_T is determined as the solution to the following minimization problem:

$$\min_{\{\tau_t\}_{t=1}^T} \sum_{t=1}^T (x_t - \tau_t)^2 + \lambda \sum_{t=2}^{T-1} [(\tau_{t+1} - \tau_t) - (\tau_t - \tau_{t-1})]^2. \quad (19.1.1)$$

The objective is to minimize the variance of the cyclical component $c_t \equiv x_t - \tau_t$ subject to a penalty in the second difference of τ_t , which measures the acceleration of the trend line. The parameter λ controls the degree of smoothness of the trend component. In the limit, as $\lambda \rightarrow \infty$, the trend component will coincide with a linear deterministic trend. At the other extreme, for $\lambda = 0$, $x_t = \tau_t$. For quarterly U.S. real GDP data, Hodrick and Prescott recommend $\lambda = 1600$; for the corresponding annual data, $\lambda = 400$. These choices are ad hoc. In practice, it may be useful to experiment with different degrees of smoothness and to verify whether the implied cyclical component matches extraneous information about the business cycle. For further discussion of the choice of λ the reader is referred to Ravn and Uhlig (2002).

There are two important caveats. One is that the HP filter described above is a two-sided MA filter that is not designed to estimate the trend component at the edges of the sample. Standard code available for the HP filter makes some ad hoc adjustments to overcome this problem, but in general we have to be cautious in interpreting results for the first and last observation. Another potential limitation of the HP filter is that it is applied to each series individually. To the extent that often several macroeconomic aggregates share a common trend, as would be the case in modeling output, consumption, and investment, it makes sense to filter these series jointly. Such a multivariate generalization of the HP filter was proposed by Kozicki (1999).

19.1.2 Band-Pass Filters

The HP filter is a low-pass filter in that it allows fluctuations that last less than a certain number of years to pass the filter and prevents longer-lasting fluctuations from passing. Under suitable choices for λ we can focus on fluctuations lasting less than, say, eight years which is commonly considered the maximum length of a business cycle. The cycle in this case is defined as all fluctuations that are not part of the trend. Typically, however, we would not consider fluctuations lasting less than, say, two years to be part of the business cycle. This observation led to the development of band-pass filters. Band-pass filters can be designed to filter out all fluctuations that last less than two years or longer

than eight years, for example, allowing us to focus on the variability of the data at business cycle frequencies (see, e.g., Canova 1998; Baxter and King 1999). The HP filter may be viewed as a special case of a band-pass filter.

19.1.3 Potential Shortcomings of Trend Filters

All trend filters are inherently ad hoc. They define the trend rather than estimating some prespecified trend component. More importantly, they may produce distortions in the filtered data. A well-known example of the dangers of using trend filters is provided by early business cycle analysis. When researchers, starting in the 1920s, smoothed noisy raw data for real economic activity using simple moving average filters, they detected various apparent cycles and waves of different lengths that received considerable attention in the literature at the time. For example, Kuznets (1961) obtained a business cycle of 15 to 25 years duration, named the Kuznets cycle, after applying a low-order moving average filter to U.S. macroeconomic time series. Subsequently, Howrey (1968) showed that the same cycle arises, when Kuznets' filter is applied to artificially generated white noise data (which have no cycles by construction), calling into question the existence of a Kuznets cycle in the U.S. data. This experience has increased awareness of the potential limitations of applying trend filters to macroeconomic data.

While some degree of distortion is inevitable, one criterion of the quality of a trend filter is that it must provide reasonable trend components when applied to data generated from a range of processes thought to be representative of macroeconomic time series data. The HP filter, while specified in an ad hoc manner, performs about as well in this regard as do other more recently developed trend filters (see Christiano and Fitzgerald 2003). Nevertheless, there is considerable evidence of the use of the HP filter producing spurious dynamic relations among time series (see, e.g., Hamilton 2016).

19.1.4 Trend-Filtered Variables in VAR Models

Trend-filtered variables obtained by using the HP filter or other band-pass filters may be accommodated in VAR analysis as follows. Let x_t denote a trending time series such as the log of real output. Suppose we decompose x_t into its trend component, τ_t , and the cyclical component, $c_t \equiv x_t - \tau_t$, defined as the deviation from trend such that trend and cycle are uncorrelated. Once the trend filter has been applied to x_t and the business cycle component c_t has been extracted, we replace x_t in the set of variables to be modeled as a VAR process by c_t and fit a VAR model to the new set of variables. Of course, it is also possible to detrend more than one time series in this manner.

Although this approach is straightforward, there is increased recognition in the literature that statistical filters that can be represented as a symmetric,

two-sided moving average of the raw data inevitably distort the estimates of the impulse responses and that applying such filters to some variables in the model but not to others creates further misspecification biases (see Canova 2014). Thus, the use of band-pass filters in VAR analysis should be avoided.

19.1.5 Choosing between Different Trend Models

The choice of the detrending method in a particular application by necessity requires judgement. There rarely is only one correct treatment of the trend. One way of discriminating between different detrending methods is to evaluate the implied cyclical component against extraneous information such as the NBER business cycle dates. A detrending method that implies that the U.S. economy was operating above trend during an NBER recession, for example, would be considered implausible. Alternative data-based business cycle dating methods have been discussed, for example, in Harding and Pagan (2002).

19.1.6 Combining Different Trend Specifications

A common situation in applied work is that some model variables are trending, whereas others are not. Consider, for example, the problem of constructing a VAR model for

$$y_t = \begin{pmatrix} gnp_t \\ \pi_t \\ i_t \end{pmatrix},$$

where gnp_t denotes the log level of real GNP, π_t the inflation rate, and i_t the nominal interest rate. Inspection shows that gnp_t is subject to a linear trend, whereas π_t and i_t are not. One common choice in this situation is to replace gnp_t by detrended gnp_t , using linear detrending or a suitable band-pass filter, and postulating a VAR model with intercept for

$$\begin{pmatrix} \widetilde{gnp}_t \\ \pi_t \\ i_t \end{pmatrix},$$

where \widetilde{gnp}_t denotes detrended log real GNP. Alternatively, in the presence of a stochastic trend in gnp_t , we would replace \widetilde{gnp}_t in this model by Δgnp_t .

In the presence of a deterministic time trend in gnp_t , another choice is to include a deterministic time trend in all three equations of a VAR model for the original y_t . This approach obviously is infeasible when working with band-pass filters.

It may seem that a third choice would be to specify a VAR model for y_t that includes both an intercept and a deterministic time trend in each equation with the coefficients on the time trends in equations two and three restricted to

zero in estimation. Such a model could be estimated by restricted GLS or, less efficiently, by equation-by-equation LS. This approach is not valid in general, however, because lags of the trending variable gnp_t appear in equations two and three.

19.2 Seasonality

Seasonality in macroeconomic data refers to systematic and recurrent variation in the data within the year. For example, automobile output in Europe tends to drop every summer, as workers go on vacation and plants all but close down. Likewise, construction activity in the Northern hemisphere slows every winter, while the consumption of natural gas surges. Similarly, monthly retail sales spike every year in December, as Christmas approaches, and airfares and motel rates increase during tourist season. Unmodeled seasonality tends to violate the constant parameter assumption of standard linear VAR models. There are a number of remedies depending on the type of seasonal variation in the data. Typically, the objective is to construct the VAR impulse responses controlling for seasonality, which facilitates comparisons with responses in economic models that do not contain seasonality.

19.2.1 Deterministic Seasonal Variation in VAR Models

Just as we have distinguished between deterministic and stochastic trend models, it is necessary to distinguish between deterministic seasonal variation and stochastic seasonal variation. The most common form of deterministic seasonal variation involves adding seasonal dummies to the VAR model. For example, consider the stable autoregression

$$A(L)y_t = v_i + u_t,$$

where v_i is the intercept associated with the i^{th} season and u_t is white noise. Quarterly models would typically include four seasonal dummies, and monthly models would include twelve seasonal dummies, where each dummy variable takes on a value of 1 for the quarter (or month) in question and zero otherwise.¹ While it is common to include one dummy for each quarter (or month), one dummy would obviously be enough if we knew that there is no seasonality in the other quarters (or months). Models with seasonal dummies impart a deterministic pattern. Seasonality is perfectly predictable and, every year, the seasonal effect is the same. Although seasonal dummies are often used when modeling $I(0)$ processes, it is worth stressing that it is equally possible for an $I(1)$ process in levels to have deterministic seasonals.

¹ If all seasonal dummies are included in the VAR model, the intercept must be dropped to avoid a singularity in the regressor matrix. Alternatively, one may retain the intercept, but drop one of the seasonal dummies.

19.2.2 Stochastic Seasonal Variation in VAR Models

Stationary Seasonal Processes. Alternatively, one can model seasonal effects as random, which allows seasonality to be less than perfectly predictable and permits seasonal effects to evolve over time. Such randomness is appealing because there is no reason to expect seasonality to be time-invariant. For example, one would expect seasonal patterns in airfares to evolve with changes in market structure and seasonal energy consumption patterns to evolve with new technologies. Some forms of stochastic seasonality are easy to model. For example, for a VAR process

$$A_s(L^s)y_t = v + u_t,$$

where $A_s(L^s) = I_K - A_{s,1}L^s - \dots - A_{s,r}L^{r \times s}$ is a matrix operator which involves only lags that are multiples of the seasonal period s , the autocorrelation function of y_t spikes at lags $s, 2s, 3s, \dots, rs$. The seasonal period is for example $s = 4$ for quarterly data and $s = 12$ for monthly data. To allow for more sophisticated nonseasonal autocovariance, a multiplicative model of the form

$$A(L)A_s(L^s)y_t = v + u_t$$

may be considered, where $A(L)$ is a standard VAR operator of order p , say. Multiplying out the operator product shows that a standard VAR process with sufficiently large lag order nests the purely seasonal VAR or multiplicative model, if all roots of the VAR operator are outside the unit circle, but some roots are complex pairs with seasonal periodicities. This means that even a standard VAR model may generate seasonality following a stationary stochastic process, provided we include enough lags. One drawback of such models is that stationary seasonal processes tend not to exhibit the regularity commonly associated with seasonal effects, unless the seasonal roots approach the unit circle. This observation suggests that this seasonal model without the addition of seasonal dummies will be of limited relevance for applied work.

Integrated Seasonal Processes. Further complications arise when the seasonal variation is known to follow unit root processes. In that case, the appropriate transformation of the data is the seasonal-differencing operator:

$$\Delta_s \equiv (1 - L^s),$$

where $s = 4$ for quarterly data and $s = 12$ for monthly data. Rather than expressing the data as quarter-to-quarter growth rates, as one would in non-seasonal quarterly VAR models, applying the seasonal differencing operator to series in logs, one constructs the growth rate for the current quarter or month relative to the same quarter or month one year ago. These transformed data are often referred to as year-on-year growth rates in the press and in government

reports. Having transformed the relevant data in this fashion, one proceeds by fitting a standard VAR model with intercept to seasonally differenced data or untransformed data, as appropriate.

Once we allow for a seasonal unit root, an obvious concern is that the seasonal components are likely to be cointegrated across at least some variables in the VAR model. Seasonal cointegration means that shocks in this VAR model have a permanent effect on the seasonal pattern in each variable. In that case, LS estimates of the VAR model for the seasonally differenced data would not enforce the long-run properties of the variables because there is no seasonal error correction term. One way of addressing this problem would be to specify a seasonally cointegrated VAR model. There are different possible specifications, depending on the assumptions regarding the roots on the unit circle and the deterministic terms. For example, Lee (1992) specifies a quarterly model of the form

$$\Delta_4 y_t = \nu + \sum_{i=1}^4 \Pi_i z_{it-1} + \sum_{j=1}^{p-4} \Gamma_j \Delta_4 y_{t-j} + u_t,$$

where

$$\begin{aligned} z_{1t} &= (1 + L + L^2 + L^3)y_t, \\ z_{2t} &= (1 - L + L^2 - L^3)y_t, \\ z_{3t} &= L(1 - L^2)y_t, \\ z_{4t} &= (1 - L^2)y_t \end{aligned}$$

capture roots on the unit circle of different frequencies (see also Johansen and Schaumburg 1999). A corresponding model with seasonal intercepts is discussed by Franses and Kunst (1999). A brief discussion of seasonal cointegration can also be found in Ghysels and Osborn (2001, chapter 3).

In practice, the existence of seasonal unit roots and seasonal cointegration is rarely, if ever, known, and standard statistical tests lack the ability to discriminate reliably between different seasonal models. This creates a problem because imposing incorrect seasonal structure in estimation will invalidate the VAR model estimates. A convenient alternative is to specify the VAR model allowing for at least 12 monthly lags (or at least four quarterly lags) in levels fit to seasonally unadjusted data in levels with deterministic seasonals and possibly deterministic trends included in the VAR model (see Diebold 1993). Apart from the loss of efficiency and the small-sample bias associated with the addition of redundant deterministic regressors, the LS estimates of that model will remain consistent whether the data exhibit any combination of a deterministic trend, a stochastic trend, cointegration, deterministic seasonality, stochastic seasonality, seasonal integration, or seasonal cointegration, mirroring the discussion in Chapter 3. No knowledge of the nature of the seasonal process

is required as long as the seasonal DGP is nested within the VAR model. Of course, such a VAR model may be simplified at the expense of generality to the extent that we have additional information about the data that can be imposed in estimation.

19.2.3 Synthesis

When dealing with seasonal $I(0)$ data, the standard remedy is to allow for a sufficient number of autoregressive lags to capture stochastic seasonal dynamics combined with seasonal dummies to capture deterministic seasonal dynamics. When dealing with seasonal $I(1)$ data, there are two main modeling choices. One is seasonal differencing of the data; the other is imposing deterministic seasonality on the $I(1)$ process. Osborn (1993) stresses that seasonal unit roots imply a failure of cointegration between the four quarters of a given time series. This means that observations for different quarters in a given year may move arbitrarily far apart. This implication of seasonal differencing is *a priori* implausible and difficult to rationalize from an economic point of view. Consequently, Osborn (1993) recommends against using models of seasonal differences in favor of models that superimpose a deterministic seasonal pattern on a conventional $I(1)$ process. The latter models also seem to provide a better description of typical macroeconomic variables than models based on seasonal differences, although it is unclear whether existing statistical procedures can discriminate between these models. In contrast, Osborn (2002) concludes that the default model for forecasting seasonal economic time series for horizons of one or two years should be one with year-on-year differences. Although a good forecasting model is not necessarily the best model for other purposes, Osborn's analysis shows that the choice between using year-on-year growth rates or deterministic seasonality models for series with strong seasonality is not straightforward and has to be decided case by case.

19.2.4 Periodic Seasonal VAR Models

Both the seasonally differenced model and the model with seasonal dummies impose that the slope parameters of the model and its variance-covariance structure are time invariant across seasons. Some researchers have argued that a more general model would allow all model parameters to vary with the season (see, e.g., Osborn 1991, 1993). The periodic VAR model may be viewed as a generalization of the VAR model with seasonal dummies. That model may alternatively be written as:

$$y_t = n_{1t}v_1 + \cdots + n_{st}v_s + A_1y_{t-1} + \cdots + A_py_{t-p} + u_t,$$

where $n_{it} \in \{0, 1\}$ and $\sum_{i=1}^s n_{it} = 1$. In other words, $n_{it} = 1$ if t coincides with the i^{th} season, and $n_{it} = 0$ otherwise. Now consider relaxing the restriction that

only the intercept varies by season. This results in the general periodic VAR model:

$$y_t = v_t + A_{1t}y_{t-1} + \cdots + A_{pt}y_{t-p} + u_t = A_t Z_{t-1} + u_t$$

with $Z_{t-1} \equiv (1, y'_{t-1}, \dots, y'_{t-p})'$,

$$A_t \equiv [v_t, A_{1t}, \dots, A_{pt}] = n_{1t}A_1 + \cdots + n_{st}A_s,$$

and

$$\Sigma_t = n_{1t}\Sigma_1 + \cdots + n_{st}\Sigma_s,$$

where the n_{it} again denote seasonal dummy variables. Periodic VAR models can be motivated based on economic models in which seasonality is an integral part of agents' decision-making. A detailed discussion of periodic VAR models can be found in section 17.3 of Lütkepohl (2005). Extensions of this model to allow for integrated and cointegrated variables are possible (see Ghysels and Osborn 2001, chapter 6; Franses and Paap 2004).

19.2.5 Seasonal TVC-VAR Models

Periodic models imply deterministic variation in the model parameters. An even less restrictive, periodic VAR model would allow the intercept terms to evolve stochastically according to a latent random walk process:

$$y_t = v_{s,t} + A_{1t}y_{t-1} + \cdots + A_{pt}y_{t-p} + u_t,$$

where

$$v_{s,t} = v_{s,t-1} + \eta_{s,t}$$

and $\eta_{s,t}$ is zero-mean white noise. From an econometric point of view, this model belongs to the class of time-varying coefficient vector autoregressive (TVC-VAR) models discussed in Chapter 18. Periodic VAR models and seasonal TVC-VAR models are rarely used in empirical work, perhaps because they are more difficult to estimate and evaluate than standard VAR models.

19.2.6 Seasonally Filtered Data in VAR Models

Many macroeconomic time series are reported in seasonally adjusted form, making allowance for seasonality in the model unnecessary.² Most empirical

² It is nevertheless good practice to verify that ostensibly seasonally adjusted data are indeed free of seasonality. Often a plot helps verify the absence or presence of seasonality. Alternatively, plots of the spectral density of the (suitably transformed) data or of their autocorrelations can be helpful. As an additional check, one could regress these data on seasonal dummies and conduct a Wald test for the inclusion of these regressors.

studies are based on such seasonally adjusted data. Statistical agencies reporting seasonally adjusted data tend to rely on procedures such as the Census Bureau's X-12-ARIMA filter or Eurostat's TRAMO-SEATS to remove seasonal effects (see Findley, Monsell, Bell, Otto, and Chen 1998; Gómez and Maravall 1997). At the core of the X-12 procedure (and its X-13 successor) is a set of moving average filters which allow the seasonal pattern to evolve over time. The seasonal filter is typically applied to one time series at a time. Notwithstanding the increasing sophistication of some of these seasonal filters, it remains true that all seasonal adjustment procedures – including those implemented by statistical agencies – have side-effects and may distort the data in ways that can change the economic interpretation of the data or the statistical significance of parameter estimates. Applied researchers have tended to ignore this concern, given the convenience of using seasonally adjusted data, which reduces the number of modeling choices to be made and eliminates the need for modifications of the basic VAR model along the lines discussed earlier.

19.2.7 Combining Seasonally Adjusted and Unadjusted Data in the Same VAR Model

Although typically users of VAR models rely on seasonally adjusted data (or on data not subject to any apparent seasonality such as interest rates), occasionally VAR models include seasonal variables that are not available in seasonally adjusted form. In addition, sometimes data have been labelled incorrectly as *seasonally adjusted* by statistical agencies yet still contain visible seasonality, making further adjustments necessary. The most common response to this problem is to allow for a sufficient number of autoregressive lags and/or to include seasonal dummies in all equations of the VAR model. Alternatively, the data in question can be deseasonalized up front to avoid the overparameterization of the VAR model when not all variables contain a seasonal component. In the latter case, the use of the X-12-ARIMA filter is another option. There is no consensus on which approach works better.

When combining variables requiring different seasonal adjustments within the same VAR model, the same caveats apply as when combining different trend specifications. When dealing with deterministic seasonality, we either must deseasonalize the data upfront, before fitting a VAR model with intercept, or we must include the same deterministic seasonal components in each equation of the VAR model along with any other deterministic terms.

19.2.8 Summary

In structural modeling we are typically not so much interested in modeling seasonality. In that situation, if seasonal data are available, there are two options. First, we may remove the seasonal component from the data by applying some

seasonal-adjustment procedure. One drawback of such procedures is that they may lead to distortions of the nonseasonal component. Another drawback is that these procedures are applied to individual series and, hence, cannot take into account the possible interaction between the seasonal components of different series.

The second possibility is to allow for seasonality within the VAR model. We discussed several ways of modeling seasonality and their potential drawbacks. For example, adding deterministic terms to account for seasonality captures only very regular types of seasonality. Incorporating seasonal differences in the model allows for stochastic seasonality, but may have implausible implications for the long-run features of the seasonal components. Yet another possibility is to specify VAR models that encompass deterministic and stochastic seasonality. Such models tend to be very large, however, which limits their usefulness for macroeconometric modeling. Thus, it is difficult to provide general guidelines for coping with seasonality in some or all variables.

Ultimately we are interested in modeling seasonality because of its implications for structural impulse response analysis. As long as modeling the seasonal component leaves the additive reduced-form error term unaffected, the identification of the structural shocks may proceed exactly as discussed in earlier chapters. To the extent that there is seasonal heteroskedasticity, standard identification methods for time-invariant VAR models fail, but we may proceed with identification by heteroskedasticity, as discussed in Chapter 14.

Provided the seasonality leaves the autoregressive coefficients unaffected, as would be the case if the seasonality can be captured by seasonal dummies, by seasonal differencing, or by adding autoregressive lags, the construction of the structural impulse responses may proceed as in the time-invariant model. Otherwise, the methods discussed in Chapter 18 may be used.

19.3 Structural Change in the Stochastic Component of the VAR Model

The premise of linear VAR models is that the model structure is time invariant. Few researchers would contend that this premise is literally correct. There are many reasons why we would expect the stochastic process governing the economy to evolve over time. A more defensible position is that the VAR model serves as a local approximation to the data generating process within a given time window.

19.3.1 Breaks in the Stochastic Component

In some cases, there are compelling reasons to suspect a discrete break in the stochastic structure. For example, we know of important changes in the U.S. monetary policy regime in 1979, when Paul Volcker assigned much higher

weight to the inflation objective and lower weight to the employment objective than his predecessors. Fitting the same VAR model of U.S. monetary policy to data from the pre- and post-Volcker era hence does not seem appropriate.

One obvious remedy is to split the sample in 1979 and to fit separate vector autoregressions to each subsample (see, e.g., Kilian and Lewis 2011). This means that, in general, users of VAR models by necessity will have to make do with fairly short samples, requiring special attention to the small-sample properties of estimators. Another challenge is that, following major structural changes the data may be contaminated by transition dynamics for years to come. For example, U.S. inflation and interest rates fell only slowly following the monetary policy shift under Paul Volcker in 1979, and remained persistently high well into the 1980s. As a consequence, it may make sense to exclude the early 1980s from the analysis, when splitting the sample, and to start a VAR analysis of the post-1979 U.S. economy only around 1985. Allowing for such transition dynamics further reduces the available sample. In light of these restrictions, a common situation in applied work is that we are unable to fit a sufficient number of lags on the subsample. In that situation a VAR analysis may not be an option.

An alternative approach is to estimate VAR models based on rolling windows of data. This approach is less appropriate when dealing with well-defined breaks in the model structure, as in the Volcker example, and more appropriate when dealing with smooth structural changes. An obvious caveat is that macroeconomists rarely have enough data for rolling windows that are long enough to allow reliable estimation. Hence, VAR estimates from rolling regressions tend to be noisy and unreliable. An added concern is that impulse responses computed from rolling regressions make no allowance for future changes in the structure of the model. A caveat that applies to both split-sample regressions and rolling regression applications of VAR models is that care must be taken that the sample period in question includes enough variation in the data to allow the identification of the structural shocks in question. Moreover, in reporting the impulse response estimates, it is important to normalize the shocks of interest to be of the same magnitude across all samples.

A third approach would be to specify a TVC-VAR model that allows all model parameters to evolve smoothly over time, as discussed in Chapter 18. An example of this strategy is Primiceri (2005). Such models are much harder to estimate reliably in practice given the profligacy of parameters, especially as the number of variables increases. TVC-VAR models require long samples and quickly become intractable when longer lags and/or additional variables are included in the model. These problems are compounded when dealing with structural representations of TVC-VAR models. For example, such models can allow for time variation in the parameters within a given monetary policy rule, but it would be much harder to allow for a change in the monetary policy rule from targeting monetary aggregates toward targeting interest rates,

as discussed in Kozicki and Tinsley (2009). Moreover, TVC-VAR models are nonlinear models, which complicates the interpretation of the evolution of the impulse response functions over time and makes it impossible to construct historical decompositions or conventional forecast error variance decompositions. It also restricts the range of strategies that can be used for identifying structural shocks.

An alternative class of nonlinear models is the regime-switching VAR model we already discussed in Chapter 18 (see, e.g., Sims, Waggoner, and Zha 2008). Such models have been used to model the evolution of monetary policy, in particular (see Sims and Zha 2006b). Regime switching models with alternating regimes evolving stochastically according to a latent Markov-switching process allow for changes in the stochastic component of the VAR much like TVC-VAR models, but imply a greater degree of regularity in that the model alternates between a small number of policy regimes. Whether this degree of regularity is found in the real world is an open question. An alternative view would be that monetary policy regimes change over time without any apparent tendency to revert back to previous regimes. In that case, the use of a regime-switching model would be questionable.

For these reasons, traditional linear structural VAR models estimated on samples chosen to avoid structural breaks will remain one of the primary tools of analysis. An obvious question is how to detect these breaks. Although there are Chow-type tests for structural change that could be used to detect likely breakpoints in the slope parameters of the VAR model, as discussed in Chapter 2, these tests tend to have low power especially after accounting for data mining across possible breakpoints. This problem is compounded when the number of breaks is unknown or the break does not occur all at once. Moreover, tests for structural change tend to confound persistent transitory dynamics arising from unusual sequences of structural VAR shocks with structural change.

Hence, a better strategy in practice is to base the selection of the sample primarily on extraneous information about economic institutions. An example of this strategy is provided in Alquist, Kilian, and Vigfusson (2013) who discuss how VAR models of the relationship between the price of oil and U.S. real GDP may be undermined by the inclusion of pre-1973 data, given the changes in the institutional structure of the oil market in late 1973.

19.3.2 Smooth Structural Change in the Stochastic Component

Perhaps the most pernicious form of structural change in practice is smooth structural change. In many cases, we can view linear VAR models as an approximation to a process that is evolving slowly and gradually over time, as long as the sample is not too long. In some cases, such approximations are not credible. For example, in modeling the effects of energy price shocks on the growth

rate of U.S. consumption, economic theory tells us that the magnitude of the effect should depend on how large energy expenditures are as a fraction of total expenditures. This energy share has fluctuated considerably since the 1970s. One approach to modeling this form of nonlinearity would be to specify a TVC-VAR model as in Chapter 18. Alternatively, we can convert this nonlinear model into a linear model by weighting the percent change in energy prices by the nominal expenditure share, as proposed by Edelstein and Kilian (2009). This allows us to rely on standard linear VAR models.

A similar approach was taken by Ramey and Vine (2011) in modeling the nonlinearities arising from the regulation of U.S. gasoline prices in the 1970s and early 1980s. Price ceilings are costly to consumers because they create lines at gas stations. These queuing costs are part of the price of gasoline. If they are ignored, the relationship between gasoline prices and the economy may become nonlinear. Ramey and Vine addressed this problem by quantifying the queuing costs and adding them to the observed gasoline price, allowing them to fit a linear VAR model.

The particular adjustments proposed by Edelstein and Kilian (2009) and Ramey and Vine (2011) could be combined, of course (see, e.g., Baumeister and Kilian 2017). Whether the linear VAR framework can be preserved in this fashion depends on the economic context and on the ingenuity of the researcher. If no suitable transformations are available, the only alternative would be a TVC-VAR model or some other model that allows for changing coefficients.

Bibliography

- Abraham, B. (1982): “Temporal aggregation and time series,” *International Statistical Review*, 50, 285–291.
- Abraham, K. G., and J. C. Haltiwanger (1995): “Real wages and the business cycle,” *Journal of Economic Literature*, 33, 1215–1264.
- Ahn, S. K. (1988): “Distribution of residual autocovariances in multivariate autoregressive models with structured parameterization,” *Biometrika*, 75, 590–593.
- Ahn, S. K., and G. C. Reinsel (1988): “Nested reduced-rank autoregressive models for multiple time series,” *Journal of the American Statistical Association*, 83, 849–856.
- (1990): “Estimation of partially nonstationary multivariate autoregressive models,” *Journal of the American Statistical Association*, 85, 813–823.
- Akaike, H. (1973): “Information theory and an extension of the maximum likelihood principle,” in *2nd International Symposium on Information Theory*, ed. by B. Petrov, and F. Csáki, pp. 267–281. Académiai Kiadó, Budapest.
- (1974): “A new look at the statistical model identification,” *IEEE Transactions on Automatic Control*, AC-19, 716–723.
- Alesina, A., C. Favero, and F. Giavazzi (2015): “The output effect of fiscal consolidations,” *Journal of International Economics*, 96, 19–42.
- Alessi, L., M. Barigozzi, and M. Capasso (2011): “Non-fundamentalness in structural econometric models: A review,” *International Statistical Review*, 79, 16–47.
- Alquist, R., and L. Kilian (2010): “What do we learn from the price of crude oil futures?” *Journal of Applied Econometrics*, 25, 539–573.
- Alquist, R., L. Kilian, and R. J. Vigfusson (2013): “Forecasting the price of oil,” in *Handbook of Economic Forecasting*, vol. 2, ed. by G. Elliott and A. Timmermann, pp. 427–507. Elsevier, Amsterdam.
- Amemiya, T., and R. Y. Wu (1972): “The effect of aggregation on prediction in the autoregressive model,” *Journal of the American Statistical Association*, 67, 628–632.
- Amengual, D., and M. W. Watson (2007): “Consistent estimation of the number of dynamic factors in a large N and T panel,” *Journal of Business and Economic Statistics*, 25, 91–96.

- Amir Ahmadi, P., and H. Uhlig (2009): “Measuring the dynamic effects of monetary policy shocks: A Bayesian FAVAR approach with sign restriction,” working paper, Humboldt University.
- Amisano, G., and C. Giannini (1997): *Topics in Structural VAR Econometrics*, 2nd ed. Springer-Verlag, Berlin.
- An, S., and F. Schorfheide (2007): “Bayesian analysis of DSGE models,” *Econometric Reviews*, 26, 113–172.
- Andersen, T. G., T. Bollerslev, F. X. Diebold, and C. Vega (2003): “Micro effects of macro announcements: Real-time price discovery in foreign exchange,” *American Economic Review*, 93, 38–62.
- Anderson, B. D. O., and J. B. Moore (1979): *Optimal Filtering*. Prentice-Hall, Englewood Cliffs.
- Anderson, S. T., R. Kellogg, and S. W. Salant (2016): “Hotelling under pressure,” *Journal of Political Economy*, forthcoming.
- Anderson, T. W. (1951): “Estimating linear restrictions on regression coefficients for multivariate normal distributions,” *Annals of Mathematical Statistics*, 22, 327–351.
- (1999): “Asymptotic distribution of the reduced rank regression estimator under general conditions,” *Annals of Statistics*, 27, 1141–1154.
- (2002): “Canonical correlation analysis and reduced rank regression in autoregressive models,” *Annals of Statistics*, 30, 1134–1154.
- (2003): *An Introduction to Multivariate Statistical Analysis*, 3rd ed. John Wiley, New York.
- Andrews, D. W. K. (1987): “Asymptotic results for generalized Wald tests,” *Econometric Theory*, 3, 348–358.
- (1991): “Heteroskedasticity and autocorrelation consistent covariance matrix estimation,” *Econometrica*, 59, 817–858.
- (1993): “Tests for parameter instability and structural change with unknown change point,” *Econometrica*, 61, 821–856.
- Andrews, D. W. K., and X. Cheng (2012): “Estimation and inference with weak, semi-strong, and strong identification,” *Econometrica*, 80, 2153–2211.
- Andrews, D. W. K., and W. Ploberger (1994): “Optimal tests when a nuisance parameter is present only under the alternative,” *Econometrica*, 62, 1383–1414.
- Andrews, I., and A. Mikusheva (2015): “Maximum likelihood inference in weakly identified dynamic stochastic general equilibrium models,” *Quantitative Economics*, 6, 123–152.
- Anselin, L. (2006): “Spatial econometrics,” in *Palgrave Handbook of Econometrics, Econometric Theory*, ed. by T. C. Mills and K. Patterson, vol. 1, pp. 901–969. Palgrave Macmillan, Basingstoke.
- Antolín-Díaz, J., and J. Rubio-Ramírez (2016): “Narrative sign restrictions for SVARs,” working paper, Emory University.
- Aoki, M. (1987): *State Space Modeling of Time Series*. Springer-Verlag, Berlin.
- Arezki, R., V. A. Ramey, and L. Sheng (2017): “News shocks in open economies: Evidence from giant oil discoveries,” *Quarterly Journal of Economics*, 132, 103–155.

- Arias, J. E., J. F. Rubio-Ramírez, and D. F. Waggoner (2013): “Inference based on SVARs identified with sign and zero restrictions: Theory and applications,” working paper, Federal Reserve Board.
- (2015): “Inference based on SVARs identified with sign and zero restrictions: Theory and applications,” working paper, Federal Reserve Board.
- Aruoba, S. B., F. X. Diebold, J. Nalewaik, F. Schorfheide, and D. Song (2016): “Improving GDP measurement: A measurement-error perspective,” *Journal of Econometrics*, 191, 384–397.
- Atanasova, C. (2003): “Credit market imperfections and business cycle dynamics: A nonlinear approach,” *Studies in Nonlinear Dynamics & Econometrics*, 7 (4), Article 5.
- Auerbach, A. J., and Y. Gorodnichenko (2012): “Measuring the output responses to fiscal policy,” *American Economic Journal: Economic Policy*, 4, 1–27.
- Bacchicocchi, E., E. Castelnuovo, and L. Fanelli (2013): “Gimme a break! Identification and estimation of the macroeconomic effects of monetary policy shocks in the U.S.,” working paper, University of Milan.
- Bacchicocchi, E., and L. Fanelli (2015): “Identification in structural vector autoregressive models with structural changes, with an application to US monetary policy,” *Oxford Bulletin of Economics and Statistics*, 77, 761–779.
- Bai, J. (2003): “Inferential theory for factor models of large dimensions,” *Econometrica*, 71, 135–171.
- (2004): “Estimating cross-section common stochastic trends in nonstationary panel data,” *Journal of Econometrics*, 122, 137–183.
- Bai, J., and K. Li (2012): “Maximum likelihood estimation and inference for approximate factor models of high dimension,” Discussion Paper No. 42099, MPRRA.
- Bai, J., K. Li, and L. Lu (2016): “Estimation and inference of FAVAR models,” *Journal of Business and Economic Statistics*, 34, 620–641.
- Bai, J., and S. Ng (2002): “Determining the number of factors in approximate factor models,” *Econometrica*, 70, 191–221.
- (2006): “Confidence intervals for diffusion index forecasts and inference for factor augmented regressions,” *Econometrica*, 74, 1133–1150.
- (2007): “Determining the number of primitive shocks in factor models,” *Journal of Business and Economic Statistics*, 25, 52–60.
- (2008): “Large dimensional factor analysis,” *Foundations and Trends in Econometrics*, 3, 89–163.
- (2013): “Principal components estimation and identification of static factors,” *Journal of Econometrics*, 176, 18–29.
- Baillie, R. T. (1979): “Asymptotic prediction mean squared error for vector autoregressive models,” *Biometrika*, 66, 675–678.
- Balabanova, Z., and R. Brüggemann (2017): “External information and monetary policy transmission in new EU member states: Results from FAVAR models,” *Macroeconomic Dynamics*, 21, 311–335.
- Balke, N. S. (2000): “Credit and economic activity: Credit regimes and nonlinear propagation of shocks,” *Review of Economics and Statistics*, 82, 344–349.

- Balke, N. S., and T. B. Fomby (1997): "Threshold cointegration," *International Economic Review*, 38, 627–645.
- Banbura, M., D. Giannone, and L. Reichlin (2010): "Large Bayesian vector autoregressions," *Journal of Applied Econometrics*, 25, 71–92.
- Banerjee, A., and M. Marcellino (2008): "Factor-augmented error correction models," Working Paper ECO No. 2008/15, European University Institute.
- Banerjee, A., M. Marcellino, and I. Masten (2014a): "Forecasting with factor-augmented error correction models," *International Journal of Forecasting*, 30, 589–612.
- (2014b): "Structural FECM: Cointegration in large-scale structural FAVAR models," CEPR Discussion Paper No. 9858.
- Barhoumi, K., O. Darné, and L. Ferrara (2014): "Dynamic factor models: A review of the literature," *OECD Journal: Journal of Business Cycle Measurement and Analysis*, 2013/2, 73–107.
- Barigozzi, M., M. Lippi, and M. Luciani (2016): "Non-stationary dynamic factor models for large data sets," working paper, London School of Economics.
- Barnichon, R., and C. Matthes (2016): "Gaussian mixture approximations of impulse responses and the non-linear effects of monetary shocks," working paper, CREI, Universitat Pompeu Fabra.
- Barsky, R. B., and L. Kilian (2001): "Do we really know that oil caused the Great Stagflation? A monetary alternative," *NBER Macroeconomics Annual*, 16, 137–183.
- (2004): "Oil and the macroeconomy since the 1970s," *Journal of Economic Perspectives*, 18, 115–134.
- Barsky, R. B., and E. Sims (2011): "News shocks and business cycles," *Journal of Monetary Economics*, 58, 273–289.
- Basawa, I. V., A. K. Mallik, W. P. Cormick, J. H. Reeves, and R. L. Taylor (1991): "Bootstrapping unstable first-order autoregressive processes," *Annals of Statistics*, 19, 1098–1101.
- Bastianin, A., and M. Manera (2017): "How does stock market volatility react to oil shocks," *Macroeconomic Dynamics*, forthcoming.
- Basu, S., J. G. Fernald, and M. S. Kimball (2006): "Are technology improvements contractionary?" *American Economic Review*, 96, 1418–1448.
- Bauer, D., and A. Maynard (2012): "Persistence-robust surplus-lag Granger causality testing," *Journal of Econometrics*, 169, 293–300.
- Baumeister, C., and L. Benati (2013): "Unconventional monetary policy and the Great Recession: Estimating the macroeconomic effects of a spread compression at the zero lower bound," *International Journal of Central Banking*, 9, 165–212.
- Baumeister, C., and J. D. Hamilton (2015a): "Sign restrictions, structural vector autoregressions, and useful prior information," *Econometrica*, 83, 1963–1999.
- (2015b): "Optimal inference about impulse-response functions and historical decompositions in incompletely identified vector autoregressions," working paper, University of California, San Diego.
- (2015c): "Structural interpretation of vector autoregressions with incomplete identification: Revisiting the role of oil supply and demand shocks," working paper, University of California, San Diego.

- Baumeister, C., and L. Kilian (2012): "Real-time forecasts of the real price of oil," *Journal of Business and Economic Statistics*, 30, 326–336.
- (2014): "Real-time analysis of oil price risks using forecast scenarios," *IMF Economic Review*, 63, 119–145.
- (2016a): "Forty years of oil price fluctuations: Why the price of oil may still surprise us," *Journal of Economic Perspectives*, 30, 131–158.
- (2016b): "Understanding the decline in the price of oil since June 2014," *Journal of the Association of Environmental and Resource Economists*, 3, 131–158.
- (2016c): "A general approach to recovering market expectations from futures prices with an application to crude oil," working paper, University of Michigan.
- (2017): "Lower oil prices and the U.S. economy: Is this time different?" *Brookings Papers on Economic Activity*, Fall 2016, 287–336.
- Baumeister, C., and G. Peersman (2013): "The role of time-varying price elasticities in accounting for volatility changes in the crude oil market," *Journal of Applied Econometrics*, 28, 1087–1109.
- Bäurle, G. (2013): "Structural dynamic factor analysis using prior information from macroeconomic theory," *Journal of Business and Economic Statistics*, 31, 136–150.
- Baxter, M., and R. G. King (1999): "Measuring business cycles: Approximate band pass filters for economic time series," *Review of Economics and Statistics*, 81, 575–593.
- Beaudry, P., D. Nam, and J. Wang (2012): "Do mood swings drive business cycles and is it rational?" working paper, University of British Columbia.
- Beaudry, P., and F. Portier (2006): "Stock prices, news, and economic fluctuations," *American Economic Review*, 96, 1293–1307.
- (2014): "News driven business cycles: Insights and challenges," *Journal of Economic Literature*, 52, 993–1074.
- Benati, L., and T. A. Lubik (2014): "Sales, inventories, and real interest rates: A century of stylized facts," *Journal of Applied Econometrics*, 29, 1210–1222.
- Benati, L., and H. Mumtaz (2007): "U.S. evolving macroeconomic dynamics: A structural investigation," Working Paper No. 746, European Central Bank.
- Benkowitz, A., H. Lütkepohl, and M. H. Neumann (2000): "Problems related to confidence intervals for impulse responses of autoregressive processes," *Econometric Reviews*, 19, 69–103.
- Berkowitz, J., I. Birgean, and L. Kilian (2000): "On the finite-sample accuracy of non-parametric resampling algorithms for economic time series," *Advances in Econometrics*, 14, 77–105.
- Berkowitz, J., and L. Kilian (2000): "Recent developments in bootstrapping time series," *Econometric Reviews*, 19, 1–48.
- Bernanke, B. S. (1983): "Irreversibility, uncertainty, and cyclical investment," *Quarterly Journal of Economics*, 98, 85–106.
- (1986): "Alternative explanations of the money-income correlation," *Carnegie-Rochester Conference Series on Public Policy*, 25, 49–99.
- Bernanke, B. S., and A. Blinder (1992): "The federal funds rate and the channels of monetary transmission," *American Economic Review*, 82, 901–921.
- Bernanke, B. S., and J. Boivin (2003): "Monetary policy in a data-rich environment," *Journal of Monetary Economics*, 50, 525–546.

- Bernanke, B. S., J. Boivin, and P. Eliasz (2005): "Measuring monetary policy: A factor augmented vector autoregressive (FAVAR) approach," *Quarterly Journal of Economics*, 120, 387–422.
- Bernanke, B. S., M. Gertler, and M. W. Watson (1997): "Systematic monetary policy and the effects of oil price shocks," *Brookings Papers on Economic Activity*, 1, 91–142.
- Bernanke, B. S., and I. Mihov (1995): "Measuring monetary policy," NBER Working Paper No. 5145.
- (1998a): "The liquidity effect and long-run neutrality," *Carnegie-Rochester Conference Series on Public Policy*, 49, 149–194.
- (1998b): "Measuring monetary policy," *Quarterly Journal of Economics*, 113, 869–902.
- Bessler, D. A., and S. Lee (2002): "Money and prices: U.S. data 1869–1914 (A study with directed graphs)," *Empirical Economics*, 27, 427–446.
- Beveridge, S., and C. R. Nelson (1981): "A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle,'" *Journal of Monetary Economics*, 7, 151–174.
- Binning, A. (2013): "Underidentified SVAR models: A framework for combining short-run and long-run restrictions with sign restrictions," working paper, Norges Bank.
- Binswanger, M. (2004): "How do stock prices respond to fundamental shocks?" *Finance Research Letters*, 1, 90–99.
- Bjørnland, H. C., and K. Leitmo (2009): "Identifying the interdependence between US monetary policy and the stock market," *Journal of Monetary Economics*, 56, 275–282.
- Blanchard, O. J. (1989): "A traditional interpretation of macroeconomic fluctuations," *American Economic Review*, 79, 1146–1164.
- (2002): "Comments on 'Do We Really Know that Oil Caused the Great Stagflation? A Monetary Alternative' by Robert Barsky and Lutz Kilian," in *NBER Macroeconomics Annual*, ed. by B. Bernanke and K. Rogoff, pp. 183–192. MIT Press, Cambridge, MA.
- Blanchard, O. J., and R. Perotti (2002): "An empirical investigation of the dynamic effects of shocks to government spending and taxes on output," *Quarterly Journal of Economics*, 117, 1329–1368.
- Blanchard, O. J., and D. Quah (1989): "The dynamic effects of aggregate demand and supply disturbances," *American Economic Review*, 79, 654–673.
- (1993): "The dynamic effects of aggregate demand and supply disturbances: Reply," *American Economic Review*, 83, 653–658.
- Blanchard, O. J., and M. W. Watson (1986): "Are business cycles all alike?" in *The American Business Cycle*, ed. by R. Gordon, pp. 123–179. NBER and University of Chicago Press, Chicago.
- Boivin, J., and M. P. Giannoni (2009): "Global forces and monetary policy effectiveness," in *International Dimensions of Monetary Policy*, ed. by J. Galí and M. Gertler, chap. 8, pp. 429–478. University of Chicago Press, Chicago.
- Boivin, J., M. P. Giannoni, and I. Mihov (2009): "Sticky prices and monetary policy: Evidence from disaggregated US data," *American Economic Review*, 99, 350–384.

- Boivin, J., and S. Ng (2005): "Understanding and comparing factor-based forecasts," *International Journal of Central Banking*, 1, 117–151.
- (2006): "Are more data always better for factor analysis?" *Journal of Econometrics*, 132, 169–194.
- Bordo, M. D., and T. F. Helbling (2010): "International business cycle synchronization in historical perspective," NBER Working Paper No. 16103.
- Boschen, J. F., and L. O. Mills (1995): "The relation between money market and narrative indicators of monetary policy," *Economic Inquiry*, 23, 24–45.
- Bose, A. (1988): "Edgeworth correction by bootstrap in autoregressions," *Annals of Statistics*, 16, 1709–1722.
- Boswijk, H. P., and J. A. Doornik (2004): "Identifying, estimating and testing restricted cointegrated systems: An overview," *Statistica Neerlandica*, 58, 440–465.
- Bouakez, H., F. Chihi, and M. Normandin (2014): "Measuring the effects of fiscal policy," *Journal of Economic Dynamics and Control*, 47, 123–151.
- Bouakez, H., B. Essid, and M. Normandin (2013): "Stock returns and monetary policy: Are there any ties?" *Journal of Macroeconomics*, 36, 33–50.
- Bouakez, H., and M. Normandin (2010): "Fluctuations in the foreign exchange market: How important are monetary policy shocks?" *Journal of International Economics*, 81, 139–153.
- Braun, P. A., and S. Mittnik (1993): "Misspecifications in vector autoregressions and their effects on impulse responses and variance decompositions," *Journal of Econometrics*, 59, 319–341.
- Breidt, F. J., R. A. Davis, K.-S. Lii, and M. Rosenblatt (1991): "Maximum likelihood estimation for noncausal autoregressive processes," *Journal of Multivariate Analysis*, 36, 175–198.
- Breitung, J. (2015): "The analysis of macroeconomic panel data," in *The Oxford Handbook of Panel Data*, ed. by B. Baltagi, chap. 15, pp. 453–492. Oxford University Press, Oxford.
- Breitung, J., and S. Eickmeier (2006): "Dynamic factor models," *Allgemeines Statistisches Archiv*, 90, 27–42.
- (2011): "Testing for structural breaks in dynamic factor models," *Journal of Econometrics*, 163, 71–84.
- Breitung, J., and U. Pigorsch (2013): "A canonical correlation approach for selecting the number of dynamic factors," *Oxford Bulletin of Economics and Statistics*, 75, 23–36.
- Breitung, J., and J. Tenhofen (2011): "GLS estimation of dynamic factor models," *Journal of the American Statistical Association*, 106, 1150–1166.
- Breusch, T. S. (1978): "Testing for autocorrelation in dynamic linear models," *Australian Economic Papers*, 17, 334–355.
- Breusch, T. S., and P. Schmidt (1988): "Alternative forms of the Wald test: How long is a piece of string," *Communications in Statistics, Theory and Methods*, 17, 2789–2795.
- Brewer, K. R. W. (1973): "Some consequences of temporal aggregation and systematic sampling for ARMA and ARMAX models," *Journal of Econometrics*, 1, 133–154.
- Brillinger, D. R. (1975): *Time Series Data Analysis and Theory*. Holt, Rinehart and Winston, New York.

- Brockwell, P. J., and R. A. Davis (1987): *Time Series: Theory and Methods*. Springer-Verlag, New York.
- Brown, R. L., J. Durbin, and J. M. Evans (1975): “Techniques for testing the constancy of regression relationships over time,” *Journal of the Royal Statistical Society B*, 37, 149–192.
- Brüggemann, R., C. Jentsch, and C. Trenkler (2016): “Inference in VARs with conditional heteroskedasticity of unknown form,” *Journal of Econometrics*, 191, 69–85.
- Brüggemann, R., and H. Lütkepohl (2001): “Lag selection in subset VAR models with an application to a U.S. monetary system,” in *Econometric Studies: A Festschrift in Honour of Joachim Frohn*, ed. by R. Friedman, L. Knüppel, and H. Lütkepohl, pp. 107–128. LIT-Verlag, Münster.
- (2005): “Practical problems with reduced-rank ML estimators for cointegration parameters and a simple alternative,” *Oxford Bulletin of Economics and Statistics*, 67, 673–690.
- Brüggemann, R., H. Lütkepohl, and P. Saikkonen (2006): “Residual autocorrelation testing for vector error correction models,” *Journal of Econometrics*, 134, 579–604.
- Caldara, D., and C. Kamps (2017): “The analytics of SVARs: A unified framework to measure fiscal multipliers,” *Review of Economic Studies*, 84, 1015–1040.
- Camacho, M. (2004): “Vector smooth transition regression models for US GDP and the composite index of leading indicators,” *Journal of Forecasting*, 23, 173–196.
- Campbell, J. Y., and R. J. Shiller (1987): “Cointegration and tests of present value models,” *Journal of Political Economy*, 95, 1062–1088.
- Campbell, S. D. (2002): “Specification testing and semiparametric estimation of regime switching models: An examination of the US short term interest rate,” Working Paper No. 2002-26, Department of Economics, Brown University.
- Candelon, B., and H. Lütkepohl (2001): “On the reliability of Chow-type tests for parameter constancy in multivariate dynamic models,” *Economics Letters*, 73, 155–160.
- Caner, M., and X. Han (2014): “Selecting the correct number of factors in approximate factor models: The large panel case with group bridge estimators,” *Journal of Business and Economic Statistics*, 32, 359–374.
- Caner, M., and L. Kilian (2001): “Size distortions of tests of the null hypothesis of stationarity: Evidence and implications for the PPP debate,” *Journal of International Money and Finance*, 20, 639–657.
- Canova, F. (1991): “The sources of financial crisis: Pre- and post-Fed evidence,” *International Economic Review*, 32, 689–713.
- (1993): “Modelling and forecasting exchange rates with a Bayesian time-varying coefficient model,” *Journal of Economic Dynamics and Control*, 17, 233–261.
- (1998): “Detrending and business cycle facts,” *Journal of Monetary Economics*, 41, 475–512.
- (2007): *Methods for Applied Macroeconomic Research*. Princeton University Press, Princeton.
- (2014): “Bridging DSGE models and the raw data,” *Journal of Monetary Economics*, 67, 1–15.

- Canova, F., and M. Ciccarelli (2009): "Estimating multicountry VAR models," *International Economic Review*, 50, 929–959.
- (2013): "Panel vector autoregressive models: A survey," *Advances in Econometrics*, 32, 205–246.
- Canova, F., M. Ciccarelli, and E. Ortega (2012): "Do institutional changes affect business cycles? Evidence from Europe," *Journal of Economic Dynamics and Control*, 36, 1520–1533.
- Canova, F., and G. De Nicolo (2002): "Monetary disturbances matter for business cycle fluctuations in the G7," *Journal of Monetary Economics*, 49, 1131–1159.
- Canova, F., and L. Gambetti (2009): "Structural changes in the US economy: Is there a role for monetary policy?" *Journal of Economic Dynamics and Control*, 33, 477–490.
- Canova, F., and M. Hamidi Sahneh (2016): "Are small-scale SVARs useful for business cycle analysis? Revisiting nonfundamentalness," working paper, BI Norwegian Business School.
- Canova, F., D. López-Salido, and C. Michelacci (2010): "The effects of technology shocks on hours and output: A robustness analysis," *Journal of Applied Econometrics*, 25, 755–773.
- Canova, F., and M. Paustian (2011): "Business cycle measurement with some theory," *Journal of Monetary Economics*, 58, 345–361.
- Canova, F., and F. J. Pérez Forero (2015): "Estimating overidentified, nonrecursive, time-varying coefficients structural vector autoregressions," *Quantitative Economics*, 6, 359–384.
- Canova, F., and L. Sala (2009): "Back to square one: Identification issues in DSGE models," *Journal of Monetary Economics*, 56, 431–449.
- Carrasco, M., L. Hu, and W. Ploberger (2014): "Optimal test for Markov switching parameters," *Econometrica*, 82, 765–784.
- Carriero, A., G. Kapetanios, and M. Marcellino (2009): "Forecasting exchange rates with a large Bayesian VAR," *International Journal of Forecasting*, 25, 400–417.
- (2012): "Forecasting government bond yields with large Bayesian vector autoregressions," *Journal of Banking and Finance*, 36, 2026–2047.
- Cattell, R. B. (1966): "A srree test for the number of factors," *Multivariate Behavioral Research*, 1, 245–276.
- Cavanagh, C. L., G. Elliott, and J. H. Stock (1995): "Inference in models with nearly integrated regressors," *Econometric Theory*, 11, 1131–1147.
- Chamberlain, G., and M. Rothschild (1983): "Arbitrage factor structure, a mean-variance analysis of large asset markets," *Econometrica*, 51, 1281–1304.
- Chan, K. S. (1993): "Consistency and limiting distribution of the least squares estimator of a threshold autoregressive model," *Annals of Statistics*, 21, 520–533.
- Chan, K.-S., and L.-H. Ho (2004): "On the unique representation of non-Gaussian multivariate linear processes," Technical Report No. 341, University of Iowa.
- Chari, V. V., L. J. Christiano, and M. Eichenbaum (1998): "Expectation traps and discretion," *Journal of Economic Theory*, 81, 462–492.
- Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2008): "Are structural models with long-run restrictions useful in developing business cycle theories?" *Journal of Monetary Economics*, 55, 1337–1352.

- Chaudhuri, P. (1996): "On a geometric notion of quantiles for multivariate data," *Journal of the American Statistical Association*, 91, 862–872.
- Chaudhuri, S., and E. Zivot (2011): "A new method of projection-based inference in GMM with weakly identified nuisance parameters," *Journal of Econometrics*, 164, 239–251.
- Chen, B., J. Choi, and C. Escanciano (2017): "Testing for fundamental vector moving average representations," *Quantitative Economics*, 8, 149–180.
- Chen, L., J. J. Dolado, and J. Gonzalo (2014): "Detecting big structural breaks in large factor models," *Journal of Econometrics*, 180, 30–48.
- Cheung, Y.-W., and K. S. Lai (1993): "Finite-sample sizes of Johansen's likelihood ratio tests for cointegration," *Oxford Bulletin of Economics and Statistics*, 55, 313–328.
- Chevillon, G., S. Mavroeidis, and Z. Zhan (2015): "Robust inference in structural VARs with long-run restrictions," working paper, Oxford University.
- Chib, S. (2001): "Markov Chain Monte Carlo methods: Computation and inference," in *Handbook of Econometrics*, ed. by J. J. Heckman and E. E. Leamer, vol. 5, chap. 57, pp. 3569–3649. Elsevier, Amsterdam.
- Chib, S., and E. Greenberg (1995): "Understanding the Metropolis-Hastings algorithm," *American Statistician*, 49, 327–335.
- Choi, I. (2012): "Efficient estimation of factor models," *Econometric Theory*, 28, 274–308.
- Christ, C. F. (1994): "The Cowles Commission's contributions to econometrics at Chicago, 1939–1955," *Journal of Economic Literature*, 32, 30–59.
- Christiano, L. J. (1992): "Searching for a break in GNP," *Journal of Business and Economic Statistics*, 10, 237–250.
- (2002): "Solving dynamic equilibrium models by a method of undetermined coefficients," *Computational Economics*, 20, 21–55.
- Christiano, L. J., M. Eichenbaum, and C. L. Evans (1996): "The effects of monetary policy shocks: Evidence from the flow of funds," *Review of Economics and Statistics*, 78, 16–34.
- (1999): "Monetary policy shocks: What have we learned and to what end?" in *Handbook of Macroeconomics*, ed. by J. Taylor and M. Woodford, vol. 1, pp. 65–148. Elsevier, Amsterdam.
- (2005): "Nominal rigidities and the dynamic effects of a shock to monetary policy," *Journal of Political Economy*, 113, 1–45.
- Christiano, L. J., M. Eichenbaum, and R. J. Vigfusson (2004): "The response of hours to a technology shock: Evidence based on direct measures of technology," *Journal of the European Economic Association*, 2, 381–395.
- (2006a): "Alternative procedures for estimating vector autoregressions identified with long-run restrictions," *Journal of the European Economic Association*, 4, 475–483.
- (2006b): "Assessing structural VARs," *NBER Macroeconomics Annual*, 21, 1–72.
- Christiano, L. J., and T. J. Fitzgerald (2003): "The band pass filter," *International Economic Review*, 44, 435–465.
- Chudik, A., and M. H. Pesaran (2011): "Infinite-dimensional VARs and factor models," *Journal of Econometrics*, 163, 4–22.

- (2016): “Theory and practice of GVAR modeling,” *Journal of Economic Surveys*, 30, 165–197.
- Chung, H., and E. M. Leeper (2007): “What has financed government debt?” NBER Working Paper No. 13425.
- Ciccarelli, M., and A. Rebucci (2003): “Bayesian VARs – A survey of the recent literature,” *Rivista di Politica Economica*, 93(5), 47–112.
- Cochrane, J. H. (1994): “Permanent and transitory components of GNP and stock prices,” *Quarterly Journal of Economics*, 109, 241–265.
- Cochrane, J. H., and M. Piazzesi (2002): “The Fed and interest rates: A high-frequency identification,” *American Economic Review*, 92, 90–95.
- Cogley, T., and J. M. Nason (1995): “Output dynamics in real business cycle models,” *American Economic Review*, 79, 733–748.
- Cogley, T., and T. J. Sargent (2001): “Evolving post-World War II U.S. inflation dynamics,” *NBER Macroeconomics Annual*, 16, 331–373.
- (2005): “Drifts and volatilities: Monetary policies and outcomes in the post WWII US,” *Review of Economic Dynamics*, 8, 262–302.
- Coglianese, J., L. W. Davis, L. Kilian, and J. H. Stock (2017): “Anticipation, tax avoidance, and the price elasticity of gasoline demand,” *Journal of Applied Econometrics*, 32, 1–15.
- Connor, G., and R. A. Korajczyk (1986): “Performance measurement with the arbitrage pricing theory: A new framework for analysis,” *Journal of Financial Economics*, 15, 373–394.
- (1993): “A test for the number of factors in an approximate factor model,” *Journal of Finance*, 48, 1263–1291.
- Cooley, T. F., and M. Dwyer (1998): “Business cycle analysis without much theory: A look at structural VARs,” *Journal of Econometrics*, 83, 57–88.
- Cooley, T. F., and S. F. Leroy (1985): “Atheoretical macroeconomics: A critique,” *Journal of Monetary Economics*, 16, 283–308.
- Cooley, T. F., and E. C. Prescott (1995): “Economic growth and business cycles,” in *Frontiers of Business Cycle Research*, ed. by T. Cooley, pp. 1–38. Princeton University Press, Princeton.
- Cowles, M. K., and B. P. Carlin (1996): “Markov Chain Monte Carlo convergence diagnostics: A comparative review,” *Journal of the American Statistical Association*, 91, 883–904.
- Cragg, J. G., and S. G. Donald (1997): “Inferring the rank of a matrix,” *Journal of Econometrics*, 76, 223–250.
- Croushore, D., and C. L. Evans (2006): “Data revisions and the identification of monetary policy shocks,” *Journal of Monetary Economics*, 53, 1135–1160.
- Cushman, D. O., and T. Zha (1997): “Identifying monetary policy in a small open economy under flexible exchange rates,” *Journal of Monetary Economics*, 39, 433–448.
- D’Amico, S., and M. Farka (2011): “The Fed and the stock market: An identification based on intraday futures data,” *Journal of Business and Economic Statistics*, 29, 126–137.
- Datta, S. (1996): “On asymptotic properties of bootstrap for AR(1) processes,” *Journal of Statistical Planning and Inference*, 53, 361–374.
- Davidson, J. (1994): *Stochastic Limit Theory*. Oxford University Press, Oxford.

- Davidson, J. E. H., D. F. Hendry, F. Srba, and S. Yeo (1978): “Econometric modelling of the aggregate time-series relationship between consumer’s expenditure and income in the United Kingdom,” *Economic Journal*, 88, 661–692.
- Davidson, R., and E. Flachaire (2008): “The wild bootstrap, tamed at last,” *Journal of Econometrics*, 146, 162–169.
- Davis, L. W., and L. Kilian (2011): “Estimating the effect of a gasoline tax on carbon emissions,” *Journal of Applied Econometrics*, 26, 1187–1214.
- Davis, R. A., and L. Song (2010): “Noncausal vector AR processes with application to financial time series,” working paper, Columbia University.
- (2012): “Noncausal vector AR processes with application to economic time series,” working paper, Columbia University.
- De Mol, C., D. Giannone, and L. Reichlin (2008): “Forecasting using a large number of predictors – Is Bayesian regression a valid alternative to principal components?” *Journal of Econometrics*, 146, 318–328.
- Dedola, N., and S. Neri (2007): “What does a technology shock do? A VAR analysis with model-based sign restrictions,” *Journal of Monetary Economics*, 54, 512–549.
- Dees, S., F. Di Mauro, M. H. Pesaran, and V. Smith (2007): “Exploring the international linkages of the euro area: A global VAR analysis,” *Journal of Applied Econometrics*, 22, 1–38.
- DeJong, D. N., and C. Dave (2011): *Structural Macroeconomics*, 2nd ed. Princeton University Press, Princeton.
- Del Negro, M., and C. Otron (2007): “99 Luftballons: Monetary policy and the house price boom across U.S. states,” *Journal of Monetary Economics*, 54, 1962–1985.
- Del Negro, M., and F. Schorfheide (2004): “Priors from general equilibrium models for VARs,” *International Economic Review*, 45, 643–673.
- (2008): “Forming priors for DSGE models (and how it affects the assessment of nominal rigidities),” *Journal of Monetary Economics*, 55, 1191–1206.
- (2011): “Bayesian macroeconomics,” in *The Oxford Handbook of Bayesian Econometrics*, ed. by J. Geweke, G. Koop, and H. van Dijk, pp. 293–389. Oxford University Press, Oxford.
- (2013): “DSGE model-based forecasting,” in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann. Elsevier, Amsterdam.
- Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): “On the fit of New Keynesian models,” *Journal of Business and Economic Statistics*, 25, 123–143.
- Demetrescu, M., H. Lütkepohl, and P. Saikkonen (2009): “Testing for the cointegrating rank of a vector autoregressive process with uncertain deterministic trend term,” *Econometrics Journal*, 12, 414–435.
- Demiralp, S., and K. D. Hoover (2003): “Searching for the causal structure of a vector autoregression,” *Oxford Bulletin of Economics and Statistics*, 65, 745–767.
- Demiralp, S., K. D. Hoover, and S. J. Perez (2008): “A bootstrap method for identifying and evaluating a structural vector autoregression,” *Oxford Bulletin of Economics and Statistics*, 70, 509–533.

- Diebold, F. X. (1993): "Discussion: The effect of seasonal adjustment filters on tests for a unit root," *Journal of Business and Economic Statistics*, 55, 99–103.
- Diebold, F. X., and C. Chen (1996): "Testing structural stability with endogenous breakpoint: A size comparison of analytic and bootstrap procedures," *Journal of Econometrics*, 70, 221–241.
- Diebold, F. X., and L. Kilian (2001): "Measuring predictability: Theory and macroeconomic applications," *Journal of Applied Econometrics*, 16, 657–669.
- Dios Tena, J. de, and A. R. Tremayne (2009): "Modelling monetary transmission in UK manufacturing industry," *Economic Modelling*, 26, 1053–1066.
- Doan, T., R. Litterman, and C. A. Sims (1984): "Forecasting and conditional projection using realistic prior distributions," *Econometric Reviews*, 3, 1–100.
- Dolado, J. J., and H. Lütkepohl (1996): "Making Wald tests work for cointegrated VAR systems," *Econometric Reviews*, 15, 369–386.
- Doornik, J. A., and H. Hansen (1994): "A practical test of multivariate normality," working paper, Oxford University.
- Doornik, J. A., and D. F. Hendry (1997): *Modelling Dynamic Systems Using PcFiml 9.0 for Windows*. International Thomson Business Press, London.
- Dornbusch, R. (1976): "Expectations and exchange-rate dynamics," *Journal of Political Economy*, 84, 1161–1176.
- Dotsey, M., and M. Reid (1992): "Oil shocks, monetary policy, and economic activity," *Economic Review of the Federal Reserve Bank of Richmond*, 78, 14–27.
- Doz, C., D. Giannone, and L. Reichlin (2011): "A two-step estimator for large approximate dynamic factor models based on Kalman filtering," *Journal of Econometrics*, 164, 188–205.
- Drudi, R., A. Guay, and E. Renault (2007): "Indirect inference and calibration of dynamic stochastic general equilibrium models," *Journal of Econometrics*, 136, 397–430.
- Dufour, J.-M., L. Khalaf, and M. Kichian (2013): "Identification-robust analysis of DSGE and structural macroeconomic models," *Journal of Monetary Economics*, 60, 340–350.
- Dufour, J.-M., and E. Renault (1998): "Short run and long run causality in time series: Theory," *Econometrica*, 66, 1099–1125.
- Dufour, J.-M., and D. Stevanović (2013): "Factor-augmented VARMA models with macroeconomic applications," *Journal of Business and Economic Statistics*, 31, 491–506.
- Dufour, J.-M., and M. Taamouti (2005): "Projection-based statistical inference in linear structural models with possibly weak instruments," *Econometrica*, 73, 1351–1365.
- Durbin, J., and S. J. Koopman (2012): *Time Series Analysis by State Space Methods*, 2nd rev. ed. Oxford University Press, Oxford.
- Edelstein, P., and L. Kilian (2009): "How sensitive are consumer expenditures to retail energy prices?" *Journal of Monetary Economics*, 56, 766–779.
- Edgerton, D., and G. Shukur (1999): "Testing autocorrelation in a system perspective," *Econometric Reviews*, 18, 343–386.
- Efron, B. (1979): "Bootstrap methods: Another look at the jackknife," *Annals of Statistics*, 7, 1–26.

- (1982): *The Jackknife, the Bootstrap, and Other Resampling Plans*. Society for Industrial and Applied Mathematics, Philadelphia, PA.
- (1987): “Better bootstrap confidence intervals,” *Journal of the American Statistical Association*, 82, 171–185.
- Efron, B., and R. J. Tibshirani (1993): *An Introduction to the Bootstrap*. Chapman and Hall, New York.
- Ehrmann, M., M. Fratzscher, and R. Rigobon (2011): “Stocks, bonds, money markets and exchange rates: Measuring international financial transmission,” *Journal of Applied Econometrics*, 26, 948–974.
- Eichenbaum, M., and C. Evans (2005): “Nominal rigidities and the dynamic effects of a shock to monetary policy,” *Journal of Political Economy*, 113, 1–45.
- Eichenbaum, M., and C. L. Evans (1995): “Some empirical evidence of shocks to monetary policy to exchange rates,” *Quarterly Journal of Economics*, 110, 975–1010.
- Eichler, M. (2012): “Causal inference in time series analysis,” in *Causality: Statistical Perspectives and Applications*, ed. by C. Berzuini, P. Dawid, and L. Bernadinelli, pp. 327–354. John Wiley, Chichester.
- Eickmeier, S. (2007): “Business cycle transmission from the US to Germany — A structural factor approach,” *European Economic Review*, 51, 521–551.
- Eickmeier, S., and B. Hofmann (2013): “Monetary policy, housing booms, and financial (im)balances,” *Macroeconomic Dynamics*, 17, 830–860.
- Elder, J. (2003): “An impulse-response function for a vector autoregression with multivariate GARCH-in-mean,” *Economics Letters*, 79, 21–26.
- (2004): “Another perspective on the effects of inflation uncertainty,” *Journal of Money, Credit and Banking*, 36, 911–928.
- Elder, J., and A. Serletis (2010): “Oil price uncertainty,” *Journal of Money, Credit and Banking*, 42, 1137–1159.
- Elliott, G. (1998): “On the robustness of cointegration methods when regressors almost have unit roots,” *Econometrica*, 66, 149–158.
- Elliott, G., U. K. Müller, and M. W. Watson (2015): “Nearly optimal tests when a nuisance parameter is present under the null hypothesis,” *Econometrica*, 83, 771–811.
- Elliott, G., and J. H. Stock (1994): “Inference in time series regression when the order of integration of a regressor is unknown,” *Econometric Theory*, 10, 672–700.
- Enders, W., and B. S. Lee (1997): “Accounting for real and nominal exchange rate movements in the post-Bretton Woods period,” *Journal of International Money and Finance*, 16, 233–254.
- Engle, R. F., and C. W. J. Granger (1987): “Co-integration and error correction: Representation, estimation, and testing,” *Econometrica*, 55, 251–276.
- Engle, R. F., D. F. Hendry, and J.-F. Richard (1983): “Exogeneity,” *Econometrica*, 51, 277–304.
- Erceg, C. J., L. Guerrieri, and C. Gust (2005): “Can long-run restrictions identify technology shocks?” *Journal of the European Economic Association*, 3, 1237–1278.
- Faust, J. (1998): “The robustness of identified VAR conclusions about money,” *Carnegie-Rochester Conference Series on Public Policy*, 49, 207–244.

- Faust, J., and E. M. Leeper (1997): "When do long-run identifying restrictions give reliable results?" *Journal of Business and Economic Statistics*, 15, 345–353.
- Faust, J., J. Rogers, E. Swanson, and J. H. Wright (2003): "Identifying the effects of monetary policy shocks on exchange rates using high-frequency data," *Journal of the European Economic Association*, 1, 1031–1057.
- Faust, J., E. T. Swanson, and J. H. Wright (2004): "Identifying VARs based on high frequency futures data," *Journal of Monetary Economics*, 51, 1107–1131.
- Favero, C. A. (2001): *Applied Macroeometrics*. Oxford University Press, Oxford.
- Favero, C., and F. Giavazzi (2012): "Measuring tax multipliers: The narrative method in fiscal VARs," *American Economic Journal: Economic Policy*, 4, 69–94.
- Favero, C. A., M. Marcellino, and F. Neglia (2005): "Principal components at work: The empirical analysis of monetary policy with large data sets," *Journal of Applied Econometrics*, 20, 603–620.
- Fernald, J. (2007): "Trend breaks, long-run restrictions, and contractionary technology improvements," *Journal of Monetary Economics*, 54, 2467–2485.
- Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2006): "Economic and VAR shocks: What can go wrong?" *Journal of the European Economic Association*, 4, 466–474.
- (2007): "How structural are structural parameters?" *NBER Macroeconomics Annual*, 22, 83–137.
- Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of understanding VARs," *American Economic Review*, 97, 1021–1026.
- Findley, D. F., B. C. Monsell, W. R. Bell, M. C. Otto, and B.-C. Chen (1998): "New capabilities and methods of the X-12-ARIMA seasonal-adjustment program," *Journal of Business and Economic Statistics*, 16, 127–177.
- Fisher, J. (2006): "The dynamic effects of neutral and investment-specific technology shocks," *Journal of Political Economy*, 114, 413–451.
- Fisher, L. A., H.-S. Huh, and A. R. Pagan (2016): "Econometric methods for modelling systems with a mixture of I(1) and I(0) variables," *Journal of Applied Econometrics*, 31, 892–911.
- Forni, M., and L. Gambetti (2010): "The dynamic effects of monetary policy: A structural factor model approach," *Journal of Monetary Economics*, 57, 203–216.
- (2014): "Sufficient information in structural VARs," *Journal of Monetary Economics*, 66, 124–136.
- Forni, M., D. Giannone, M. Lippi, and L. Reichlin (2009): "Opening the black box: Structural factor models with large cross-sections," *Econometric Theory*, 25, 1319–1347.
- Forni, M., M. Hallin, M. Lippi, and L. Reichlin (2000): "The generalized dynamic factor model: Identification and estimation," *Review of Economics and Statistics*, 82, 540–554.
- (2004): "Generalized dynamic factor model consistency and rates," *Journal of Econometrics*, 119, 231–255.
- (2005): "The generalized dynamic factor model: One-sided estimation and forecasting," *Journal of the American Statistical Association*, 100, 830–840.

- Forni, M., and M. Lippi (2001): “The generalized dynamic factor model: Representation theory,” *Econometric Theory*, 17, 1113–1141.
- Forni, M., and L. Reichlin (1998): “Let’s get real: A factor analytical approach to disaggregated business cycle dynamics,” *Review of Economic Studies*, 65, 453–473.
- Foroni, C., E. Ghysels, and M. Marcellino (2013): “Mixed-frequency vector autoregressive models,” *Advances in Econometrics*, 31, 247–272.
- Foroni, C., and M. Marcellino (2016): “Mixed frequency structural vector autoregressive models,” *Journal of the Royal Statistical Society A*, 179, 403–425.
- Franchi, M., and A. Vidotto (2013): “A check for finite-order VAR representations of DSGE models,” *Economics Letters*, 120, 100–103.
- Francis, N., M. Owyang, J. E. Roush, and R. DiCecio (2014): “A flexible finite-horizon alternative to long-run restrictions with an application to technology shocks,” *Review of Economics and Statistics*, 96, 638–647.
- Francis, N., and V. A. Ramey (2005): “Is the technology-driven real business cycle hypothesis dead? Shocks and aggregate fluctuations revisited,” *Journal of Monetary Economics*, 52, 1379–1399.
- (2009): “Measures of hours per capita with implications for the technology-hours debate,” *Journal of Money, Credit and Banking*, 41, 1071–1097.
- Francq, C., and H. Raïssi (2007): “Multivariate portmanteau test for autoregressive models with uncorrelated but nonindependent errors,” *Journal of Time Series Analysis*, 28, 454–470.
- Franses, P. H., and R. M. Kunst (1999): “On the role of seasonal intercepts in seasonal cointegration,” *Oxford Bulletin of Economics and Statistics*, 61, 409–433.
- Franses, P. H., and R. Paap (2004): *Periodic Time Series Models*. Oxford University Press, Oxford.
- Freedman, D. A. (1981): “Bootstrapping regression models,” *Annals of Statistics*, 9, 1218–1228.
- Friedman, M., and A. J. Schwartz (1963): *A Monetary History of the United States, 1867–1960*. Princeton University Press, Princeton.
- Fry, R., and A. R. Pagan (2011): “Sign restrictions in structural vector autoregressions: A critical review,” *Journal of Economic Literature*, 49, 938–960.
- Fujita, S. (2011): “Dynamics of worker flows and vacancies: Evidence from the sign restriction approach,” *Journal of Applied Econometrics*, 26, 89–121.
- Gafarov, B., M. Meier, and J. L. Montiel Olea (2015a): “Delta method inference for a class of set-identified SVARs,” working paper, New York University.
- (2015b): “Projection inference for set-identified SVARs,” working paper, New York University.
- Galí, J. (1992): “How well does the IS-LM model fit postwar US data?” *Quarterly Journal of Economics*, 107, 709–738.
- Galí, J. (1999): “Technology, employment, and the business cycle: Do technology shocks explain aggregate fluctuations?” *American Economic Review*, 89, 249–271.
- Gallant, A. R., D. A. Hsieh, and G. E. Tauchen (1991): “On fitting a recalcitrant series: The Pound/Dollar exchange rate, 1974–1983,” in *Nonparametric and Semiparametric Methods in Econometrics and Statistics*, ed. by W. A. Barnett, J. Powell, and G. E. Tauchen, pp. 199–240. Cambridge University Press, Cambridge.

- Gallant, A. R., and D. W. Nychka (1987): "Seminonparametric maximum likelihood estimation," *Econometrica*, 55, 363–390.
- Gallant, A. R., P. E. Rossi, and G. Tauchen (1993): "Nonlinear dynamic structures," *Econometrica*, 61, 871–907.
- Gallant, A. R., and G. E. Tauchen (1989): "Seminonparametric estimation of conditional constrained heterogeneous processes: Asset pricing applications," *Econometrica*, 57, 1091–1120.
- (1994): "SNP: A program for nonparametric time series analysis, Version 8.3, User's Guide," Department of Economics, University of North Carolina.
- Galvao, A. B., and M. Marcellino (2014): "The effects of the monetary policy stance on the transmission mechanism," *Studies in Nonlinear Dynamics and Econometrics*, 18, 217–236.
- Garcia, R. (1998): "Asymptotic null distribution of the likelihood ratio test in Markov switching models," *International Economic Review*, 39, 763–788.
- Gefang, D. (2012): "Money-output causality revisited – A Bayesian logistic smooth transition VECM perspective," *Oxford Bulletin of Economics and Statistics*, 74, 131–151.
- Gefang, D., and R. Strachan (2010): "Nonlinear impacts of international business cycles on the U.K. – A Bayesian smooth transition VAR approach," *Studies in Nonlinear Dynamics and Econometrics*, 14, Article 2.
- Gelfand, A. E., and A. F. M. Smith (1990): "Sampling based approaches to calculating marginal densities," *Journal of the American Statistical Association*, 85, 398–409.
- George, E. I., D. Sun, and S. Ni (2008): "Bayesian stochastic search for VAR model restrictions," *Journal of Econometrics*, 142, 553–580.
- Gertler, M., and P. Karadi (2015): "Monetary policy surprises, credit costs, and economic activity," *American Economic Journal: Macroeconomics*, 7, 44–76.
- Geweke, J. (2001): "Bayesian econometrics and forecasting," *Journal of Econometrics*, 100, 11–15.
- (2005): *Contemporary Bayesian Econometrics and Statistics*. Wiley, New York.
- Geweke, J., and C. Whiteman (2006): "Bayesian forecasting," in *Handbook of Economic Forecasting, Volume 1*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann, pp. 3–80. Elsevier, Amsterdam.
- Ghysels, E., and D. R. Osborn (2001): *The Econometric Analysis of Seasonal Time Series*. Cambridge University Press, Cambridge.
- Giacomini, R. (2013): "The relationship between DSGE and VAR models," *Advances in Econometrics*, 32, 1–25.
- Giacomini, R., and T. Kitagawa (2015): "Robust inference about partially identified SVARs," working paper, University College London.
- Giacomini, R., and R. Ragusa (2014): "Theory-coherent forecasting," *Journal of Econometrics*, 182, 145–155.
- Giannone, D., M. Lenza, and G. Primiceri (2015): "Prior selection for vector autoregressions," *Review of Economics and Statistics*, 97, 436–451.
- Giannone, D., and L. Reichlin (2006): "Does information help recovering structural shocks from past observations?" *Journal of the European Economic Association*, 4, 455–465.

- Giannone, D., L. Reichlin, and L. Sala (2004): "Monetary policy in real time," *NBER Macroeconomics Annual*, 19, 161–200.
- Godfrey, L. G. (1978): "Testing for higher order serial correlation in regression equations when the regressors include lagged dependent variables," *Econometrica*, 46, 1303–1313.
- Gómez, V., and A. Maravall (1997): "Programs Tramo and Seats: Instructions for the user," manual, Madrid.
- Gonçalves, S., and L. Kilian (2004): "Bootstrapping autoregressions with conditional heteroskedasticity of unknown form," *Journal of Econometrics*, 123, 89–120.
- (2007): "Asymptotic and bootstrap inference for AR(∞) processes with conditional heteroskedasticity," *Econometric Reviews*, 26, 609–641.
- Gonçalves, S., and B. Perron (2014): "Bootstrapping factor-augmented regression models," *Journal of Econometrics*, 182, 156–173.
- Gonzalo, J., and J.-Y. Pitarkis (1999): "Dimensionality effect in cointegration analysis," in *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, ed. by R. Engle and H. White, pp. 212–229. Oxford University Press, Oxford.
- (2002): "Lag length estimation in large dimensional systems," *Journal of Time Series Analysis*, 23, 401–423.
- Gordon, D. B., and E. M. Leeper (1994): "The dynamic impacts of monetary policy: An exercise in tentative identification," *Journal of Political Economy*, 102, 1228–1247.
- Gospodinov, N. (2003): "Asymptotic confidence intervals for impulse responses of near-integrated processes," working paper, Concordia University.
- (2004): "Asymptotic confidence intervals for impulse responses of near-integrated processes," *Econometrics Journal*, 7, 505–527.
- (2010): "Inference in nearly nonstationary SVAR models with long-run identifying restrictions," *Journal of Business and Economic Statistics*, 28, 1–12.
- Gospodinov, N., A. M. Herrera, and E. Pesavento (2013): "Unit roots, cointegration and pre-testing in VAR models," *Advances in Econometrics*, 31, 81–115.
- Gospodinov, N., A. Maynard, and E. Pesavento (2011): "Sensitivity of impulse responses to small low frequency co-movements: Reconciling the evidence on the effects of technology shocks," *Journal of Business and Economic Statistics*, 29, 455–467.
- Gourieroux, C., and A. Monfort (2014): "Revisiting identification in structural VAR models," working paper, CREST.
- Granger, C. W. J. (1969): "Investigating causal relations by econometric models and cross-spectral methods," *Econometrica*, 37, 424–438.
- (1980): "Testing for causality: A personal viewpoint," *Journal of Economic Dynamics and Control*, 2, 329–352.
- (1981): "Some properties of time series data and their use in econometric model specification," *Journal of Econometrics*, 16, 121–130.
- Granger, C. W. J., and R. Joyeux (1980): "An introduction to long-memory time series models and fractional differencing," *Journal of Time Series Analysis*, 1, 15–29.
- Granger, C. W. J., and P. Newbold (1977): *Forecasting Economic Time Series*, 1st ed. Academic Press, New York.

- Granger, C. W. J., and T. Teräsvirta (1993): *Modelling Nonlinear Economic Relationships*. Oxford University Press, Oxford.
- Gregory, A. W., and M. R. Veall (1985): "Formulating Wald tests of nonlinear restrictions," *Econometrica*, 53, 1465–1468.
- Guerron-Quintana, P., A. Inoue, and L. Kilian (2013): "Frequentist inference in weakly identified DSGE models," *Quantitative Economics*, 4, 197–229.
- (2017): "Impulse response matching estimators for DSGE models," *Journal of Econometrics*, 196, 144–155.
- Gürkaynak, R. S., B. Kisacikoglu, and B. Rossi (2013): "Do DSGE models forecast more accurately out-of-sample than VAR models?" *Advances in Econometrics*, 32, 27–79.
- Gust, C., and R. J. Vigfusson (2009): "The power of long-run VARs," International Finance Discussion Paper 978, Board of Governors of the Federal Reserve System.
- Hall, A. R., A. Inoue, J. M. Nason, and B. Rossi (2012): "Information criteria for impulse response function matching estimation of DSGE models," *Journal of Econometrics*, 170, 499–518.
- Hall, P. (1992): *The Bootstrap and Edgeworth Expansion*. Springer-Verlag, New York.
- Hall, P., and J. L. Horowitz (1996): "Bootstrap critical values for tests based on generalized method of moments estimators," *Econometrica*, 64, 891–916.
- Hallin, M., and R. Liška (2007): "Determining the number of factors in the general dynamic factor model," *Journal of the American Statistical Association*, 102, 603–617.
- (2011): "Dynamic factors in the presence of blocks," *Journal of Econometrics*, 163, 29–41.
- Hamilton, J. D. (1989): "A new approach to the economic analysis of nonstationary time series and the business cycle," *Econometrica*, 57, 357–384.
- (1994): *Time Series Analysis*. Princeton University Press, Princeton, NJ.
- (1996): "This is what happened to the oil price-macroeconomy relationship," *Journal of Monetary Economics*, 38, 215–220.
- (1997): "Measuring the liquidity effect," *American Economic Review*, 87, 80–97.
- (2003): "What is an oil shock?" *Journal of Econometrics*, 113, 363–398.
- (2009): "Causes and consequences of the oil shock of 2007–08," *Brookings Papers on Economic Activity*, 40, 215–283.
- (2016): "Why you should never use the Hodrick-Prescott filter," working paper, University of California at San Diego.
- Hamilton, J. D., and A. M. Herrera (2004): "Oil shocks and aggregate economic behavior: The role of monetary policy," *Journal of Money, Credit and Banking*, 36, 265–286.
- Hammersley, J. M., and D. C. Handscomb (1964): *Monte Carlo Methods*. Methuen, London.
- Han, X., and A. Inoue (2015): "Tests for parameter instability in dynamic factor models," *Econometric Theory*, 31, 1117–1152.
- Hannan, E. J., and M. Deistler (1988): *The Statistical Theory of Linear Systems*. Wiley, New York.
- Hannan, E. J., and B. G. Quinn (1979): "The determination of the order of an autoregression," *Journal of the Royal Statistical Society B*, 41, 190–195.

- Hansen, B. E. (1992): “The likelihood ratio test under nonstandard conditions: Testing the Markov switching model of GNP,” *Journal of Applied Econometrics*, 7, 61–82 (Erratum: 11, 195–198).
- (1997a): “Approximate asymptotic *p* values for structural-change tests,” *Journal of Business and Economic Statistics*, 15, 60–67.
- (1997b): “Inference in TAR Models,” *Studies in Nonlinear Dynamics and Econometrics*, 2, 1–14.
- (1999): “The grid bootstrap and the autoregressive model,” *Review of Economics and Statistics*, 81, 594–607.
- (2000): “Sample splitting and threshold estimation,” *Econometrica*, 68, 575–603.
- Hansen, L. P. (1982): “Large sample properties of generalized method of moments estimators,” *Econometrica*, 50, 1029–1054.
- Hansen, L. P., and T. J. Sargent (1980): “Formulating and estimating dynamic linear rational expectations models,” *Journal of Economic Dynamics and Control*, 2, 7–46.
- (1991): “Two difficulties in interpreting vector autoregressions,” in *Rational Expectations Econometrics*, ed. by L. P. Hansen and T. J. Sargent, pp. 77–120. Westview Press, Boulder.
- Hansen, P. R. (2003): “Structural changes in the cointegrated vector autoregressive model,” *Journal of Econometrics*, 114, 261–295.
- Hanson, M. S. (2004): “The ‘price puzzle’ reconsidered,” *Journal of Monetary Economics*, 51, 1385–1413.
- Harding, D., and A. R. Pagan (2002): “Dissecting the cycle: A methodological investigation,” *Journal of Monetary Economics*, 49, 365–381.
- Härdle, W., H. Lütkepohl, and R. Chen (1997): “A review of nonparametric time series analysis,” *International Statistical Review*, 65, 49–72.
- Härdle, W., and A. Tsybakov (1997): “Local polynomial estimators of the volatility function in nonparametric autoregression,” *Journal of Econometrics*, 81, 223–242.
- Härdle, W., A. Tsybakov, and L. Yang (1998): “Nonparametric vector autoregression,” *Journal of Statistical Planning and Inference*, 68, 221–245.
- Harvey, A. C. (1989): *Forecasting, Structural Time Series Models, and the Kalman Filter*. Cambridge University Press, Cambridge.
- Hausman, J. A., and W. K. Newey (1995): “Nonparametric estimation of exact consumers surplus and deadweight loss,” *Econometrica*, 63, 1445–1476.
- He, C., and T. Teräsvirta (1999): “Fourth moment structure of the GARCH(p, q) process,” *Econometric Theory*, 15, 824–846.
- Heinlein, R., and H.-M. Krolzig (2013): “Monetary policy and exchange rates: A balanced two-country cointegrated VAR model approach,” working paper, University of Kent.
- Helbling, T., and T. Bayoumi (2003): “Are they all in the same boat? The 2000–2001 growth slowdown and the G7-business cycle linkages,” Working Paper No. 03/46, International Monetary Fund.
- Hendry, D. F., and T. von Ungern-Sternberg (1981): “Liquidity and inflation effects on consumer’s expenditure,” in *Essays in the Theory and Measurement of Consumer’s Behavior*, ed. by A. S. Deaton. Cambridge University Press, Cambridge.
- Herbst, E., and F. Schorfheide (2016): *Bayesian Estimation of DSGE Models*. Princeton University Press, Princeton.

- Herrera, A. M., L. G. Lagalo, and T. Wada (2011): "Oil price shocks and industrial production: Is the relationship linear?" *Macroeconomic Dynamics*, 15, 472–497.
- Herwartz, H. (2015): "Structural VAR modelling with independent innovations—An analysis of macroeconomic dynamics in the Euro Area based on a novel identification scheme," working paper, University of Göttingen.
- Herwartz, H., and H. Lütkepohl (2011): "Generalized least squares estimation for cointegration parameters under conditional heteroskedasticity," *Journal of Time Series Analysis*, 32, 281–291.
- (2014): "Structural vector autoregressions with Markov switching: Combining conventional with statistical identification of shocks," *Journal of Econometrics*, 183, 104–116.
- Herwartz, H., and M. Plödt (2016): "The macroeconomic effects of oil price shocks: Evidence from a statistical identification approach," *Journal of International Money and Finance*, 61, 30–44.
- Hodrick, R. J., and E. C. Prescott (1997): "Postwar U.S. business cycles: An empirical investigation," *Journal of Money, Credit and Banking*, 29, 1–16.
- Hoover, K. D. (2005): "Automatic inference of the contemporaneous causal order of a system of equations," *Econometric Theory*, 21, 69–77.
- Hoover, K. D., S. Demiralp, and S. J. Perez (2009): "Empirical identification of the vector autoregression: The causes and effects of US M2," in *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, ed. by J. Castle, and N. Shephard, pp. 37–58. Oxford University Press, Oxford.
- Hoover, K. D., and S. J. Perez (1994): "Post hoc ergo propter once more. An evaluation of 'Does monetary policy matter?' in the spirit of James Tobin," *Journal of Monetary Economics*, 34, 47–74.
- Hosking, J. R. M. (1980): "The multivariate portmanteau statistic," *Journal of the American Statistical Association*, 75, 602–608.
- (1981a): "Equivalent forms of the multivariate portmanteau statistic," *Journal of the Royal Statistical Society B*, 43, 261–262.
- (1981b): "Fractional differencing," *Biometrika*, 68, 165–176.
- (1981c): "Lagrange-multiplier tests of multivariate time series models," *Journal of the Royal Statistical Society B*, 43, 219–230.
- Howrey, E. P. (1968): "A spectrum analysis of the long swing hypothesis," *International Economic Review*, 9, 228–252.
- Hsiao, C. (1979): "Autoregressive modeling of Canadian money and income data," *Journal of the American Statistical Association*, 74, 553–560.
- (1981): "Autoregressive modeling and money-income causality detection," *Journal of Monetary Economics*, 7, 85–106.
- Huang, J., and Y. Pawitan (2000): "Quasi-likelihood estimation of non-invertible moving average processes," *Scandinavian Journal of Statistics*, 27, 689–702.
- Hubrich, K., H. Lütkepohl, and P. Saikkonen (2001): "A review of systems cointegration tests," *Econometric Reviews*, 20, 247–318.
- Hubrich, K., and T. Teräsvirta (2013): "Thresholds and smooth transitions in vector autoregressive models," *Advances in Econometrics*, 32, 273–326.
- Hubrich, K., D. Waggoner, and T. Zha (2016): "Monetary policy and financial stability: Unconventional monetary policy, leveraging, and financial stress," working paper, Federal Reserve Board.

- Hurvich, C. M., and C.-L. Tsai (1993): “A corrected Akaike information criterion for vector autoregressive model selection,” *Journal of Time Series Analysis*, 14, 271–279.
- Hussain, S. M., and S. Malik (2016): “Asymmetric effects of exogenous tax changes,” *Journal of Economic Dynamics and Control*, 69, 268–300.
- Hyvärinen, A., J. Karhunen, and E. Oja (2001): *Independent Component Analysis*. Wiley, New York.
- Ing, C.-K. (2003): “Multistep prediction in autoregressive processes,” *Econometric Theory*, 19, 254–279.
- (2004): “Selecting optimal multistep predictors for autoregressive processes of unknown order,” *Annals of Statistics*, 32, 693–722.
- (2007): “Accumulated prediction errors, information criteria and optimal forecasting for autoregressive time series,” *Annals of Statistics*, 35, 1238–1277.
- Ing, C.-K., J.-L. Lin, and S.-H. Yu (2009): “Toward optimal multistep forecasts in nonstationary autoregressions,” *Bernoulli*, 15, 402–437.
- Ing, C.-K., C.-Y. Sin, and S.-H. Yu (2010): “Prediction errors in nonstationary autoregressions of infinite order,” *Econometric Theory*, 26, 774–803.
- Ing, C.-K., and C. Z. Wei (2003): “On same realization prediction in an infinite-order autoregressive process,” *Journal of Multivariate Analysis*, 85, 130–155.
- (2005): “Order selection for same-realization prediction in autoregressive processes,” *Annals of Statistics*, 33, 2423–2474.
- Ingram, B. F., and C. H. Whiteman (1994): “Supplanting the ‘Minnesota’ prior: Forecasting macroeconomic time series using real business cycle model priors,” *Journal of Monetary Economics*, 34, 497–510.
- Inoue, A., and L. Kilian (2002a): “Bootstrapping autoregressive processes with possible unit roots,” *Econometrica*, 70, 377–391.
- (2002b): “Bootstrapping smooth functions of slope parameters and innovation variances in VAR(∞) models,” *International Economic Review*, 43, 309–332.
- (2003): “The continuity of the limit distribution in the parameter of interest is not essential for the validity of the bootstrap,” *Econometric Theory*, 19, 944–961.
- (2004): “In-sample or out-of-sample tests of predictability: Which one should we use?” *Econometric Reviews*, 23, 371–402.
- (2006): “On the selection of forecasting models,” *Journal of Econometrics*, 130, 273–306.
- (2013): “Inference on impulse response functions in structural VAR models,” *Journal of Econometrics*, 177, 1–13.
- (2016): “Joint confidence sets for structural impulse responses,” *Journal of Econometrics*, 192, 421–432.
- (2017): “Corrigendum to ‘Inference on impulse response functions in structural VAR models,’ *Journal of Econometrics*, 177 (2013), 1–13,” *Journal of Econometrics*, forthcoming.
- Inoue, A., L. Kilian, and F. B. Kiraz (2009): “Do actions speak louder than words? Household expectations of inflation based on micro consumption data,” *Journal of Money, Credit and Banking*, 41, 1331–1363.
- Jacobs, J. P. A. M., and P. W. Otter (2008): “Determining the number of factors and lag order in dynamic factor models: A minimum entropy approach,” *Econometric Reviews*, 27, 385–397.

- Jarque, C. M., and A. K. Bera (1987): "A test for normality of observations and regression residuals," *International Statistical Review*, 55, 163–172.
- Jeliazkov, I. (2013): "Nonparametric vector autoregressions: Specification, estimation, and inference," *Advances in Econometrics*, 32, 327–359.
- Jentsch, C., D. N. Politis, and E. Paparoditis (2015): "Block bootstrap theory for multivariate integrated and cointegrated processes," *Journal of Time Series Analysis*, 36, 416–441.
- Jo, S. (2014): "The effect of oil price uncertainty on global real economic activity," *Journal of Money, Credit and Banking*, 46, 1113–1135.
- Johansen, S. (1988): "Statistical analysis of cointegration vectors," *Journal of Economic Dynamics and Control*, 12, 231–254.
- (1995): *Likelihood-Based Inference in Cointegrated Vector Autoregressive Models*. Oxford University Press, Oxford.
- (2002): "A small sample correction for the test of cointegrating rank in the vector autoregressive model," *Econometrica*, 70, 1929–1961.
- (2008): "A representation theory for a class of vector autoregressive models for fractional processes," *Econometric Theory*, 24, 651–676.
- Johansen, S., and M. Ø. Nielsen (2012): "Likelihood inference for a fractionally cointegrated vector autoregressive model," *Econometrica*, 80, 2667–2732.
- Johansen, S., and E. Schaumburg (1999): "Likelihood analysis of seasonal cointegration," *Journal of Econometrics*, 88, 301–339.
- Jordà, O. (2005): "Estimation and inference of impulse responses by local projection," *American Economic Review*, 95, 161–182.
- (2009): "Simultaneous confidence regions for impulse responses," *Review of Economics and Statistics*, 91, 629–647.
- Jordà, O., and M. Marcellino (2010): "Path forecast evaluation," *Journal of Applied Econometrics*, 25, 635–662.
- Judge, G. G., W. E. Griffiths, R. C. Hill, H. Lütkepohl, and T.-C. Lee (1985): *The Theory and Practice of Econometrics*, 2nd ed. John Wiley and Sons, New York.
- Jutten, C., and J. Herault (1991): "Blind separation of sources, part I: An adaptive algorithm based on neuromimetic architecture," *Signal Processing*, 24, 1–10.
- Kadiyala, K. R., and S. Karlsson (1997): "Numerical methods for estimation and inference in Bayesian VAR models," *Journal of Applied Econometrics*, 12, 99–132.
- Karlsson, S. (2013): "Forecasting with Bayesian vector autoregressions," in *Handbook of Economic Forecasting*, ed. by G. Elliott and A. Timmermann, vol. 2, pp. 791–897. Elsevier, Amsterdam.
- Kascha, C., and K. Mertens (2009): "Business cycle analysis and VARMA models," *Journal of Economic Dynamics and Control*, 33, 267–282.
- Kashyap, A. K., J. C. Stein, and D. W. Wilcox (1993): "Monetary policy and credit conditions: Evidence from the composition of external finance," *American Economic Review*, 83, 78–98.
- Keating, J. (1990): "Identifying VAR models under rational expectations," *Journal of Monetary Economics*, 25, 453–476.
- (1992): "Structural approaches to vector autoregressions," *Federal Reserve Bank of St Louis Review*, 74, 37–57.
- (1993): "Asymmetric vector autoregression," in *Proceedings of the Business and Economic Statistics Section*, pp. 68–73. American Statistical Association.

- Kehoe, P. (2006): "Comment on 'Assessing Structural VARs,'" *NBER Macroeconomics Annual*, 21, 73–96.
- Kilian, L. (1998a): "Accounting for lag order uncertainty in autoregressions: The endogenous lag order bootstrap algorithm," *Journal of Time Series Analysis*, 19, 531–548.
- (1998b): "Confidence intervals for impulse responses under departures from normality," *Econometric Reviews*, 17, 1–29.
- (1998c): "Small-sample confidence intervals for impulse response functions," *Review of Economics and Statistics*, 80, 218–230.
- (1999a): "Exchange rates and monetary fundamentals: What do we learn from long-horizon regressions?" *Journal of Applied Econometrics*, 14, 491–510.
- (1999b): "Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses," *Review of Economics and Statistics*, 81, 652–660.
- (2001): "Impulse response analysis in vector autoregressions with unknown lag order," *Journal of Forecasting*, 20, 161–179.
- (2006): "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," CEPR Discussion Paper No. 5994.
- (2007): "Comment on 'On the Fit of New Keynesian Models' by Del Negro, Schorfheide, Smets and Wouters," *Journal of Business and Economic Statistics*, 25, 156–159.
- (2008a): "The economic effects of energy price shocks," *Journal of Economic Literature*, 46, 871–909.
- (2008b): "Exogenous oil supply shocks: How big are they and how much do they matter for the U.S. economy?" *Review of Economics and Statistics*, 90, 216–240.
- (2008c): "A comparison of the effects of exogenous oil supply shocks on output and inflation in the G7 countries," *Journal of the European Economic Association*, 6, 78–121.
- (2009): "Not all oil price shocks are alike: Disentangling demand and supply shocks in the crude oil market," *American Economic Review*, 99, 1053–1069.
- (2010): "Explaining fluctuations in gasoline prices: A joint model of the global crude oil market and the US retail gasoline market," *Energy Journal*, 31, 87–104.
- (2013): "Structural vector autoregressions," in *Handbook of Research Methods and Applications on Empirical Macroeconomics*, ed. by N. Hashimzade and M. Thornton, pp. 515–554. Edward Elger, Cheltenham.
- (2014): "Oil price shocks: Causes and consequences," *Annual Review of Resource Economics*, 6, 133–154.
- (2016): "The impact of the shale oil revolution on U.S. oil and gas prices," *Review of Environmental Economics and Policy*, 10, 185–205.
- (2017): "The impact of the fracking boom on Arab oil producers," *Energy Journal*, forthcoming.
- Kilian, L., and P.-L. Chang (2000): "How accurate are confidence intervals for impulse responses in large VAR models?" *Economics Letters*, 69, 299–307.
- Kilian, L., and U. Demiroglu (2000): "Residual-based tests for normality in autoregressions: Asymptotic theory and simulation evidence," *Journal of Business and Economic Statistics*, 18, 40–50.

- Kilian, L., and B. Hicks (2013): "Did unexpectedly strong economic growth cause the oil price shock of 2003–2008?" *Journal of Forecasting*, 32, 385–394.
- Kilian, L., and V. Ivanov (2005): "A practitioners guide to lag order selection for VAR impulse response analysis," *Studies in Nonlinear Dynamics and Econometrics*, 9, Article 2.
- Kilian, L., and Y. J. Kim (2009): "Do local projections solve the bias problem in impulse response inference?" CEP Discussion Paper No. 7266.
- (2011): "How reliable are local projection estimators of impulse responses?" *Review of Economics and Statistics*, 93, 1460–1466.
- Kilian, L., and T. K. Lee (2014): "Quantifying the speculative component in the real price of oil: The role of global oil inventories," *Journal of International Money and Finance*, 42, 71–87.
- Kilian, L., and L. T. Lewis (2011): "Does the Fed respond to oil price shocks?" *Economic Journal*, 121, 1047–1072.
- Kilian, L., and S. Manganelli (2007): "Quantifying the risk of deflation," *Journal of Money, Credit and Banking*, 39, 561–590.
- Kilian, L., and D. P. Murphy (2012): "Why agnostic sign restrictions are not enough: Understanding the dynamics of oil market VAR models," *Journal of the European Economic Association*, 10, 1166–1188.
- (2014): "The role of inventories and speculative trading in the global market for crude oil," *Journal of Applied Econometrics*, 29, 454–478.
- Kilian, L., and L. E. Ohanian (2002): "Unit roots, trend breaks and transitory dynamics: A macroeconomic perspective," *Macroeconomic Dynamics*, 6, 614–632.
- Kilian, L., and C. Park (2009): "The impact of oil price shocks on the U.S. stock market," *International Economic Review*, 50, 1267–1287.
- Kilian, L., and C. Vega (2011): "Do energy prices respond to US macroeconomic news? A test of the hypothesis of predetermined energy prices," *Review of Economics and Statistics*, 93, 660–671.
- Kilian, L., and R. J. Vigfusson (2011a): "Are the responses of the U.S. economy asymmetric in energy price increases and decreases?" *Quantitative Economics*, 2, 419–453.
- (2011b): "Nonlinearities in the oil price-output relationship," *Macroeconomic Dynamics*, 15, 337–363.
- (2017): "The role of oil price shocks in causing U.S. recessions," *Journal of Money, Credit and Banking*, forthcoming.
- Kilian, L., and T. Zha (2002): "Quantifying the uncertainty about the half-life of deviations from PPP," *Journal of Applied Econometrics*, 17, 107–125.
- Kim, J. H. (1999): "Asymptotic and bootstrap prediction regions for vector autoregression," *International Journal of Forecasting*, 15, 393–403.
- (2004): "Bias-corrected bootstrap prediction regions for vector autoregression," *Journal of Forecasting*, 23, 141–154.
- Kim, S., N. Shephard, and S. Chib (1998): "Stochastic volatility: Likelihood inference and comparison with ARCH models," *Review of Economic Studies*, 65, 361–393.
- King, R. G., C. I. Plosser, J. H. Stock, and M. W. Watson (1991): "Stochastic trends and economic fluctuations," *American Economic Review*, 81, 819–840.

- King, R. G., and M. W. Watson (1997): "Testing long-run neutrality," *Federal Reserve Bank of Richmond Economic Quarterly*, 83, 69–101.
- Kitagawa, T., J. L. Montiel Olea, and J. Payne (2015): "Calibrated credible sets for nondifferentiable functions," working paper, New York University.
- Kleibergen, F., and R. Paap (2002): "Priors, posteriors and Bayes factors for a Bayesian analysis of cointegration," *Journal of Econometrics*, 111, 223–249.
- Kleibergen, F., and H. K. van Dijk (1994): "On the shape of the likelihood/posterior in cointegration models," *Econometric Theory*, 10, 514–551.
- Klein, L. R. (1950): *Economic Fluctuations in the United States, 1921–1941*. Cowles Commission for Research in Economics, Monograph No. 11. John Wiley and Sons, New York.
- Klein, L. R., and A. Goldberger (1955): *An Econometric Model of the United States, 1929–1952*. North-Holland, Amsterdam.
- Kloek, T., and H. K. van Dijk (1978): "Bayesian estimates of equation system parameters: An application of integration by Monte Carlo," *Econometrica*, 46, 1–20.
- Koltchinskii, V. I. (1997): "M-estimation, convexity and quantiles," *Annals of Statistics*, 25, 435–477.
- Koop, G. (2003): *Bayesian Econometrics*. Wiley, Chichester.
- (2013a): "Forecasting with medium and large Bayesian VARs," *Journal of Applied Econometrics*, 28, 177–203.
- (2013b): "Using VARs and TVP-VARs with many macroeconomic variables," Discussion Paper SIRE-DP-2013-35, University of Strathclyde.
- Koop, G., and D. Korobilis (2009): "Bayesian multivariate time series methods for empirical macroeconomics," *Foundations and Trends in Econometrics*, 3, 267–358.
- Koop, G., R. Leon-Gonzalez, and R. W. Strachan (2009): "On the evolution of the monetary policy transmission mechanism," *Journal of Economic Dynamics and Control*, 33, 997–1017.
- Koop, G., M. H. Pesaran, and S. M. Potter (1996): "Impulse response analysis in non-linear multivariate models," *Journal of Econometrics*, 74, 119–147.
- Koop, G., R. Strachan, H. van Dijk, and M. Villani (2005): "Bayesian approaches to cointegration," in *Palgrave Handbook of Econometrics, Volume 1: Econometric Theory*, ed. by K. Patterson and T. C. Mills, pp. 871–898. Palgrave Macmillan, Hounds-mills.
- Korobilis, D. (2013): "VAR forecasting using Bayesian variable selection," *Journal of Applied Econometrics*, 28, 204–230.
- Kose, M. A., C. Otrok, and C. H. Whiteman (2003): "International business cycles: World, region, and country-specific factors," *American Economic Review*, 93, 1216–1239.
- (2008): "Understanding the evolution of world business cycles," *Journal of International Economics*, 75, 110–130.
- Kozicki, S. (1999): "Multivariate detrending under common trend restrictions: Implications for business cycle research," *Journal of Economic Dynamics and Control*, 23, 997–1028.
- Kozicki, S., and P. Tinsley (2009): "Perhaps the FOMC did what it said it did: An alternative interpretation of the Great Inflation," *Journal of Monetary Economics*, 56, 842–855.

- Krämer, W., W. Ploberger, and R. Alt (1988): “Testing for structural change in dynamic models,” *Econometrica*, 56, 1355–1369.
- Kreiss, J.-P., E. Paparoditis, and D. N. Politis (2011): “On the range of validity of the autoregressive sieve bootstrap,” *Annals of Statistics*, 39, 2103–2011.
- Krolzig, H.-M. (1997): *Markov-Switching Vector Autoregressions: Modelling, Statistical Inference, and Application to Business Cycle Analysis*. Springer-Verlag, Berlin.
- Kulikov, D., and A. Netšunajev (2013): “Identifying monetary policy shocks via heteroskedasticity: A Bayesian approach,” Bank of Estonia Working Papers WP2013-9, Bank of Estonia.
- Künsch, H. R. (1989): “The jackknife and the bootstrap for general stationary observations,” *Annals of Statistics*, 17, 1217–1241.
- Kurmann, A., and E. Mertens (2014): “Stock prices, news, and economic fluctuations: Comment,” *American Economic Review*, 104, 1439–1445.
- Kuttner, K. (2001): “Monetary policy surprises and interest rates: Evidence from the Fed funds futures market,” *Journal of Monetary Economics*, 47, 523–544.
- Kuznets, S. S. (1961): *Capital in the American Economy: Its Formation and Financing*. National Bureau of Economic Research, New York.
- Kydland, F. E., and E. C. Prescott (1982): “Time to build and aggregate fluctuations,” *Econometrica*, 50, 1345–1370.
- Lanne, M., and H. Lütkepohl (2008): “Identifying monetary policy shocks via changes in volatility,” *Journal of Money, Credit and Banking*, 40, 1131–1149.
- (2010): “Structural vector autoregressions with nonnormal residuals,” *Journal of Business and Economic Statistics*, 28, 159–168.
- (2014): “A statistical comparison of alternative identification schemes for monetary policy shocks,” in *Contributions to Mathematics, Statistics, Econometrics, and Finance – Essays in Honour of Professor Seppo Pynnönen*, ed. by J. Knif, and B. Pape, Acta Wasenia 296, Statistics 7, pp. 137–152. University of Vaasa, Vaasa, Finland.
- Lanne, M., H. Lütkepohl, and K. Maciejowska (2010): “Structural vector autoregressions with Markov switching,” *Journal of Economic Dynamics and Control*, 34, 121–131.
- Lanne, M., M. Meitz, and P. Saikkonen (2017): “Identification and estimation of non-Gaussian structural vector autoregressions,” *Journal of Econometrics*, 196, 288–304.
- Lanne, M., and P. Saikkonen (2007): “A multivariate generalized orthogonal factor GARCH model,” *Journal of Business and Economic Statistics*, 25, 61–75.
- (2011): “Noncausal autoregressions for economic time series,” *Journal of Time Series Econometrics*, 3, Article 2.
- (2013): “Noncausal vector autoregression,” *Econometric Theory*, 29, 447–481.
- Leamer, E. E. (1978): *Specification Searches: Ad hoc Inference with Nonexperimental Data*. Wiley-Interscience, New York.
- (1985): “Vector autoregressions for causal inference?” *Carnegie-Rochester Conference Series on Public Policy*, 22, 255–304.
- Lee, B.-S. (1995): “Fundamentals and bubbles in asset prices: Evidence from US and Japanese asset prices,” *Asia-Pacific Financial Markets*, 2, 89–122.
- Lee, H. S. (1992): “Maximum likelihood inference and cointegration,” *Journal of Econometrics*, 54, 1–47.

- Leeb, H., and B. M. Pötscher (2005): “Model selection and inference: Facts and fiction,” *Econometric Theory*, 21, 21–59.
- (2006): “Can one estimate the conditional distribution of post-model selection estimators?” *Annals of Statistics*, 34, 2554–2591.
- Leeper, E. M. (1997): “Narrative and VAR approaches to monetary policy: Common identification problems,” *Journal of Monetary Economics*, 40, 641–657.
- Leeper, E. M., and C. A. Sims (1994): “Toward a modern macro model usable for policy analysis,” *NBER Macroeconomics Annual*, 9, 81–118.
- Leeper, E. M., C. A. Sims, and T. Zha (1996): “What does monetary policy do?” *Brookings Papers on Economic Activity*, 2, 1–63.
- Leeper, E. M., T. B. Walker, and S.-C. S. Yang (2013): “Fiscal foresight and information flows,” *Econometrica*, 81, 1115–1145.
- Leeper, E. M., and T. Zha (2003): “Modest policy interventions,” *Journal of Monetary Economics*, 50, 1673–1700.
- León, C. A., J.-C. Massé, and L.-P. Rivest (2006): “A statistical model for random rotations,” *Journal of Multivariate Analysis*, 97, 412–430.
- Lewis, R., and G. C. Reinsel (1985): “Prediction of multivariate time series by autoregressive model fitting,” *Journal of Multivariate Analysis*, 16, 393–411.
- Li, W. K., and A. I. McLeod (1981): “Distribution of the residual autocorrelations in multivariate ARMA time series models,” *Journal of the Royal Statistical Society B*, 43, 231–239.
- Lippi, M., and L. Reichlin (1993): “The dynamic effects of aggregate demand and supply disturbances: Comment,” *American Economic Review*, 83, 644–652.
- (1994): “VAR analysis, nonfundamental representations, Blaschke matrices,” *Journal of Econometrics*, 63, 307–325.
- Litterman, R. B. (1986): “Forecasting with Bayesian vector autoregressions – Five years of experience,” *Journal of Business and Economic Statistics*, 4, 25–38.
- Liu, P., and T. Konstantinos (2012): “DSGE model restrictions for structural VAR identification,” *International Journal of Central Banking*, 8, 61–95.
- Liu, R. Y. (1988): “Bootstrap procedure under some non-i.i.d. models,” *Annals of Statistics*, 16, 1696–1708.
- Liu, R. Y., J. M. Parelius, and K. Singh (1999): “Multivariate analysis by data depth: Descriptive statistics, graphics, and inference,” *Annals of Statistics*, 27, 783–858.
- Lomnicki, Z. A. (1961): “Tests for departure from normality in the case of linear stochastic processes,” *Metrika*, 4, 37–62.
- Lucas, R. E. J. (1976): “Econometric policy evaluation: A critique,” *Carnegie-Rochester Conference Series on Public Policy*, 1, 19–46.
- Lunsford, K. G. (2015a): “Identifying structural VARs with a proxy variable and a test for a weak proxy,” working paper, Federal Reserve Bank of Cleveland.
- (2015b): “Inference in proxy VARs,” working paper, Federal Reserve Bank of Cleveland.
- Lütkepohl, H. (1982a): “Non-causality due to omitted variables,” *Journal of Econometrics*, 19, 367–378.
- (1982b): “Differencing multiple time series: Another look at Canadian money and income data,” *Journal of Time Series Analysis*, 3, 235–243.
- (1985): “Comparison of criteria for estimating the order of a vector autoregressive process,” *Journal of Time Series Analysis*, 6, 35–52.

- (1987): *Forecasting Aggregated Vector ARMA Processes*. Springer-Verlag, Berlin.
 - (1988): “Asymptotic distribution of the moving-average coefficients of an estimated vector autoregressive process,” *Econometric Theory*, 4, 77–85.
 - (1990): “Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models,” *Review of Economics and Statistics*, 72, 116–125.
 - (1991): *Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.
 - (1993): “Testing for causation between two variables in higher dimensional VAR models,” in *Studies in Applied Econometrics*, ed. by H. Schneeweiss and K. F. Zimmermann, pp. 75–91. Physica-Verlag, Heidelberg.
 - (1996): *Handbook of Matrices*. John Wiley & Sons, Chichester.
 - (2004): “Vector autoregressive and vector error correction models,” in *Applied Time Series Econometrics*, ed. by H. Lütkepohl and M. Krätsig, pp. 86–158. Cambridge University Press, Cambridge.
 - (2005): *New Introduction to Multiple Time Series Analysis*. Springer-Verlag, Berlin.
 - (2006): “Vector autoregressive models,” in *Palgrave Handbook of Econometrics, Volume 1, Econometric Theory*, ed. by T. C. Mills and K. Patterson, pp. 477–510. Palgrave Macmillan, Basingstoke.
 - (2008): “Problems related to over-identifying restrictions for structural vector error correction models,” *Economics Letters*, 99, 512–515.
 - (2009): “Econometric analysis with vector autoregressive models,” in *Handbook of Computational Econometrics*, ed. by D. A. Belsley and E. J. Kontoghiorghes, pp. 281–319. Wiley, Chichester.
 - (2013): “Vector autoregressive models,” in *Handbook of Research Methods and Applications on Empirical Macroeconomics*, ed. by N. Hashimzade and M. Thornton, pp. 139–164. Edward Elger, Cheltenham.
 - (2013a): “Reducing confidence bands for simulated impulse responses,” *Statistical Papers*, 54, 1131–1145.
 - (2013b): “Identifying structural vector autoregressions via changes in volatility,” *Advances in Econometrics*, 32, 169–203.
 - (2014): “Fundamental problems with nonfundamental shocks,” in *Essays in Non-linear Time Series Econometrics*, ed. by N. Haldrup, M. Meitz, and P. Saikkonen, chap. 8, pp. 198–214. Oxford University Press, Oxford.
- Lütkepohl, H., and G. Milunovich (2016): “Testing for identification in SVAR-GARCH models,” *Journal of Economic Dynamics and Control*, 73, 241–258.
- Lütkepohl, H., and A. Netšunajev (2014): “Disentangling demand and supply shocks in the crude oil market: How to check sign restrictions in structural VARs,” *Journal of Applied Econometrics*, 29, 479–496.
- (2015): “Structural vector autoregressions with smooth transition in variances: The interaction between U.S. monetary policy and the stock market,” working paper, DIW Berlin.
 - (2017): “Structural vector autoregressions with heteroskedasticity: A review of different volatility models,” *Econometrics and Statistics*, 1, 2–18.
- Lütkepohl, H., and D. S. Poskitt (1991): “Estimating orthogonal impulse responses via vector autoregressive models,” *Econometric Theory*, 7, 487–496.

- (1996): “Testing for causation using infinite order vector autoregressive processes,” *Econometric Theory*, 12, 61–87.
- Lütkepohl, H., and H.-E. Reimers (1992a): “Impulse response analysis of cointegrated systems,” *Journal of Economic Dynamics and Control*, 16, 53–78.
- (1992b): “Granger-causality in cointegrated VAR processes: The case of the term structure,” *Economics Letters*, 40, 263–268.
- Lütkepohl, H., and P. Saikkonen (1999): “Order selection in testing for the cointegrating rank of a VAR process,” in *Cointegration, Causality, and Forecasting. A Festschrift in Honour of Clive W.J. Granger*, ed. by R. F. Engle, and H. White, pp. 168–199. Oxford University Press, Oxford.
- Lütkepohl, H., P. Saikkonen, and C. Trenkler (2004): “Testing for the cointegrating rank of a VAR process with level shift at unknown time,” *Econometrica*, 72, 647–662.
- Lütkepohl, H., A. Staszewska-Bystrova, and P. Winker (2015a): “Comparison of methods for constructing joint confidence bands for impulse response functions,” *International Journal of Forecasting*, 31, 782–798.
- (2015b): “Confidence bands for impulse responses: Bonferroni versus Wald,” *Oxford Bulletin of Economics and Statistics*, 77, 800–821.
- Lütkepohl, H., and A. Velinov (2016): “Structural vector autoregressions: Checking identifying long-run restrictions via heteroskedasticity,” *Journal of Economic Surveys*, 30, 377–392.
- Magdalinos, T., and P. C. B. Phillips (2009): “Limit theory for cointegrated systems with moderately integrated and moderately explosive regressors,” *Econometric Theory*, 25, 482–526.
- Magnus, J. R., and H. Neudecker (1988): *Matrix Differential Calculus with Applications in Statistics and Econometrics*. John Wiley, Chichester.
- Mammen, E. (1993): “Bootstrap and wild bootstrap for high dimensional linear models,” *Annals of Statistics*, 21, 255–285.
- Mann, H. B., and A. Wald (1943): “On the statistical treatment of linear stochastic difference equations,” *Econometrica*, 11, 173–220.
- Mansour, J. M. (2003): “Do national business cycles have an international origin?” *Empirical Economics*, 28, 223–247.
- Marcellino, M., J. H. Stock, and M. W. Watson (2006): “A comparison of direct and iterated multistep AR methods for forecasting macroeconomic time series,” *Journal of Econometrics*, 135, 499–526.
- Mertens, E. (2012): “Are spectral estimators useful for long-run restrictions in SVARs?” *Journal of Economic Dynamics and Control*, 36, 1831–1844.
- Mertens, K., and M. O. Ravn (2012): “Empirical evidence on the aggregate effects of anticipated and unanticipated U.S. tax policy shocks,” *American Economic Journal: Economic Policy*, 4, 145–181.
- (2014): “A reconciliation of SVAR and narrative estimates of tax multipliers,” *Journal of Monetary Economics*, 68, S1–S19.
- Mikusheva, A. (2007): “Uniform inference in autoregressive models,” *Econometrica*, 75, 1411–1452.
- (2012): “One-dimensional inferences in autoregressive models with the potential presence of a unit root,” *Econometrica*, 80, 173–212.

- Milunovich, G., and M. Yang (2013): "On identifying structural VAR models via ARCH effects," *Journal of Time Series Econometrics*, 5, 117–131.
- Moench, E., and S. Ng (2011): "A hierarchical factor analysis of U.S. housing market dynamics," *Econometrics Journal*, 14, C1–C24.
- Moneta, A. (2008): "Graphical causal models and VARs: An empirical assessment of the real business cycles hypothesis," *Empirical Economics*, 35, 275–300.
- Moneta, A., D. Entner, P. O. Hoyer, and A. Coad (2013): "Causal inference by independent component analysis: Theory and applications," *Oxford Bulletin of Economics and Statistics*, 75, 705–730.
- Montiel Olea, J. L., J. H. Stock, and M. W. Watson (2015a): "Inference in structural VARs with external instruments," working paper, New York University.
- (2015b): "Uniform inference in SVARs identified with external instruments," working paper, New York University.
- Moon, H. R., and F. Schorfheide (2012): "Bayesian and frequentist inference in partially identified models," *Econometrica*, 80, 755–782.
- Moon, H. R., F. Schorfheide, and E. Granziera (2013): "Inference for VARs identified with sign restrictions," working paper, University of Pennsylvania.
- Mork, K. A. (1989): "Oil and the macroeconomy. When prices go up and down: An extension of Hamilton's results," *Journal of Political Economy*, 97, 740–744.
- Mountford, A., and H. Uhlig (2009): "What are the effects of fiscal policy shocks?" *Journal of Applied Econometrics*, 24, 960–992.
- Müller, U. K. (2005): "Size and power of tests of stationarity in highly autocorrelated time series," *Journal of Econometrics*, 128, 195–213.
- Mumtaz, H., and P. Surico (2009): "The transmission of international shocks: A factor-augmented VAR approach," *Journal of Money, Credit and Banking*, 41, 71–100.
- Nelson, C. R., and C. I. Plosser (1982): "Trends and random walks in macroeconomic time series: Some evidence and implications," *Journal of Monetary Economics*, 10, 139–162.
- Netšunajev, A. (2013): "Reaction to technology shocks in Markov-switching structural VARs: Identification via heteroskedasticity," *Journal of Macroeconomics*, 36, 51–62.
- Ng, S., and P. Perron (2005): "A note on the selection of time series models," *Oxford Bulletin of Economics and Statistics*, 67, 115–134.
- Nicholls, D. F., and A. R. Pagan (1983): "Heteroskedasticity in models with lagged dependent variables," *Econometrica*, 51, 1233–1242.
- Nicholls, D. F., and A. L. Pope (1988): "Bias in estimation of multivariate autoregression," *Australian Journal of Statistics*, 30A, 296–309.
- Nickelsburg, G. (1985): "Small-sample properties of dimensionality statistics for fitting VAR models to aggregate economic data: A Monte Carlo study," *Journal of Econometrics*, 28, 183–192.
- Normandin, M., and L. Phaneuf (2004): "Monetary policy shocks: Testing identification conditions under time-varying conditional volatility," *Journal of Monetary Economics*, 51, 1217–1243.
- Nyblom, J. (1989): "Testing for the constancy of parameters over time," *Journal of the American Statistical Association*, 84, 223–230.

- Oliner, S. D., and G. D. Rudebusch (1995): “Is there a broad credit channel for monetary policy?” *Federal Reserve Bank of San Francisco Review*, 1, 3–13.
- Onatski, A. (2009): “Testing hypotheses about the number of factors in large factor models,” *Econometrica*, 77, 1447–1479.
- Onatski, A., and H. Uhlig (2012): “Unit roots in white noise,” *Econometric Theory*, 28, 485–508.
- Osborn, D. R. (1991): “The implications of periodically varying coefficients for seasonal time-series processes,” *Journal of Econometrics*, 48, 373–384.
- (1993): “Discussion: Seasonal cointegration: The Japanese consumption function,” *Journal of Business and Economic Statistics*, 21, 547–563.
- (2002): “Unit-root versus deterministic representation of seasonality for forecasting,” in *A Companion to Economic Forecasting*, ed. by M. P. Clements and D. F. Hendry, pp. 409–431. Blackwell, Oxford.
- Osterwald-Lenum, M. (1992): “A note with fractiles of the asymptotic distribution of the maximum likelihood cointegration rank test statistics: Four cases,” *Oxford Bulletin of Economics and Statistics*, 54, 461–472.
- Otrok, C., and C. H. Whiteman (1998): “Bayesian leading indicators: Measuring and predicting economic conditions in Iowa,” *International Economic Review*, 39, 997–1014.
- Ouliaris, S., and A. R. Pagan (2016): “A method for working with sign restrictions in structural equation modelling,” *Oxford Bulletin of Economics and Statistics*, 78, 605–622.
- Ouliaris, S., A. Pagan, and J. Restrepo (2014): “Analysing two agnostic techniques for recovering impulse response functions from structural vector autoregressions,” working paper, University of Sydney.
- Pagan, A. (1984): “Econometric issues in the analysis of regressions with generated regressors,” *International Economic Review*, 25, 221–248.
- Pagan, A., and M. H. Pesaran (2008): “Econometric analysis of structural systems with permanent and transitory shocks,” *Journal of Economic Dynamics and Control*, 32, 3376–3395.
- Pagan, A., and T. Robinson (2016): “Investigating the relationship between DSGE and SVAR models,” working paper, University of Melbourne.
- Pagan, A. R., and J. C. Robertson (1998): “Structural models of the liquidity effect,” *Review of Economics and Statistics*, 80, 202–217.
- Paparoditis, E. (1996): “Bootstrapping autoregressive and moving average parameter estimates of infinite order vector autoregressive processes,” *Journal of Multivariate Analysis*, 57, 277–296.
- Park, J. Y., and P. C. B. Phillips (1988): “Statistical inference in regressions with integrated processes: Part 1,” *Econometric Theory*, 4, 468–497.
- (1989): “Statistical inference in regressions with integrated processes: Part 2,” *Econometric Theory*, 5, 95–131.
- Pascual, L., J. Romo, and E. Ruiz (2004): “Bootstrap predictive inference for ARIMA processes,” *Journal of Time Series Analysis*, 25, 449–465.
- Paulsen, J. (1984): “Order determination of multivariate autoregressive time series with unit roots,” *Journal of Time Series Analysis*, 5, 115–127.

- Paulsen, J., and D. Tjøstheim (1985): "On the estimation of residual variance and order in autoregressive time series," *Journal of the Royal Statistical Society B*, 36, 105–117.
- Paustian, M. (2007): "Assessing sign restrictions," *B.E. Journal of Macroeconomics*, 7, Article 23.
- Peña, D., and G. E. P. Box (1987): "Identifying a simplifying structure in time series," *Journal of the American Statistical Association*, 82, 836–843.
- Pearl, J. (2000): *Causality: Models, Reasoning, and Inference*. Cambridge University Press, Cambridge.
- Peersman, G. (2005): "What caused the millennium slowdown? Evidence based on autoregressions," *Journal of Applied Econometrics*, 20, 185–207.
- Pesaran, M. H. (2015): *Time Series and Panel Data Econometrics*. Oxford University Press, Oxford.
- Pesaran, M. H., A. Pick, and A. Timmermann (2011): "Variable selection, estimation and inference for multi-period forecasting problems," *Journal of Econometrics*, 164, 173–187.
- Pesaran, M. H., T. Schuermann, and S. M. Weiner (2004): "Modeling regional interdependencies using a global error-correcting macroeconomic model," *Journal of Business and Economic Statistics*, 22, 129–162.
- Pesaran, M. H., and Y. Shin (1998): "Generalized impulse response analysis in linear multivariate models," *Economics Letters*, 58, 17–29.
- Pesavento, E., and B. Rossi (2006): "Small-sample confidence intervals for multivariate impulse response functions at long horizons," *Journal of Applied Econometrics*, 21, 1135–1155.
- Phillips, P. C. B. (1991): "To criticize the critics: An objective Bayesian analysis of stochastic trends," *Journal of Applied Econometrics*, 6, 333–364.
- (1998): "Impulse response and forecast error variance asymptotics in nonstationary VARs," *Journal of Econometrics*, 83, 21–56.
- (2014): "On confidence intervals for autoregressive roots and predictive regression," *Econometrica*, 82, 1177–1195.
- Phillips, P. C. B., and W. Ploberger (1996): "An asymptotic theory of Bayesian inference for time series," *Econometrica*, 64, 381–412.
- Piazzesi, M., and E. Swanson (2008): "Futures prices as risk-adjusted forecasts of monetary policy," *Journal of Monetary Economics*, 55, 677–691.
- Piffer, M., and M. Podstawska (2016): "Identifying uncertainty shocks using the price of gold," DIW Discussion Paper 1549.
- Plagborg-Møller, M. (2016): "Bayesian inference on structural impulse response functions," working paper, Harvard University.
- Pope, A. L. (1990): "Biases of estimators in multivariate non-Gaussian autoregressions," *Journal of Time Series Analysis*, 11, 249–258.
- Poskitt, D. S., and Yao, W. (2017), Vector autoregressions and macroeconomic modelling: An error taxonomy, *Journal of Business and Economic Statistics*, 35, 407–419.
- Prescott, E. C. (1986): "Theory ahead of business cycle measurement," *Carnegie-Rochester Conference Series on Public Policy*, 24, 11–44.

- Primiceri, G. E. (2005): "Time varying structural vector autoregressions and monetary policy," *Review of Economic Studies*, 72, 821–852.
- Psaradakis, Z., and N. Spagnolo (2003): "On the determination of the number of regimes in Markov-switching autoregressive models," *Journal of Time Series Analysis*, 24, 237–252.
- (2006): "Joint determination of the state dimension and autoregressive order for models with Markov regime switching," *Journal of Time Series Analysis*, 27, 753–766.
- Qu, Z. (2014): "Inference in dynamic stochastic general equilibrium models with possible weak identification," *Quantitative Economics*, 5, 457–494.
- Quinn, B. G. (1980): "Order determination for a multivariate autoregression," *Journal of the Royal Statistical Society B*, 42, 182–185.
- Ramey, V. A. (2011): "Identifying government spending shocks: It's all in the timing," *Quarterly Journal of Economics*, 126, 1–50.
- Ramey, V. A., and M. D. Shapiro (1998): "Costly capital reallocation and the effects of government spending," *Carnegie-Rochester Conference Series on Public Policy*, 48, 145–194.
- Ramey, V. A., and D. J. Vine (2011): "Oil, automobiles and the U.S. economy: How much have things really changed?" *NBER Macroeconomics Annual*, 25, 333–367.
- Rapach, D. E. (2001): "Macro shocks and real stock prices," *Journal of Economics and Business*, 53, 5–26.
- Ravenna, F. (2007): "Vector autoregressions and reduced form representations of DSGE models," *Journal of Monetary Economics*, 54, 2048–2064.
- Ravn, M. O., and H. Uhlig (2002): "On adjusting the Hodrick-Prescott filter for the frequency of observations," *Review of Economics and Statistics*, 84, 371–376.
- Reinsel, G. (1980): "Asymptotic properties of prediction errors for the multivariate autoregressive model using estimated parameters," *Journal of the Royal Statistical Society B*, 42, 328–333.
- (1983): "Some results on multivariate autoregressive index models," *Biometrika*, 70, 145–156.
- (1993): *Elements of Multivariate Time Series Analysis*. Springer-Verlag, New York.
- Reinsel, G. C., and R. P. Velu (1998): *Multivariate Reduced-Rank Regression, Theory and Applications*. Springer-Verlag, New York.
- Rigobon, R. (2003): "Identification through heteroskedasticity," *Review of Economics and Statistics*, 85, 777–792.
- Rigobon, R., and B. Sack (2003): "Measuring the reaction of monetary policy to the stock market," *Quarterly Journal of Economics*, 118, 639–669.
- (2004): "The impact of monetary policy on asset prices," *Journal of Monetary Economics*, 51, 1553–1575.
- Rissanen, J. (1978): "Modeling by shortest data description," *Automatica*, 14, 465–471.
- (1986): "Stochastic complexity and modeling," *Annals of Statistics*, 14, 1080–1100.
- Robertson, J. C., and E. W. Tallman (2001): "Improving federal funds rate forecasts in VAR models used for policy analysis," *Journal of Business and Economic Statistics*, 19, 324–330.
- Robinson, P. M. (1983): "Non-parametric estimation for time series models," *Journal of Time Series Analysis*, 4, 185–208.

- Rogers, J. H., C. Scotti, and J. H. Wright (2015): "Unconventional monetary policy and international risk premia," working paper, Federal Reserve Board.
- Romer, C. D., and D. H. Romer (1989): "Does monetary policy matter? A new test in the spirit of Friedman and Schwartz," *NBER Macroeconomics Annual*, 4, 121–170.
- (2004): "A new measure of monetary shocks: Derivation and implications," *American Economic Review*, 94, 1055–1084.
- (2010): "The macroeconomic effects of tax changes: Estimates based on a new measure of fiscal shocks," *American Economic Review*, 100, 763–801.
- Rotemberg, J., and M. Woodford (1996): "Imperfect competition and the effects of energy price increases on economic activity," *Journal of Money, Credit and Banking*, 28, 550–577.
- (1997): "An optimization-based econometric framework for the evaluation of monetary policy," *NBER Macroeconomics Annual*, 12, 297–346.
- Rothman, P., D. van Dijk, and P. H. Franses (2001): "Multivariate STAR analysis of money-output relationship," *Macroeconomic Dynamics*, 5, 506–532.
- Rozanov, Y. A. (1967): *Stationary Random Processes*. Holden-Day, San Francisco.
- Rubio-Ramírez, J. F., D. Waggoner, and T. Zha (2005): "Markov-switching structural vector autoregressions: Theory and applications," working paper, Federal Reserve Bank of Atlanta.
- (2010): "Structural vector autoregressions: Theory of identification and algorithms for inference," *Review of Economic Studies*, 77, 665–696.
- Rudebusch, G. (1998): "Do measures of monetary policy in a VAR make sense?" *International Economic Review*, 39, 907–931.
- Ruge-Murcia, F. (2007): "Methods to estimate dynamic stochastic general equilibrium models," *Journal of Economic Dynamics and Control*, 31, 2599–2636.
- Runkle, D. E. (1987): "Vector autoregressions and reality," *Journal of Business and Economic Statistics*, 5, 437–442.
- Saikkonen, P. (1992): "Estimation and testing of cointegrated systems by an autoregressive approximation," *Econometric Theory*, 8, 1–27.
- Saikkonen, P., and H. Lütkepohl (1996): "Infinite-order cointegrated vector autoregressive processes: Estimation and inference," *Econometric Theory*, 12, 814–844.
- (2000a): "Testing for the cointegrating rank of a VAR process with structural shifts," *Journal of Business and Economic Statistics*, 18, 451–464.
- (2000b): "Trend adjustment prior to testing for the cointegrating rank of a vector autoregressive process," *Journal of Time Series Analysis*, 21, 435–456.
- Saikkonen, P., H. Lütkepohl, and C. Trenkler (2006): "Break date estimation for VAR processes with level shift with an application to cointegration testing," *Econometric Theory*, 22, 15–68.
- Salmon, M. (1982): "Error correction mechanisms," *Economic Journal*, 92, 615–629.
- Samaranayake, V. A., and D. P. Hasza (1988): "Properties of predictors for multivariate autoregressive models with estimated parameters," *Journal of Time Series Analysis*, 9, 361–383.
- Sargan, J. D. (1964): "Wages and prices in the United Kingdom: A study in econometric methodology," in *Econometric Analysis for National Economic Planning*, ed. by P. E. Hart, G. Mills, and J. K. Whitaker, pp. 25–63. Butterworths, London.

- Sargent, T. J., and C. A. Sims (1977): “Business cycle modeling without pretending to have too much a priori economic theory,” in *New Methods in Business Cycle Research: Proceedings from a Conference*, ed. by C. A. Sims, pp. 45–109. Federal Reserve Bank of Minneapolis, Minneapolis.
- Sbrana, G., and A. Silvestrini (2009): “What do we know about comparing aggregate and disaggregate forecasts?” CORE Discussion Paper No. 2009/20, Université Catholique de Louvain.
- Scheffé, H. (1953): “A method for judging all contrasts in the analysis of variance,” *Biometrika*, 40, 87–104.
- Scholl, A., and H. Uhlig (2008): “New evidence on the puzzles: Results from agnostic identification on monetary policy and exchange rates,” *Journal of International Economics*, 76, 1–13.
- Schorfheide, F. (2005): “VAR forecasting under misspecification,” *Journal of Econometrics*, 128, 99–136.
- Schwarz, G. (1978): “Estimating the dimension of a model,” *Annals of Statistics*, 6, 461–464.
- Sentana, E., and G. Fiorentini (2001): “Identification, estimation and testing of conditionally heteroskedastic factor models,” *Journal of Econometrics*, 102, 143–164.
- Serfling, R. J. (1980): *Approximation Theorems of Mathematical Statistics*. John Wiley, New York.
- Shaman, P., and R. A. Stine (1988): “The bias of autoregressive coefficient estimators,” *Journal of the American Statistical Association*, 83, 842–848.
- Shapiro, M. D. (1994): “Federal Reserve policy: Cause and effect,” in *Monetary Policy*, ed. by N. Mankiw, pp. 307–334. University of Chicago Press, Chicago.
- Shapiro, M. D., and M. W. Watson (1988): “Sources of business cycle fluctuations,” *NBER Macroeconomics Annual*, 3, 111–148.
- Shimmura, T., and T. Yoshimura (2013): “Circadian clock determines the timing of rooster crowing,” *Current Biology*, 23, R231–R233.
- Siegfried, N. A. (2002): “An information-theoretic extension to structural VAR modelling,” working paper, Hamburg University.
- Sims, C. A. (1972): “Money, income and causality,” *American Economic Review*, 62, 540–552.
- (1980a): “Macroeconomics and reality,” *Econometrica*, 48, 1–48.
- (1980b): “Comparison of interwar and postwar business cycles: Monetarism reconsidered,” *American Economic Review*, 70, 250–257.
- (1986): “Are forecasting models usable for policy analysis?” *Minneapolis Federal Reserve Bank Quarterly Review*, 10, 2–16.
- (1987): “Comment on ‘Vector Autoregressions and Reality’ by Runkle,” *Journal of Business and Economic Statistics*, 5, 443–449.
- (1988): “Bayesian skepticism on unit root econometrics,” *Journal of Economic Dynamics and Control*, 12, 463–474.
- (1989): “Models and their uses,” *American Journal of Agricultural Economics*, 71, 75–102.
- (1992): “Interpreting the macroeconomic time series facts: The effects of monetary policy,” *European Economic Review*, 36, 975–1011.

- (1993): “A nine-variable probabilistic macroeconomic forecasting model,” in *Business Cycles, Indicators and Forecasting*, ed. by J. Stock, and M. Watson. NBER Studies in Business Cycles Volume 28, 179–214.
- (1998): “Comment on Glenn Rudebusch’s ‘Do Measures of Monetary Policy in a VAR Make Sense?’” *International Economic Review*, 39, 933–941.
- (2007): “Comment on ‘On the Fit of New Keynesian Models’ by Del Negro, Schorfheide, Smets and Wouters,” *Journal of Business and Economic Statistics*, 25, 152–154.
- Sims, C. A., J. H. Stock, and M. W. Watson (1990): “Inference in linear time series models with some unit roots,” *Econometrica*, 58, 113–144.
- Sims, C. A., and H. Uhlig (1991): “Understanding unit rooters: A helicopter tour,” *Econometrica*, 59, 1591–1600.
- Sims, C. A., D. F. Waggoner, and T. Zha (2008): “Methods for inference in large multiple-equation Markov-switching models,” *Journal of Econometrics*, 146, 255–274.
- Sims, C. A., and T. Zha (1998): “Bayesian methods for dynamic multivariate models,” *International Economic Review*, 39, 949–968.
- (1999): “Error bands for impulse responses,” *Econometrica*, 67, 1113–1155.
- (2006a): “Does monetary policy generate recessions?” *Macroeconomic Dynamics*, 10, 231–272.
- (2006b): “Were there regime switches in U.S. monetary policy?” *American Economic Review*, 96, 54–81.
- Sims, E. R. (2012): “News, non-invertibility, and structural VARs,” *Advances in Econometrics*, 28, 81–135.
- Sin, C.-Y., and H. White (1996): “Information criteria for selecting possibly misspecified parametric models,” *Journal of Econometrics*, 71, 207–225.
- Singleton, K. J. (1988): “Econometric issues in the analysis of equilibrium business cycle models,” *Journal of Monetary Economics*, 21, 361–386.
- Smets, F., and R. Wouters (2007): “Shocks and frictions in US business cycles: A Bayesian DSGE approach,” *American Economic Review*, 97, 586–606.
- Song, S., and P. J. Bickel (2011): “Large vector auto regressions,” SFB 649 Discussion Paper No. 2011-048, Humboldt University.
- Spearman, C. (1904): “‘General-intelligence,’ objectively determined and measured,” *American Journal of Psychology*, 15, 201–293.
- Spirites, P., C. Glymour, and R. Scheines (2000): *Causation, Prediction, and Search*, 2nd ed. MIT Press, Cambridge, MA.
- Staszewska, A. (2007): “Representing uncertainty about impulse response paths: The use of heuristic optimization methods,” *Computational Statistics and Data Analysis*, 52, 121–132.
- Stewart, G. W. (1980): “The efficient generation of random orthogonal matrices with an application to condition estimators,” *SIAM Journal on Numerical Analysis*, 17, 403–409.
- Stock, J. H. (1991): “Confidence intervals for the largest autoregressive root in U.S. macroeconomic time series,” *Journal of Monetary Economics*, 28, 435–459.
- (1996): “VAR, error correction and pretest forecasts at long horizons,” *Oxford Bulletin of Economics and Statistics*, 58, 685–701.

- Stock, J. H., and M. W. Watson (2002a): "Forecasting using principal components from a large number of predictors," *Journal of the American Statistical Association*, 97, 1167–1179.
- (2002b): "Macroeconomic forecasting using diffusion indexes," *Journal of Business and Economic Statistics*, 20, 147–162.
- (2005): "Implications of dynamic factor models for VAR analysis," working paper, Princeton University.
- (2006): "Forecasting with many predictors," in *Handbook of Economic Forecasting*, ed. by G. Elliott, C. W. J. Granger, and A. Timmermann, pp. 515–554. Elsevier, Amsterdam.
- (2009): "Forecasting in dynamic factor models subject to structural instability," in *The Methodology and Practice of Econometrics: A Festschrift in Honour of David F. Hendry*, ed. by J. Castle and N. Shephard, pp. 1–57, chap. 7. Oxford University Press, Oxford.
- (2011): "Dynamic factor models," in *Oxford Handbook of Economic Forecasting*, ed. by M. P. Clements and D. F. Hendry, pp. 35–60. Oxford University Press, Oxford.
- (2012): "Disentangling the channels of the 2007–09 recession," *Brookings Papers on Economic Activity*, Spring, 81–135.
- (2016): "Factor models and structural vector autoregressions in macroeconomics," in *Handbook of Macroeconomics*, ed. by J. B. Taylor and H. Uhlig, vol. 2, pp. 415–525. Elsevier, Amsterdam.
- Stone, J. V. (2004): *Independent Component Analysis: A Tutorial Introduction*. MIT Press, Cambridge, MA.
- Strachan, R. W. (2003): "Valid Bayesian estimation of the cointegrating error correction model," *Journal of Business and Economic Statistics*, 21, 185–195.
- Strachan, R. W., and B. Inder (2004): "Bayesian analysis of the error correction model," *Journal of Econometrics*, 123, 307–325.
- Strohsal, T., and E. Weber (2015): "Time-varying international stock market interaction and the identification of volatility signals," *Journal of Banking and Finance*, 56, 28–36.
- Strongin, S. (1995): "The identification of monetary policy disturbances: Explaining the liquidity puzzle," *Journal of Monetary Economics*, 35, 463–498.
- Swanson, N. R., and C. W. J. Granger (1997): "Impulse response functions based on a causal approach of residual orthogonalization in vector autoregressions," *Journal of the American Statistical Association*, 92, 357–367.
- Taylor, J. (1993): "Discretion versus policy rules in practice," *Carnegie-Rochester Conference Series on Public Policy*, 39, 195–214.
- Taylor, M. P. (2004): "Estimating structural macroeconomic shocks through long-run recursive restrictions on vector autoregressive models: The problem of identification," *International Journal of Finance and Economics*, 9, 229–244.
- Teräsvirta, T., D. Tjøstheim, and C. W. J. Granger (2010): *Modelling Nonlinear Economic Time Series*. Oxford University Press, Oxford.
- Theil, H. (1963): "On the use of incomplete prior information in regression analysis," *Journal of the American Statistical Association*, 58, 401–414.
- Tiao, G. C. (1972): "Asymptotic behaviour of temporal aggregates of time series," *Biometrika*, 59, 525–531.

- Tibshirani, R. (1996): "Regression shrinkage and selection via the lasso," *Journal of the Royal Statistical Society B*, 58, 267–288.
- Tinbergen, J. (1939): *Statistical Testing of Business Cycle Theories. Vol. II: Business Cycles in the United States of America, 1919–1932*. League of Nations, Geneva.
- Tobin, J. (1970): "Money and income: Post hoc ergo propter hoc?" *Quarterly Journal of Economics*, 84, 310–317.
- Toda, H. Y. (1994): "Finite sample properties of likelihood ratio tests for cointegrating ranks when linear trends are present," *Review of Economics and Statistics*, 76, 66–79.
- (1995): "Finite sample performance of likelihood ratio tests for cointegrating ranks in vector autoregressions," *Econometric Theory*, 11, 1015–1032.
- Toda, H. Y., and P. C. B. Phillips (1993): "Vector autoregressions and causality," *Econometrica*, 61, 1367–1393.
- Toda, H. Y., and T. Yamamoto (1995): "Statistical inference in vector autoregressions with possibly integrated processes," *Journal of Econometrics*, 66, 225–250.
- Todd, R. M. (1990): "Vector autoregression evidence on monetarism: Another look at the robustness debate," *Federal Reserve Bank of Minneapolis Quarterly Review*, Spring, 19–37.
- Tsay, R. S. (1984): "Order selection in nonstationary autoregressive models," *Annals of Statistics*, 12, 1425–1433.
- Tschernig, R. (2004): "Nonparametric time series modeling," in *Applied Time Series Econometrics*, ed. by H. Lütkepohl and M. Krätsig, chap. 7, pp. 243–288. Cambridge University Press, Cambridge.
- Tschernig, R., E. Weber, and R. Weigand (2013): "Long-run identification in a fractionally integrated system," *Journal of Business and Economic Statistics*, 31, 438–450.
- Tso, M. K.-S. (1981): "Reduced-rank regression and canonical analysis," *Journal of the Royal Statistical Society B*, 43, 183–189.
- Uhlig, H. (1994): "What macroeconomists should know about unit roots: A Bayesian perspective," *Econometric Theory*, 10, 645–671.
- (1997): "Bayesian vector autoregressions with stochastic volatility," *Econometrika*, 65, 59–73.
- (2005): "What are the effects of monetary policy on output? Results from an agnostic identification procedure," *Journal of Monetary Economics*, 52, 381–419.
- van der Weide, R. (2002): "GO-GARCH: A multivariate generalized orthogonal GARCH model," *Journal of Applied Econometrics*, 17, 549–564.
- Velinov, A. (2013): "Can stock price fundamentals really be captured? Using Markov switching in heteroskedasticity models to test identification restrictions," Discussion Paper No. 1350, DIW Berlin.
- Velinov, A., and W. Chen (2015): "Do stock prices reflect their fundamentals? New evidence in the aftermath of the financial crisis," *Journal of Economics and Business*, 80, 1–20.
- Velu, R. P., G. C. Reinsel, and D. W. Wichern (1986): "Reduced rank models for multiple time series," *Biometrika*, 73, 105–118.
- Venditti, F. (2013): "From oil to consumer energy prices: How much asymmetry along the way?" *Energy Economics*, 40, 468–473.
- Villani, M. (2009): "Steady state prior for vector autoregressions," *Journal of Applied Econometrics*, 24, 630–650.

- Waggoner, D. F., and T. Zha (1999): “Conditional forecasts in dynamic multivariate models,” *Review of Economics and Statistics*, 81, 639–651.
- (2003): “A Gibbs sampler for structural vector autoregressions,” *Journal of Economic Dynamics and Control*, 28, 349–366.
- Wallis, K. (1977): “Multiple time series analysis and the final form of econometric models,” *Econometrica*, 45, 1481–1497.
- (1980): “Econometric implications of the rational expectations hypothesis,” *Econometrica*, 48, 49–74.
- Watson, M. W. (1994): “Vector autoregressions and cointegration,” in *Handbook of Econometrics*, ed. by R. Engle and D. McFadden, vol. 4, pp. 2844–2915. Elsevier, Amsterdam.
- (2006): “Comment on ‘Assessing Structural VARs’,” *NBER Macroeconomics Annual*, 21, 97–102.
- Weber, E. (2010): “Structural conditional correlation,” *Journal of Financial Econometrics*, 8, 392–407.
- Wei, C. Z. (1992): “On predictive least squares principles,” *Annals of Statistics*, 20, 1–42.
- Wei, W. W. S. (1981): “Effect of systematic sampling on ARIMA models,” *Communications in Statistics – Theory and Methods A*, 10, 2389–2398.
- Weise, C. L. (1999): “The asymmetric effects of monetary policy: A nonlinear vector autoregression approach,” *Journal of Money, Credit and Banking*, 31, 85–108.
- Whittle, P. (1953): “The analysis of multiple stationary time series,” *Journal of the Royal Statistical Society B*, 15, 125–139.
- Wold, H. (1938): *A Study in the Analysis of Stationary Time-Series*. Almqvist and Wiksell, Uppsala.
- Wolf, M., and D. Wunderli (2015): “Bootstrap joint prediction regions,” *Journal of Time Series Analysis*, 36, 352–376.
- Woodford, M. (2003): *Interest Rates and Prices: Foundations of a Theory of Monetary Policy*. Princeton University Press, Princeton.
- Woźniak, T., and M. Droumaguet (2015): “Assessing monetary policy models: Bayesian inference for heteroskedastic structural VARs,” working paper, University of Melbourne.
- Wright, J. H. (2000): “Confidence intervals for univariate impulse responses with a near unit root,” *Journal of Business and Economic Statistics*, 18, 368–373.
- (2012): “What does monetary policy do to long-term interest rates at the zero lower bound?” *Economic Journal*, 122, F447–F466.
- Wu, C. F. J. (1986): “Jackknife, bootstrap and other resampling methods in regression analysis,” *Annals of Statistics*, 14, 1261–1295.
- Yamamoto, Y. (2016): “Bootstrap inference for impulse response functions in factor-augmented vector autoregressions,” working paper, Hitotsubashi University.
- Yao, W., T. Kam, and F. Vahid (2017): “On weak identification in structural VARMA models,” *Economics Letters*, 156, 1–6.
- Yeh, A. B. (1998): “A bootstrap procedure in linear regression with nonstationary errors,” *Canadian Journal of Statistics*, 26, 149–160.
- Zellner, A., and F. Palm (1974): “Time series analysis and simultaneous equation econometric models,” *Journal of Econometrics*, 2, 17–54.

Notation and Abbreviations

The following list specifies general notational conventions used in the book. Occasionally in the text, a symbol has a meaning that differs from the one specified in this list when confusion is unlikely.

General Symbols and Notation

=	equals
\equiv	equals by definition, is defined as
\propto	is proportional to
\Rightarrow	implies
\Leftrightarrow	is equivalent to
\sim	is distributed as
$\stackrel{a}{\sim}$	is approximately distributed in large samples
$\stackrel{iid}{\sim}$	is independently, identically distributed as
\forall	for all
\exists	there exists
\in	element of
\subset	subset of
\cup	union
\cap	intersection
\sum	summation sign
\prod	product sign
\rightarrow	converges to, approaches
\xrightarrow{p}	converges in probability to
$\xrightarrow{a.s.}$	converges almost surely to
$\xrightarrow{q.m.}$	converges in quadratic mean to
\xrightarrow{d}	converges in distribution to
$\stackrel{iid}{\sim}$	independently, identically distributed
\lim	limit
plim	probability limit

max	maximum
min	minimum
sup	supremum, least upper bound
log	natural logarithm
exp	exponential function
$ z $	absolute value or modulus of z
K	dimension of a stochastic process or time series
T	sample size, time series length
\mathbb{R}	real numbers
\mathbb{R}^m	m -dimensional Euclidean space
\mathbb{C}	complex numbers
\mathbb{Z}	integers
\mathbb{N}	positive integers
\mathbb{P}	probability
H_0	null hypothesis
H_1	alternative hypothesis
$I(\cdot)$	indicator function
L	lag operator
Δ	differencing operator
E	expectation operator
$l(\cdot)$	likelihood function
$\log l$	log-likelihood function
$[x]$	largest integer smaller or equal to $x \in \mathbb{R}$
1968m10	October 1968
1968q3	third quarter of 1968

Abbreviations

AD	aggregate demand
AIC	Akaike information criterion
AR	autoregression
ARCH	autoregressive conditional heteroskedasticity
ARMA	autoregressive moving average
AS	aggregate supply
AVAR	asymmetric vector autoregression
BVAR	Bayesian vector autoregression
Corr	correlation, correlation matrix
Cov	covariance, covariance matrix
CPI	consumer price index
CUSUM	cumulative sum
d.f.	degrees of freedom
DFM	dynamic factor model
DGP	data generating process
DSEM	dynamic simultaneous equations model
DSGE	dynamic stochastic general equilibrium
EM	expectation maximization
FAVAR	factor-augmented vector autoregression
FAVARMA	factor-augmented vector autoregressive moving average

FECM	factor error correction model
FIML	full information maximum likelihood
FOMC	Federal Open Market Committee
GARCH	generalized autoregressive conditional heteroskedasticity
GDFM	generalized dynamic factor model
GDP	gross domestic product
GIRF	generalized impulse response function
GLS	generalized least squares
GMM	generalized method of moments
GNP	gross national product
GO-GARCH	generalized orthogonal GARCH
GVAR	global vector autoregression
HP	Hodrick-Prescott
HPD	highest posterior density
HQC	Hannan-Quinn criterion
ICA	independent component analysis
IV	instrumental variables
LM	Lagrange multiplier
LR	likelihood ratio
LS	least squares
M1	narrow money stock
MA	moving average
MCMC	Markov Chain Monte Carlo
MGARCH	multivariate generalized autoregressive conditional heteroskedasticity
ML	maximum likelihood
MS	Markov switching
MSE	mean squared error
MSPE	mean squared prediction error
MS-VAR	Markov-switching vector autoregression
NBER	National Bureau of Economic Research
NBR	nonborrowed reserves
PC	principal components
RBC	real business cycle
SIC	Schwarz information criterion
SNP	semi-nonparametric
ST-VAR	smooth-transition vector autoregression
SVAR	structural vector autoregression
S&P 500	Standard and Poor's 500 stock price index
TFP	total factor productivity
TR	total reserves
TRAMO-SEATS	seasonal adjustment method
TVAR	threshold vector autoregression
TVC-VAR	time-varying coefficient vector autoregression
Var	variance
VAR	vector autoregression
VARMA	vector autoregressive moving average

VARX	VAR with exogenous variables
VECM	vector error correction model
X-12-ARIMA	seasonal adjustment method

Vector and Matrix Operations

M'	transpose of matrix M
M^{-1}	inverse of square matrix M
M_{\perp}	orthogonal complement of matrix M
$M^{1/2}$	square root of symmetric positive definite matrix M
M^k	k^{th} power of matrix M
MN	matrix product of matrices M and N
\otimes	Kronecker product
$\text{chol}(M)$	Cholesky factor of positive definite matrix M
$\det(M)$	determinant of matrix M
$\ M\ $	Euclidean norm of matrix M
$\text{rk}(M)$	rank of matrix M
$\text{tr}(M)$	trace of matrix M
vec	column stacking operator
vech	column stacking operator for square matrices (stacks the elements on and below the main diagonal only)
veck	column stacking operator for square matrices (stacks the elements above the main diagonal only)
$\frac{\partial \varphi}{\partial \beta'}$	vector or matrix of first order partial derivatives of function φ with respect to vector β
$\frac{\partial^2 \varphi}{\partial \beta \partial \beta'}$	matrix of second order partial derivatives of φ with respect to β (Hessian matrix)

General Matrices

\mathbf{D}_m	$m^2 \times \frac{1}{2}m(m+1)$ duplication matrix
I_m	$m \times m$ unit or identity matrix
$\mathcal{I}(\cdot)$	information matrix
$\mathcal{I}_a(\cdot)$	asymptotic information matrix
\mathbf{K}_{mn}	$mn \times mn$ commutation matrix
\mathbf{L}_m	$\frac{1}{2}m(m+1) \times m^2$ elimination matrix
0	zero or null matrix or vector of suitable dimension
$0_{m \times n}$	zero matrix of dimension $m \times n$
$\mathcal{O}(K)$	set of $K \times K$ orthogonal matrices

Distributions and General Stochastic Processes

$\mathcal{N}(\mu, \Sigma)$	(multivariate) normal distribution with mean (vector) μ and variance (covariance matrix) Σ
$\chi^2(m)$	χ^2 distribution with m degrees of freedom
$F(m, n)$	F distribution with m numerator and n denominator degrees of freedom

$t(m)$	t distribution with m degrees of freedom
$\mathcal{W}_K(\Sigma, n)$	K -dimensional Wishart distribution with parameters Σ and n
$\mathcal{IW}_K(\Sigma, n)$	K -dimensional inverse Wishart distribution with parameters Σ and n
$\mathcal{U}(a, b)$	uniform distribution on the interval (a, b)
\mathbf{W}_K	K -dimensional standard Brownian motion or Wiener process
$I(d)$	stochastic process integrated of order d
$I(1)$	stochastic process integrated of order 1
$I(0)$	stationary stochastic process
mds	martingale difference sequence

Alternative VAR Specifications

$y_t = A_1 y_{t-1} + \cdots + A_p y_{t-p} + u_t$	reduced-form VAR
$B_0 y_t = B_1 y_{t-1} + \cdots + B_p y_{t-p} + w_t$	structural-form VAR
$\Delta y_t = \boldsymbol{\Pi} y_{t-1} + \boldsymbol{\Gamma}_1 \Delta y_{t-1} + \cdots$	reduced-form VECM
$+ \boldsymbol{\Gamma}_{p-1} \Delta y_{t-p+1} + u_t$	
$B_0 \Delta y_t = \alpha^\dagger \beta y_{t-1} + \boldsymbol{\Gamma}_1^\dagger \Delta y_{t-1} + \cdots$	structural-form VECM
$+ \boldsymbol{\Gamma}_{p-1}^\dagger \Delta y_{t-p+1} + w_t$	

Vectors and Matrices Related to VAR Models

$u_t = \begin{pmatrix} u_{1t} \\ \vdots \\ u_{Kt} \end{pmatrix}$	K -dimensional white noise process, reduced-form error
$U \equiv [u_1, \dots, u_T]$	$K \times T$ matrix
$U_t \equiv \begin{pmatrix} u_t \\ 0 \\ \vdots \\ 0 \end{pmatrix}$	Kp -dimensional vector
$w_t = \begin{pmatrix} w_{1t} \\ \vdots \\ w_{Kt} \end{pmatrix}$	K -dimensional structural error vector
$y_t = \begin{pmatrix} y_{1t} \\ \vdots \\ y_{Kt} \end{pmatrix}$	K -dimensional stochastic process of observed variables
$y_{t+h t}$	h -step forecast of y_{t+h} made at point t
$Y \equiv [y_1, \dots, y_T]$	$K \times T$ matrix

$$Y_t \equiv \begin{pmatrix} y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix} \quad Kp\text{-dimensional vector}$$

$$Z_t \equiv \begin{pmatrix} 1 \\ y_t \\ \vdots \\ y_{t-p+1} \end{pmatrix} \quad (Kp + 1)\text{-dimensional vector}$$

$$Z \equiv [Z_0, \dots, Z_{T-1}] \quad (Kp + 1) \times T \text{ matrix or } [Y_0, \dots, Y_{T-1}] \quad Kp \times T \text{ matrix}$$

Matrices and Vectors Related to VARs and VECMs

$$A_i = \begin{bmatrix} a_{11,i} & \dots & a_{1K,i} \\ \vdots & \ddots & \vdots \\ a_{K1,i} & \dots & a_{KK,i} \end{bmatrix} \text{ reduced-form VAR coefficient matrix}$$

$$A \equiv [A_1, \dots, A_p] \text{ or } [\nu, A_1, \dots, A_p]$$

$$\alpha \equiv \text{vec}(A)$$

$$\mathbf{A} \equiv \begin{bmatrix} A_1 & \dots & A_{p-1} & A_p \\ I_K & & 0 & 0 \\ & \ddots & \vdots & \vdots \\ 0 & \dots & I_K & 0 \end{bmatrix}$$

$$A(L) \equiv I_K - A_1L - \dots - A_pL^p$$

$$B_i = \begin{bmatrix} b_{11,i} & \dots & b_{1K,i} \\ \vdots & \ddots & \vdots \\ b_{K1,i} & \dots & b_{KK,i} \end{bmatrix} \text{ structural-form VAR coefficient matrix}$$

$$B_0^{-1} = \begin{bmatrix} b_0^{11} & \dots & b_0^{1K} \\ \vdots & \ddots & \vdots \\ b_0^{K1} & \dots & b_0^{KK} \end{bmatrix} \text{ matrix of impact effects of structural shocks}$$

$$B(L) \equiv B_0 - B_1L - \dots - B_pL^p$$

$$\alpha \text{ loading matrix of VECM}$$

$$\beta \text{ cointegration matrix of VECM}$$

$$\Pi \equiv \alpha\beta'$$

$$\Gamma_i \text{ coefficient matrix on } i^{\text{th}} \text{ lagged difference of VECM}$$

$$\Phi_i = \begin{bmatrix} \phi_{11,i} & \dots & \phi_{1K,i} \\ \vdots & \ddots & \vdots \\ \phi_{K1,i} & \dots & \phi_{KK,i} \end{bmatrix} \text{ coefficient matrix of canonical MA representation}$$

$\Phi(L)$	$= I_K + \sum_{i=1}^{\infty} \Phi_i L^i$
Θ_i	$= \begin{bmatrix} \theta_{11,i} & \dots & \theta_{1K,i} \\ \vdots & \ddots & \vdots \\ \theta_{K1,i} & \dots & \theta_{KK,i} \end{bmatrix}$ matrix of structural impulse responses
$\Theta(L)$	$= \sum_{i=0}^{\infty} \Theta_i L^i$
Ξ	matrix of long-run effects of reduced-form shocks in VECM
Υ	$= \begin{bmatrix} \zeta_{11} & \dots & \zeta_{1K} \\ \vdots & \ddots & \vdots \\ \zeta_{K1} & \dots & \zeta_{KK} \end{bmatrix}$ matrix of long-run effects of structural shocks

Moment Matrices

Γ	$\equiv \text{plim } ZZ'/T$
$\Gamma_y(h)$	$\equiv \text{Cov}(y_t, y_{t-h})$ for a stationary process y_t
Σ_u	$\equiv \mathbb{E}(u_t u_t')$ (reduced-form white noise covariance matrix)
Σ_w	$\equiv \mathbb{E}(w_t w_t')$ (structural-form white noise covariance matrix)
Σ_y	$\equiv \mathbb{E}[(y_t - \mu)(y_t - \mu)']$ (covariance matrix of a stationary process y_t)
P	lower-triangular Cholesky decomposition of Σ_u
$\Sigma_{\hat{\alpha}}$	covariance matrix of the asymptotic distribution of $\sqrt{T}(\hat{\alpha} - \alpha)$
$\Sigma_y(h)$	MSE or forecast error covariance matrix of h -step forecast of y_t
$\Sigma_{\hat{y}}(h)$	approximate MSE matrix of h -step forecast of estimated process y_t

Author Index

- Abraham, B., 30, 223
Ahn, S.K., 53, 92, 558
Alesina, A.C., 208
Alessi, L., 591
Alquist, R., 204, 605, 639, 671
Alt, R., 71
Amemiya, T., 30
Amengual, 564
Amir Ahmadi, 560, 579, 583
An, S., 186
Andersen, T.G., 210
Anderson, B.D.O., 632
Anderson, S.T., 226, 434
Anderson, T.W., 87, 551, 552, 554, 558
Andrews, D.W.K., 33, 45, 71, 182, 188, 352
Anselin, L., 587
Antolín-Díaz, J., 452
Aoki, M., 632
Arezki, R., 543
Arias, J.E., 450, 451, 462, 472, 474, 475, 477, 478, 481, 482, 490
Aruoba, S.B., 565
Atanasova, C., 622
Auerbach, A.J., 543

Bacchicocchi, E., 525
Bai, J., 551, 552, 554, 556, 557, 560, 563, 564, 566, 567, 568
Baillie, R.T., 48
Balke, N.S., 619, 620, 622
Banbara, M., 150, 158, 159, 230, 550, 579, 580, 581, 582, 583, 588, 602
Banerjee, A., 568
Barhoumi, K., 551
Barigozzi, M., 556, 591

Barsky, R.B., 203, 243, 449, 606, 636
Basawa, I.V., 376, 382
Bastianin, A., 208
Basu, S., 208
Bauer, D., 42, 49, 381
Baumeister, C., 58, 124, 125, 126, 131, 158, 165, 211, 268, 396, 397, 409, 446, 447, 453, 455, 456, 457, 458, 459, 460, 461, 463, 464, 474, 489, 636, 672
Bäurle, G., 575
Baxter, M., 659, 661
Bayoumi, T., 579
Beaudry, P., 285, 286, 287, 449, 472, 608
Bell, W.R., 668
Benati, L., 474, 636
Benkwitz, A., 336, 362, 365, 366, 367, 368
Bera, A.K., 67
Berkowitz, J., 61, 353, 354, 410
Bernanke, B.S., 64, 136, 220, 222, 223, 229, 230, 231, 235, 253, 420, 502, 566, 569, 570, 577, 583, 638, 651
Bessler, D.A., 238
Beveridge, S., 22
Bickel, P.J., 588
Binning, A., 474
Binswanger, M., 276, 509
Birgean, I., 353, 354
Björnland, H.C., 516, 517
Blanchard, O.J., 220, 221, 223, 235, 236, 243, 269, 270, 271, 272, 278, 280, 284, 285, 293, 294, 295, 300, 309, 322, 323, 324, 387, 388, 399, 435, 573, 591, 604
Blinder, A., 220, 229
Boivin, J., 230, 231, 556, 560, 566, 567, 569, 570, 577, 578, 583

- Bollerslev, T., 210
 Bordo, M.D., 579
 Bose, A., 341
 Boswijk, H.P., 95
 Bouakez, H., 491, 523
 Box, G.E.P., 556
 Braun, P.A., 44
 Breidt, F.J., 646
 Breitung, J., 551, 554, 557, 563, 564, 565,
 585
 Breusch, T.S., 39
 Brewer, K.R.W., 30
 Brown, R.L., 71
 Brüggemann, R., 53, 54, 73, 91, 99, 100, 104,
 335, 346, 347, 348, 355, 548
 Burns, A.F., 636
 Caldara, D., 434, 450
 Camacho, M., 619
 Campbell, J.Y., 373, 507
 Candelon, B., 70, 71
 Caner, M., 377, 564
 Canova, 591
 Canova, F., 145, 189, 192, 195, 266, 267, 296,
 333, 420, 421, 426, 432, 436, 468, 469, 584,
 585, 587, 630, 634, 661, 662
 Capasso, M., 591
 Carlin, B.P., 149
 Carrasco, M., 629
 Carriero, A., 550, 581, 583
 Castelnovo, E., 525
 Cattell, R.B., 562
 Cavanagh, C.L., 106
 Chamberlain, G., 555, 558
 Chan, K.S., 621
 Chang, P.-L., 365, 377
 Chari, V.V., 324, 325, 327, 328, 329, 331
 Chauduri, S., 381, 444
 Chen, B., 601
 Chen, B.-C., 668
 Chen, C., 71
 Chen, L., 565
 Chen, R., 641
 Chen, W., 508
 Cheng, X., 182
 Cheung, Y.-W., 102
 Chevillon, G., 323
 Chib, S., 148
 Chihi, F., 524
 Choi, I., 551, 554, 555, 556, 557
 Choi, J., 601
 Christiano, L.J., 71, 177, 190, 218, 228, 230,
 232, 233, 296, 321, 324, 325, 327, 328, 329,
 330, 331, 332, 502, 534, 659, 661
 Chudik, A., 584, 586, 587
 Chung, H., 237, 591
 Ciccarelli, M., 189, 584, 585, 587, 630
 Coad, A., 238, 526, 527
 Coaglianese, J., 457
 Cochrane, J.H., 211, 234, 403, 533
 Cogley, T., 44, 184, 188, 199, 212, 220, 324,
 325, 326, 328, 630
 Connor, G., 555
 Cormick, W.P., 376, 382
 Cowles, M.K., 149
 Cragg, J.G., 536
 Croushore, D., 231
 Cushman, D.O., 74, 223
 D'Amico, S., 538, 540
 Darné, O., 551
 Datta, S., 376
 Davidson, J., 75
 Davidson, R., 345
 Davis, L.W., 457
 Davis, R.A., 223, 646, 647, 648
 De Mol, C., 579, 581
 De Nicolo, G., 421, 426
 Dedola, N., 467
 Dees, S., 586
 Deistler, M., 28, 632
 Del Negro, M., 170, 186, 190, 191, 195,
 578
 Demetrescu, M., 102
 Demiralp, S., 238
 Demiroglu, U., 68, 104
 Di Mauro, F., 586
 DiCecio, R., 451
 Diebold, F.X., 71, 189, 210, 565, 665
 Dios Tena, J.d., 622
 Doan, T., 150, 155, 158, 398
 Dolado, J.I., 42, 49, 50, 96, 107, 375, 565
 Donald, S.G., 536
 Doornik, J.A., 67, 69, 70, 95
 Dornbusch, R., 284, 468
 Dotsey, M., 208
 Doz, C., 560
 Dridi, R., 190
 Droumaguet, M., 508
 Dufour, J.-M., 50, 188, 381, 568
 Durbin, J., 71, 632
 Dwyer, M., 44, 188, 328

- Edelstein, P., 64, 227, 672
Edgerton, D., 54
Efron, B., 340, 357, 358, 359, 360
Ehrmann, M., 500, 501, 502, 504
Eichenbaum, M., 177, 190, 218, 228, 230,
232, 233, 234, 296, 321, 324, 325, 327,
328, 329, 330, 331, 332, 367, 435, 467,
502, 534
Eichler, M., 238
Eickmeier, S., 551, 563, 565, 570,
579
Elder, J., 636, 637, 638, 639
Eliasz, P., 230, 566, 569, 570, 577,
583
Elliott, G., 106, 386
Enders, W., 284
Engle, R.F., 75, 77, 92, 199
Entner, D., 238, 526, 527
Erceg, C.J., 293, 321, 322, 330, 332
Escanciano, C., 601
Essid, B., 524
Evans, C.L., 190, 218, 228, 230, 231, 232,
233, 234, 367, 435, 467, 502, 534
Evans, J.M., 71

Fanelli, L., 525
Farka, M., 538, 540
Faust, J., 293, 294, 321, 322, 421, 437, 449,
451, 534, 535, 536, 537, 538
Favero, C.A., 208, 544, 561, 566, 568, 578
Fernald, J., 208
Fernande, 591
Fernández-Villaverde, J., 177, 182, 192,
591
Ferrara, L., 551
Findley, D.F., 668
Fiorentini, G., 518
Fisher, J., 221
Fisher, L.A., 290, 291, 292, 420, 471
Fitzgerald, T.J., 659, 661
Flachaire, E., 345
Fomby, T.B., 619, 620
Forni, M., 230, 551, 560, 561, 562, 572, 573,
578, 591, 601, 602
Foroni, C., 30, 542
Franchi, M., 177
Francis, N., 296, 324, 451
Francq, C., 53
Franses, P.H., 619, 620, 665, 667
Fratzscher, M., 500, 501, 502, 504
Friedman, M., 196–197, 205, 347

Fry, R., 409, 432, 442, 444, 464, 469
Fujita, S., 466

Gafarov, B., 439
Galí, J., 223, 272, 280, 287, 288, 289, 293,
296, 299, 305, 306, 307, 324, 420
Gallant, A.R., 644, 645
Galvao, A.B., 619, 620, 622
Gambetti, L., 572, 573, 601, 630
Garcia, R., 508, 629
Gefang, D., 621
Gelfand, A., 148
George, E.I., 582
Gertler, M., 64, 136, 231, 548, 651
Geweke, J., 145, 148, 151
Ghysels, E., 30, 665, 667
Giacomini, R., 177, 195, 461, 482, 489
Giannone, D., 149, 150, 151, 158, 159, 165,
230, 550, 560, 575, 579, 580, 581, 582, 583,
588, 591, 602
Giannoni, M.P., 567, 578
Giavazzi, F., 208, 544
Glymour, C., 238
Goldberger, A., 171
Gómez, V., 668
Gonçalves, S., 44, 345, 346, 347, 348, 551,
575
Gonzalo, J., 60, 63, 103, 565
Gordon, D.B., 266, 420
Gorodnichenko, Y., 543, 544
Gospodinov, N., 108, 296, 322, 323, 330, 332,
379, 380, 381, 383, 384, 386, 387, 388
Gourieroux, C., 526, 527, 529
Granger, C.W.J., 23, 48, 72, 75, 77, 92, 197,
198, 238, 612, 619
Granziera, E., 438, 439, 490
Greenberg, E., 148
Greenspan, A., 636
Gregory, A.W., 39
Griffiths, W.E., 175, 260, 264
Guay, A., 190
Guerrieri, L., 293, 321, 322, 330, 332
Guerrón-Quintana, P., 177, 185, 188, 191
Gürkaynak, R.S., 190
Gust, C., 293, 321, 322, 330, 332

Hall, A.R., 183, 191
Hall, P., 362, 368, 418
Hallin, M., 551, 561, 562, 564, 575, 578
Haltiwanger, J.C., 223
Hamidi Sahneh, H., 591

- Hamilton, J.D., 64, 137, 138, 201, 202, 203, 211, 232, 268, 396, 397, 409, 446, 447, 453, 455, 456, 457, 458, 459, 460, 461, 463, 471, 489, 542, 547, 623, 624, 650, 651, 661
- Hammersly, J.M., 147
- Han, X., 564, 565
- Handscomb, D.C., 147
- Hannan, E.J., 632
- Hansen, B.E., 71, 382, 383, 508, 621
- Hansen, H., 67
- Hansen, L.P., 184, 253, 590
- Hansen, P.R., 70
- Harding, D., 662
- Härdle, W., 641, 643
- Harvey, A.C., 632
- Hasza, D.P., 48
- Hausman, J.A., 470
- Heinlen, R., 238
- Helbling, T., 579
- Hendry, D.F., 69, 70, 71, 75, 199
- Herault, J., 527
- Herrera, A.M., 64, 108, 137, 138, 232, 386, 658
- Herwartz, H., 92, 505, 507, 508, 510, 511, 526, 527, 528
- Hicks, B., 210
- Hill, R.C., 175, 260, 264
- Hodrick, R.J., 659, 660
- Hofmann, B., 570
- Hoover, K.D., 208, 238
- Horowitz, J.L., 418
- Hosking, J.R.M., 23, 53
- Howrey, E.P., 661
- Hoyer, P.O., 238, 526, 527
- Hsiao, C., 72
- Hsieh, D.A., 644
- Hu, L., 629
- Hubrich, K., 100, 103, 619, 623, 630
- Huh, H.-S., 290, 291, 292, 471
- Hurvich, C.M., 55
- Hussain, S.M., 658
- Hyvärinen, A., 527
- Inder, B., 159
- Ing, C.-K., 58, 61, 66
- Ingram, B.F., 170
- Inoue, A., 46, 55, 57, 58, 61, 73, 107, 177, 183, 185, 188, 191, 222, 328, 329, 339, 376, 381, 383, 402, 404, 406, 407–408, 409, 420, 432, 435, 436, 442, 444, 445, 446, 447, 453, 459, 462, 482, 483, 486, 489, 565, 618, 658
- Ivanov, V., 60
- Jarque, C.M., 67
- Jeliazkov, 643
- Jentsch, C., 335, 346, 347, 348, 355, 548
- Johansen, S., 23, 75, 79, 87, 88, 89, 95, 98, 102, 103, 307, 372
- Jordà, O., 390, 392, 393, 399
- Joyeux, R., 23
- Judge, G.G., 175, 260, 264
- Jutten, C., 527
- Kadiyala, K.R., 158, 397
- Kam, T., 44
- Kamps, C., 434, 450
- Kapetanios, G., 550, 581, 583
- Karadi, P., 548
- Karhunen, J., 527
- Karlsson, S., 158, 159, 397
- Kascha, C., 181, 182, 332
- Kashyap, A.K., 205
- Keating, J., 73, 194, 218, 223, 237, 250, 253
- Kehoe, P., 324, 325, 326, 327, 328, 329, 330, 331, 333
- Kellogg, R., 226, 434
- Keynes, J.M., 171
- Khalaf, L., 188
- Kichian, M., 188
- Kilian, L., 36, 37, 44, 46, 55, 57, 58, 59, 60, 61, 62, 63, 64, 68, 104, 107, 112, 113, 115, 119, 122, 124, 125, 126, 129, 131, 132, 134, 135, 136, 138, 158, 165, 169, 177, 182, 185, 188, 189, 190, 191, 203, 204, 208, 210, 211, 222, 223, 225, 226, 227, 232, 243, 328, 329, 339, 342, 345, 346, 347, 348, 353, 354, 357, 359, 360, 362, 363, 364, 365, 368, 373, 376, 377, 383, 391, 392, 394, 396, 399, 402, 404, 405, 406, 409, 410, 416, 419, 432, 433, 434, 435, 436, 437, 442, 444, 445, 446, 447, 451, 452, 453, 454, 455, 456, 457, 458, 459, 462, 463, 464, 469, 471, 482, 483, 486, 487, 489, 511, 527, 528, 543, 544, 547, 567, 579, 591, 604, 605, 636, 638, 639, 648, 650, 651, 652, 654, 655, 656, 657, 658, 671, 672
- Kim, J.H., 411
- Kim, Y.J., 391, 392, 399
- Kimball, 208
- King, R.G., 260, 272, 278, 279, 280, 281, 282, 303, 316, 319, 659, 661
- Kiraz, F.B., 222, 420
- Kisacikoglu, B., 190
- Kitagawa, T., 439, 461, 482, 489
- Kleibergen, F., 159
- Klein, L., 171

- Kloek, T., 147
Koltchinskii, V.I., 444
Konstantinos, T., 181
Koop, G., 150, 159, 163, 169, 550, 582, 617, 630, 634
Koopman, S.J., 632
Korajczyk, R.A., 555
Korobilis, D., 159, 163, 169, 583, 630, 634
Kose, M., 560, 573, 579
Kozicki, S., 636, 660, 671
Krämer, W., 71
Kreiss, J.-P., 339
Krolzig, H.-M., 238, 507, 622, 626, 629
Kulikov, D., 508
Künsch, H.R., 350, 359
Kunst, R.M., 665
Kurmann, A., 286, 287
Kuttner, K., 211, 533, 541
Kuznets, S.S., 661
Kydland, F.E., 185, 186, 188, 323, 325
- Lagalo, L.G., 658
Lai, K.S., 102
Lanne, M., 491, 493, 497, 500, 501, 502, 503, 505, 506, 509, 510, 519, 520, 521, 522, 523, 526, 528, 530, 646, 647, 648
Leamer, E.E., 199, 340
Lee, B.S., 284, 509
Lee, S., 238
Lee, T.-C., 175, 264
Lee, T.K., 119, 122, 126, 132, 134, 135, 451, 456, 469, 605
Leeb, H., 62, 63
Leeper, E.M., 125, 138, 185, 191, 194, 207, 231, 237, 266, 293, 294, 321, 322, 420, 591, 603, 604
Leitemo, K., 516, 517
Lenza, M., 149, 150, 151, 165, 581, 583
Leon-Gonzalez, 630
Leroy, S.F., 199, 212, 220
Lewis, R., 44, 136, 138, 232, 328, 567
Li, W.K., 53, 554, 566, 567
Lii, K.-S., 646
Lin, J.-L., 58, 66
Lippi, M., 230, 551, 556, 560, 561, 562, 578, 591, 602
Liška, R., 564, 575
Litterman, R.B., 150, 151, 155, 157, 158, 398
Liu, P., 181
Liu, R.Y., 444
Lomnicki, Z.A., 67
López-Salido, D., 296
- Lu, L., 566, 567
Lubik, T.A., 474
Lucas, R.E.J., 125, 137, 172, 183, 191, 192
Luciani, M., 556
Lunsford, K.G., 548
Lütkepohl, H., 19, 27, 28, 29, 31, 34, 41, 42, 44, 45, 48, 49, 50, 51, 53, 54, 59, 62, 66, 68, 70, 71, 73, 74, 75, 82, 85, 92, 95, 96, 99, 100, 102, 103, 104, 107, 158, 175, 180, 204, 260, 264, 276, 277, 295, 316, 328, 335, 336, 337, 339, 362, 365, 367, 368, 369, 371, 375, 384, 392, 399, 400, 402, 403, 491, 492, 493, 497, 498, 500, 501, 502, 503, 505, 506, 507, 508, 509, 510, 511, 514, 517, 521, 522, 523, 526, 528, 530, 531, 632, 638, 641, 667
- Maciejowska, K., 491, 497, 500, 505, 506, 510
Magnus, J.R., 555
Malik, S., 658
Mallik, A.K., 376, 382
Mammen, E., 346
Manera, M., 208
Manganelli, 129
Mann, H.B., 31, 613
Mansour, J.M., 579
Maravall, A., 668
Marcellino, M., 30, 66, 392, 542, 550, 561, 566, 568, 578, 581, 583, 619, 620, 622
Masten, I., 568
Mavroeidis, S., 323
Maynard, A., 42, 49, 296, 332, 381, 387, 388
McGrattan, E.R., 324, 325, 327, 328, 329, 331
McLeod, A.I., 53
Meier, M., 439
Meitz, M., 526, 528
Mertens, E., 181, 182, 208, 286, 287, 321, 332, 544
Michelacci, 296
Mihov, I., 222, 253, 420, 502, 578
Mikusheva, A., 45, 188, 383, 384, 393–394
Milunovich, G., 518, 521, 522, 523
Mittnik, S., 44
Moench, E., 575
Moneta, A., 238, 526, 527
Monfort, A., 526, 527, 529
Monsell, B.C., 668
Montiel Olea, J.L., 211, 439, 490, 544, 545, 546, 547, 548
Moon, H.R., 438, 439, 460, 490
Moore, J.B., 632
Mork, K.A., 650
Mountford, A., 449, 450, 471, 472

- Müller, U.K., 377
 Mumtaz, H., 473, 570, 636
 Murphy, D.P., 409, 433, 437, 442, 452, 454,
 455, 456, 457, 458, 463, 469, 471, 483, 487,
 511, 527, 591, 604
- Nalewaik, 565
 Nam, D., 449, 472
 Nason, J.M., 183, 191, 324, 325, 326
 Neglia, F., 561, 566, 568, 578
 Nelson, C.R., 22, 386
 Neri, S., 467
 Netšunajev, A., 492, 501, 508, 511, 514, 517,
 531
 Neudecker, H., 555
 Neumann, M.H., 336, 362, 365, 366, 367, 368
 Newbold, P., 198
 Newey, W.K., 470
 Ng, S., 56, 551, 552, 554, 556, 560, 563, 564,
 567, 575
 Ni, S., 582
 Nicholls, D.F., 36, 346
 Nickelsburg, G., 59
 Nielsen, M. Ø., 23, 98
 Normandin, M., 491, 518, 523, 524
 Nyblom, J., 72
- Ohanian, L.E., 650
 Oja, E., 527
 Oliner, S.D., 205
 Onatski, A., 44, 564
 Ortega, E., 585
 Osborn, D.R., 665, 666, 667
 Otrok, C., 560, 573, 578, 579
 Otto, M.C., 668
 Ouliaris, S., 428, 567
 Owyang, M., 451
- Paap, R., 159, 667
 Pagan, A., 182, 207, 237, 260, 289, 290, 291,
 292, 307, 309, 319, 320, 332, 346, 409, 428,
 432, 442, 444, 464, 469, 471, 567, 662
- Palm, F., 172
 Paparoditis, E., 45, 339, 353
 Parelius, J.M., 444
 Park, C., 113, 115, 226, 227
 Park, J.Y., 41
 Pascual, L., 411
 Paulsen, J., 60
 Paustian, M., 195, 333, 432, 436, 468, 469
 Payne, J., 439, 490
 Pearl, J., 238
- Peersman, G., 290, 292, 464, 471, 474, 510,
 636
 Peña, D., 556
 Perez, S.J., 208, 238
 Pérez Forero, F.J., 267, 420
 Perotti, R., 223, 236, 604
 Perron, P., 56, 551, 575
 Pesaran, M.H., 66, 289, 307, 309, 320, 392,
 584, 586, 587, 617
 Pesavento, E., 97, 108, 296, 332, 379, 380,
 381, 383, 384, 385, 386, 387, 388
 Phaneuf, L., 491, 518, 524
 Phillips, P.C.B., 41, 42, 49, 55, 97, 377, 396
 Piazzesi, M., 211, 533
 Pick, A., 66, 392
 Piffer, M., 544
 Pigorsch, U., 564
 Pitarakis, J.-Y., 60, 63, 103
 Plagborg-Møller, M., 459, 460, 461, 489
 Ploberger, W., 55, 71, 629
 Plödt, M., 527
 Plosser, C.I., 194, 272, 278, 279, 280, 281,
 303, 316, 319, 386
 Podstawska, M., 544
 Politis, D.N., 339
 Pope, A.L., 36, 329, 364, 365
 Portier, F., 285, 286, 287, 608
 Poskitt, D.S., 44, 49, 181, 328, 336, 402, 403
 Pötscher, B.M., 62, 63
 Potter, S.M., 617
 Prescott, E.C., 184, 185, 186, 188, 323, 325,
 659, 660
 Primiceri, G.E., 149, 150, 151, 165, 581, 583,
 630, 632, 633, 635, 636, 670
 Psaradakis, Z., 508, 629
- Qu, Z., 188
 Quah, D., 269, 270, 271, 272, 278, 280, 284,
 285, 293, 294, 295, 300, 309, 322, 323, 324,
 387, 388, 435, 573, 591
- Ragusa, R., 195
 Raïssi, H., 53
 Ramey, V.A., 207, 208, 210, 296, 324, 542,
 543, 591, 651, 672
 Rapach, D.E., 509
 Ravenna, F., 177, 181
 Ravn, M.O., 208, 544, 660
 Reeves, J.H., 376, 382
 Reichlin, L., 150, 158, 159, 230, 550, 551,
 560, 561, 562, 575, 578, 579, 580, 581, 583,
 588, 591, 602

- Reid, M., 208
Reimers, H.-E., 85, 105
Reinsel, G.C., 44, 48, 92, 328, 558
Renault, E., 191
Restrepo, J., 567
Richard, J.-F., 199
Rigobon, R., 491, 495, 496, 500, 501, 502, 504, 505
Rissanen, J., 55, 57
Robertson, J., 260, 319
Robertson J., 237
Robinson, P.M., 182, 332, 642
Rogers, J.H., 534, 535
Romer, C.D., 205, 206, 207, 542, 591, 607
Romer, D.H., 205, 206, 207, 542, 591, 607
Romo, J., 411
Rosenblatt, M., 646
Rossi, B., 97, 183, 190, 191, 379, 380, 381, 383, 384, 385, 386, 387
Rotemberg, J., 190, 228
Rothman, P.D., 619, 620
Rothschild, M., 555, 558
Roush, J.E., 451
Rubio-Ramírez, J.F., 177, 182, 192, 218, 246, 248, 304, 310, 312, 427, 428, 431, 450, 451, 452, 462, 472, 474, 475, 477, 478, 481, 482, 490, 506, 508, 591, 623, 624, 625
Rudebusch, G., 205, 211, 532, 533
Ruge-Marcia, F., 185
Ruiz, E., 411
Runkle, D.E., 329, 341, 344, 357, 360, 363
Sack, B., 491, 505
Saikkonen, P., 44, 49, 53, 54, 92, 96, 99, 100, 102, 103, 104, 519, 520, 521, 522, 523, 526, 528, 646, 647, 648
Sala, L., 192, 575
Salant, S.W., 226, 434
Salmon, M., 75
Samaranayake, V.A., 48
Sargan, J.D., 75
Sargent, T.J., 177, 182, 183, 184, 558, 590, 630
Sbrana, G., 599
Schaumburg, 665
Scheffé, H., 391
Scheines, R., 238
Schmidt, P., 39
Scholl, A., 435, 467
Schorfheide, F., 66, 170, 186, 190, 191, 195, 438, 439, 460, 490, 565
Schuermann, T., 584
Schwartz, A., 196–197, 205
Schwarz, G., 55
Scotti, C., 548
Sentana, E., 518
Serletis, A., 637, 638, 639
Shaman, P., 360
Shapiro, M.D., 207, 289, 542, 591
Sheng, L., 543
Shiller, R.J., 373
Shimmura, T., 198
Shin, Y., 617
Shukur, G., 54
Siegfried, N.A., 526, 527
Silvestrini, A., 599
Sims, C.A., 41, 51, 107, 108, 136, 150, 155, 158, 159, 171, 176, 185, 186, 191, 192, 194, 197, 212, 220, 221, 222, 223, 230, 231, 235, 266, 267, 268, 340, 373, 376, 381, 385, 394, 397, 398, 408, 420, 444, 449, 506, 508, 510, 558, 574, 606, 608, 623, 624, 626, 628, 629, 630, 671
Sin, C.-Y., 55, 61
Singh, K., 444
Singleton, K.J., 189
Smets, F., 170, 190, 191, 574
Smith, A.F.M., 148
Smith, V., 586
Song, S., 565, 588, 647, 648
Spagnolo, N., 508, 629
Spearman, C., 551, 552
Spirites, P., 238
Srba, F., 75
Staszewska, A., 399
Staszewska-Bystrova, A., 392, 399, 400
Stein, J.C., 205
Stevanovic, D., 568
Stine, R.A., 360
Stock, J.H., 41, 66, 97, 106, 107, 194, 211, 230, 272, 278, 279, 280, 281, 303, 316, 319, 373, 376, 380, 381, 392, 457, 544, 546, 547, 548, 551, 554, 558, 559, 560, 565, 566, 574, 576, 579
Strachan, R.W., 159, 621, 630
Strohsal, T., 525
Strongin, S., 229, 233
Sun, D., 582
Surico, P., 473, 570
Swanson, N.R., 211, 238, 533, 534, 535, 536, 537, 538
Taamouti, M., 381
Tauchen, G.E., 644, 645
Taylor, J., 237

- Taylor, M.P., 218, 294, 295
 Taylor, R.L., 376, 382
 Tenhofen, J., 551, 554, 557
 Teräsvirta, T., 72, 612, 619
 Tiao, G.C., 30
 Tibshirani, R., 340, 359
 Timmerman, A., 66, 392
 Tinbergen, J., 171
 Tinsley, P., 636, 671
 Tjøstheim, D., 60, 72, 612, 619
 Tobin, J., 197
 Toda, H.Y., 42, 49, 96, 102, 375
 Todd, R.M., 223
 Tremayne, A.R., 622
 Trenkler, C., 102, 335, 346, 347, 348, 355, 548
 Tsai, C.-L., 55
 Tschernig, R., 98, 272, 292, 293, 388, 642
 Tso, M.K.-S., 558
 Tsybakov, A., 643
 Uhlig, H., 44, 163, 394, 395, 397, 398, 421,
 432, 435, 436, 441, 443, 447, 448, 449, 450,
 451, 452, 467, 471, 472, 486, 487, 560, 579,
 583, 660
 Vahid, F., 44
 van Dijk, H.K., 147, 159, 619, 620
 Veall, M.R., 39
 Vega, C., 210, 223, 227, 638, 651
 Velinov, A., 276, 508, 510
 Velu, R.P., 558
 Venditti, F., 658
 Vidotto, A., 177
 Vigfusson, R.J., 177, 204, 296, 321, 322, 324,
 325, 327, 328, 329, 330, 331, 332, 639, 651,
 652, 654, 655, 656, 657, 671
 Villani, M., 149, 158, 159
 Vine, D.J., 651, 672
 Volcker, P., 231, 502, 636, 669, 670
 von Ungern-Sternberg, T., 75
 Wada, T., 658
 Waggoner, D.F., 124, 130, 149, 158, 192, 218,
 246, 248, 304, 310, 312, 427, 428, 431, 450,
 451, 452, 462, 472, 474, 475, 477, 478, 481,
 482, 490, 506, 508, 510, 623, 624, 625, 626,
 628, 629, 630, 671
 Wald, A., 31, 613
 Walker, T.B., 591, 603, 604
 Wallis, K.F., 184
 Wang, J., 449, 472
 Watson, M.W., 41, 64, 66, 107, 136, 177, 182,
 194, 211, 220, 223, 230, 231, 235, 260, 272,
 278, 279, 280, 281, 289, 303, 316, 319, 373,
 376, 392, 544, 546, 547, 548, 551, 554, 558,
 559, 560, 564, 565, 566, 574, 576, 579, 591,
 651
 Weber, E., 98, 272, 292, 388, 519, 525
 Wei, W.W.S., 30, 57, 61
 Weigand, R., 99, 272, 292, 388
 Weiner, S.M., 584
 Weise, C.L., 622
 White, H., 55
 Whiteman, C., 145, 151, 170, 560, 573,
 579
 Whittle, P., 460
 Wichern, D.W., 558
 Wilcox, D.W., 205
 Winker, P., 392, 399, 400
 Wold, H., 27
 Wolf, M., 392, 411
 Woodford, M., 190, 228, 403
 Wouters, R., 170, 190, 191, 574
 Wozniak, T., 508
 Wright, J.H., 379, 384, 500, 505, 534, 535,
 536, 537, 538, 548
 Wu, C.F.J., 30, 345
 Wunderli, D., 392, 411
 Yamamoto, Y., 42, 49, 96, 375, 570, 574, 576
 Yang, M., 518, 591, 603, 604, 643
 Yao, W., 44, 181
 Yeh, A.B., 354
 Yeo, S., 75
 Yoshimura, T., 198
 Yu, S.-H., 58, 61, 66
 Zellner, A., 172
 Zha, T., 74, 112, 124, 125, 130, 136, 138, 149,
 158, 159, 191, 192, 194, 218, 221, 223, 230,
 231, 246, 248, 266, 267, 268, 304, 310, 312,
 394, 397, 398, 408, 420, 427, 428, 431, 444,
 452, 506, 508, 510, 574, 623, 624, 625, 626,
 628, 629, 630, 671
 Zhan, Z., 323
 Zivot, E., 381

Subject Index

- Anderson-Rubin test, 547
Anticipation, 138–139, 202–203, 236–237, 285, 471–472, 603–607, 646
Asset prices, 201–203, 209, 224, 226, 230–232, 435, 532–542, 544, 603–605
Asymmetric responses, 638, 648–658
Asymmetric VAR models, 73–74
Asymptotic inference
 fractionally integrated processes, 49, 98–99, 292–293, 388–389
 integrated processes, 75–108, 369–377
 local-to-unity asymptotics, 97–98, 322–323, 337–385, 387–389
 robust inference in levels, 41–43, 107–108, 373–377, 381–387
 sieve approximations, 43–46
 stationary processes, 19–72, 334–369
Atheoretical VAR models, 212, 220
- Bayes factor, 144
Bayesian posterior inference
 credible sets, 143, 396–397, 406–409
 estimation, 142–143, 163, 186, 265–268
 framework, 140–145
 impulse response inference
 joint, 406–409
 pointwise, 394–397, 420
 median response function, 409, 442–447
 model averaging, 145
 posterior odds ratio, 143–144
 unit roots and cointegration in, 158–159, 397–398
Bayesian posterior simulation
 acceptance sampling, 146–147
 direct sampling, 145–146
- Gibbs sampling, 149
importance sampling, 147
Markov Chain Monte Carlo sampling, 147–148
Metropolis-Hastings sampling, 148–149
- Bayesian priors
 conjugate, 150
 diffuse, 154, 395
 flat, 153
 Gaussian for given Σ_u , 151–155
 Haar, 452–454
 hyperparameters, 150–151
 independent Gaussian-inverse Wishart, 166–169
 Litterman, 155–162
 Minnesota, 155–162
 natural conjugate, 150
 natural conjugate Gaussian-inverse Wishart, 162–166
 on structural model, 265–268, 454
 sum-of-coefficients, 158–159
 uniform, 151, 440
 uninformative, 154
- Beveridge-Nelson decomposition, 23, 80
Blaschke matrix, 593, 597
Bonferroni method, 384–385, 399–400, 438, 537–538
- Bootstrap inference
 bias-adjusted estimators, 363–365
 bias estimate, 365
 confidence interval
 Efron's percentile interval, 357–360
 equal-tailed percentile-t interval, 360–362
 finite-sample accuracy, 368–369

- Bootstrap inference (*cont.*)
 Hall's percentile interval, 362–363
 symmetric percentile-t interval, 362
 critical values of a test statistic, 411–412
 impulse response inference
 joint, 400–406
 pointwise, 340–369
 in overidentified models, 406–418
 invalidity, 366, 376
 pitfalls, 365–368
 prediction, 410–411
 standard error estimate, 357
- Bootstrap simulation methods
 blocks-of-blocks bootstrap, 351–353
 bootstrapping tuples of data, 347–348
 grid bootstrap, 382–384
 naïve block bootstrap, 348–351
 residual-based bootstrap
 block, 353–357
 iid, 341–345
 wild, 345–357
- Brownian motion, 83, 101, 381
- Censored variables, 650–658
- Cholesky decomposition, 219–221, 271, 275
- Cointegration
 common-trends representation, 79–80
 definition, 75–77
 implications for VAR model, 77–82
 matrix, 78
 near-cointegrated process, 98, 374, 384–385
 rank, 78
 rank tests, 100–104
 vector, 76–77
- Companion matrix, 25
- Counterfactuals, 121–123, 131–139, 191, 208
- Cowles Commission, 171, 187
- Delta method inference
 bootstrap standard errors, 357
 closed-form solutions for variance, 335–336
 parametric VAR models, 336–338
 sieve autoregressions, 338–339
- Distributed-lag model, 205–206
- Drift, 21
- Dynamic stochastic general equilibrium model
 calibration, 184–185
 estimation, 185–186
 forecasting, 190
 VAR representation, 177–183
- Dynamic simultaneous equations model
 final-form representation, 172
 identifying restrictions, 174–176
 reduced-form representation, 172
 structural representation, 172
- Endogeneity, 172, 177–178, 204–207, 541
- Exogeneity
 definition, 74, 171–172, 175, 193, 199–201, 543
 strict, 200
 tests of, 203–204
- Exogeneous shocks
 data-based counterfactuals, 208
 narrative approach, 205–208
 news shocks, 209–210
 shocks to expectations, 210–211
 structural VAR shocks, 211–215
- External instruments, 542–548
- Factor models
 approximate, 555, 557, 560
 dynamic, 555–560
 generalized dynamic, 560–562
 selecting the number of factors, 562–564
 static, 551–555
 structural change, 565
- Factors
 oblique, 551
 primitive, 558–559
- Forecast scenarios, 123–131
- Forward-looking behavior, 182, 201–203, 594–595, 602–607
- Fractional differencing, 23, 98–99, 292–293, 388–389
- Frequentist inference, 149–141, 143–145
- Fundamentalness
 definition, 592–594
 tests of, 593, 598, 601
- Fundamental process, 593
- Fundamental shock, 590
- Futures prices, 210–211, 532–542, 548, 598, 605, 607
- GARCH process, 24, 69, 92, 346, 492, 517–524, 528
- Generalized impulse response, 617–618
- Generalized orthogonal (GO) GARCH model, 518
- Granger causality
 bootstrap critical values, 411–412

- forward-looking asset prices, 201–203
in higher-dimensional processes, 50
in the presence of unit roots and
cointegration, 49, 85
tests, 48–50, 196–204
- Granger representation, 80–82, 273,
292–293
- Graph-theoretic approach, 238–239
- High-frequency data, 29–30, 210–211,
223–224, 532–542
- Historical decompositions
bar charts, 121–122
counterfactuals, 122–123
definition, 116–117
extensions to integrated series, 120–121
time series plots, 117–120
transients, 117–119
- Identification
data-based identification, 238–239, 491–492
exact identification, 174
order condition, 218
overidentification, 174, 253
partial identification, 174, 228, 430–432
point identification, 174
problem of, 174, 212–213
rank condition, 218
set identification, 174
underidentification, 174
- Identification based on
external instruments, 542–548
heteroskedasticity, 491–526
high-frequency futures prices, 532–542
long-run restrictions, 269–296
non-Gaussianity, 526–531
independent shocks, 526–528
uncorrelated shocks, 528–530
short-run restrictions, 216–240
sign restrictions, 421–490
- Identifying restrictions
exclusion restrictions, 218, 222–224
long-run restrictions, 269–278
short-run restrictions, 216–219
sign restrictions, 294–295, 421–436
sources of, 221–224, 421–424
- Idiosyncratic component, 551
- Impulse responses, 110–113
cumulated, 212
forecast error, 111
generalized, 617–618
- long-run, 269–278
nonlinear, 614–619
reduced-form, 111
structural, 110–113
transformations, 112
- Inference in sign-identified models
agnostic Bayesian priors, 462–463
conventional Bayesian priors, 440–442
explicit Bayesian priors, 453–461
frequentist approaches, 438–439
other non-Bayesian approaches, 463–464
penalty function approaches, 448–451
robust Bayesian priors, 461–462
using historical information, 451–452
- Instrumental variables
external instruments, 542–548
structural VAR(0) model, 258–259
structural VAR(p) model, 259–262,
307–310
weak instruments, 211, 323, 548
- Joint inference, 391–392, 398–411, 445–446,
483–488
- Kalman filter, 560, 633
- Kurtosis, 67–68
- Kuznets cycle, 661
- Lag augmentation, 42–43, 49–50, 96, 376,
378, 381
- Lag operator, 21
- Lag-order selection
Bayesian approach, 144, 169
conservative approach, 63–65
consistency, 58–59
finite-sample properties, 59–60
information criteria
Akaike information criterion, 54–57
Hannan–Quinn criterion, 54–57
Schwarz information criterion, 54–57
practical considerations, 61–66
recursive MSPE rankings, 57–58
sequential tests
LM test, 53–54
LR test, 51–52
Portmanteau test, 52–53
Wald test, 52
- Large-dimensional structural VAR models
Bayesian VAR models, 579–584
factor-augmented (FA) structural VAR
models, 565–571

- Large-dimensional structural VAR models (*cont.*)
 global VAR models, 586–587
 panel VAR models, 584–585
 spatial models, 587
 structural dynamic factor models, 571–577
- LASSO approach, 588–589
- Levels VAR models with integrated variables, 95–99
 estimation, 95–96
 fractional integration, 98–99
 local-to-unity asymptotics, 97–98
 sieve estimation, 96–97
- Loading matrix, 78, 552, 569
- Local projections, 389–393
- Long memory, 23, 42
- Long-run restrictions
 ability to recover DSGE model responses, 323–333
 combining short- and long-run restrictions, 287–292, 310–320
 estimation subject to, 299–320
 general framework, 272–278
 practical limitations, 320–323
 theoretical limitations, 292–296
 traditional framework, 269–272
- Loss function, 60, 142, 145, 184, 265–266, 399, 409, 447, 449, 458–459, 489–490
- Lucas critique, 137–138, 172, 183, 191–192, 232, 625
- Marginal likelihood, 144, 150–151, 189
- Marginal processes, 28–29, 180, 599
- Markov switching, 623–626
- Martingale difference sequence, 24, 44, 46–47, 257, 601, 640, 645
- Martingale sequence, 24
- Measurement error, 30, 179, 565
- Model diagnostics
 nonnormality, 67–68, 104
 residual ARCH, 68–69, 104
 residual autocorrelation, 67, 104
 time invariance, 69–72, 104–105
- Moore-Penrose inverse, 336
- Moving average
 fundamental, 592–601
 nonfundamental, 592–601
 representation, 26–27, 111
- Multiplier
 dynamic, 45, 590
 impact, 3, 7, 109–111, 217–219
 long-run, 272–274, 321, 387–386
- Near unit roots, 97, 108, 158–159
- Nonfundamentalness, 592–594
- Nonfundamental process, 593
- Nonfundamental shock, 592
- Nonlinear VAR models
 GARCH-in-mean, 636–640
 Markov switching, 505–511, 528–530, 622–630
 noncausal VAR model, 645–648
 nonparametric VAR model, 640–645
 thresholds and smooth transitions, 619–622
 time-varying coefficients, 630–636
- Normalization, 78, 83, 88, 216–219, 245–246, 267, 295, 414, 427, 506, 536–537, 545, 552
- Order of integration, 22
- Ornstein-Uhlenbeck process, 381
- Orthogonal factor, 551
- Orthogonalization, 219–221
- Orthogonal matrix, 313–315, 425, 327
- Overidentified structural models
 estimation, 253–254, 258, 266, 319, 420
 inference, 416–420
 testing, 254, 418
- Penalty functions, 448–452
- Policy counterfactuals, 136–139, 191–192, 231–232
- Post-model selection inference, 61–63, 106
- Predeterminedness, 199–200
- Prediction
 conditional expectation, 46, 210, 594, 615
 prediction intervals, 47, 410–411
 prediction mean squared error, 46–48, 114–116
 predictive density, 127–128, 410–411, 643
- Price puzzle, 230, 538, 578
- Principal components, 229, 554, 561, 578, 601
- Projection method, 381, 438
- Proxy VAR model, 544
- QR decomposition, 427
- Quasi-maximum likelihood estimator, 40

- Random walk, 20–21, 80, 83
Rational expectations, 176, 183–184, 194
Real business cycle model, 109, 178–179,
221, 278–280, 283, 296, 323–324
Reduced-form autoregressive representation,
25
Reduced-form estimation
 VAR model
 bias-corrected least squares, 35–38
 generalized least squares, 34–35
 least squares, 31–34
 levels VAR models with integrated
 variables, 41–43
 maximum likelihood estimation, 38–40
 VECM
 generalized least squares, 84
 least squares, 83, 95–96
 least squares with known cointegrating
 matrix, 84–85
 maximum likelihood with known
 cointegrating matrix, 90–95
 maximum likelihood with known
 cointegrating rank, 85–90
Reduced-form moving average representation,
26–27
Risk, 129–130
Romer dummies, 205–208
Roots, 21–22, 25–28, 45–46
Rotation matrix
 generation of, 426–428
 Givens rotation matrix, 426
 Householder transformation, 427–428
- Seasonality
 combining seasonal specifications, 668
 deterministic, 663
 periodical, 666–667
 seasonal differencing, 664–665
 seasonal dummies, 663
 seasonal filters, 667–668
 seasonal integration, 664–666
 seasonal lags, 664
 seasonal unit roots, 665
 stationary, 664
 stochastic, 664–666
 time-varying, 667
Semistructural VAR model, 228–234, 464–465
Shocks to
 aggregate demand, 113, 224–226, 269,
 289–290, 310, 416, 465, 473, 528,
 543
aggregate supply, 224, 237, 269, 288, 290,
 310, 465, 473
demand, 113, 213, 224, 269, 421–424, 466
efficiency, 179
energy price, 227
exchange rate, real and nominal, 284
expected productivity, 284, 472, 606
financial market expectations, 210
fiscal policy, 177, 207–208, 236, 471, 544,
 594, 603
flow demand, 433, 470, 483
flow supply, 433, 470, 483
IS curve, 237, 288–289
labor supply, 289
LM curve, 289
markup, 468
measurement error, 179, 468
monetary policy, 205, 224, 228–234, 290,
 310, 404–405, 465, 467–468, 473, 485,
 502, 523, 532–535, 538–539, 577–579,
 607, 621, 629, 635
money demand, 224, 237, 251, 288, 473
money supply, 177, 224, 237, 251, 288, 473
news, 209, 285–286, 543–544, 606
oil market-specific demand, 113, 225–226,
 416, 528, 543
oil price, 136, 226–227, 243, 289–290, 405,
 465, 638, 651–652
oil price expectation, 604–605
oil supply, 113, 208, 225–226, 416, 528,
 543, 579
preferences, 179
profitability, 466
speculative demand, 470, 604
supply, 113, 213, 269, 421–424
tastes, 468
technology, 109, 177, 208, 271, 278, 282,
 289, 299, 323–333, 467–468, 594, 606
- Shocks with
 permanent effects, 235, 279
 transitory effects, 235, 279, 283
- Short-run restrictions
 combining short- and long-run restrictions,
 287–292, 310–320
 estimation subject to, 242–268
 nonrecursive models, 223–224, 235–236
 recursive models, 219–221
 semistructural models, 227–235
- Sieve approximations, 27–28, 43–46, 51, 61,
 96–97, 181–182, 328–329, 338–339,
 392, 566

- Sign restrictions
 dynamic restrictions, 432, 435
 elasticity bounds, 432–434
 shape restrictions, 425
 static restrictions, 421–432
 verification of, 435–436
- Skewness, 67–68, 359
- Square root matrix, 68, 83, 248, 519, 520
- Stability condition, 21–22, 25
- State-space model
 measurement equation, 632
 observation equation, 559
 state equation, 632
 transition equation, 559, 563, 570, 632, 634
- State-space representation of
 DSGE model, 177, 182, 185, 188
 dynamic factor model, 558–559
 FAVAR model, 566
 time-varying coefficient model, 631–632
- Stationarity, 19, 21–22
- Structural analysis, 187
- Structural autoregressive representation, 216–219
- Structural change
 breaks in stochastic component, 669–671
 smooth structural change, 671–672
- Structural estimation
 Bayesian, 265–268, 440–464
 generalized method of moments, 253–258
 instrumental variables, 258–262, 307–310, 315–316
 maximum likelihood, 262–265, 306–307
 method of moments, 242–253, 301–306, 311–315
- Structural forecast error variance
 decompositions
 definition, 113–116
 inference, 336, 375, 369, 439
- Structural impulse responses
 linear models, 110–113
 nonlinear models, 614–619
- Structural moving average representation, 111
- Structural shocks
 correlated, 525–526
 independent, 526–528
 uncorrelated, 2, 219–220, 494, 528–530
- Structural vector error correction model
 definition, 298–299
 estimation, 307–310
- Subrotation, 474
- Subset VAR models, 72–73
- TRAMO-SEATS, 668
- Trends
 combining trend specifications, 662–663
 common, 75–77, 660
 deterministic, 20, 108
 stochastic, 20–21
 trend filters
 band pass filter, 660–661
 Hodrick-Prescott (HP) filter, 659–660
 moving-average filter, 660–661
- Unit roots
 Dickey-Fuller test, 83, 380, 384, 376
 integrated processes, 21–23
 limitations of tests for unit roots and cointegration, 106
- Variables
 cross-sectional aggregation, 180
 exogenous, 200–201
 predetermined, 199–200
 linear transformation, 29, 180
 temporal aggregation, 29–30
- Variance decomposition, 114
- VARMA model, 27–28, 43–45, 62, 172–173, 177–178, 180–182, 328–329, 392, 568
- VARX model, 74, 173, 543
- VECM
 model specification, 99–105
 recovering VAR parameters, 78, 95
 representation, 77–80
 role of intercept and trend, 80–82
 sieve approximations, 96–97
 triangular representation, 372–373
- Volatility models
 GARCH errors, 517–524
 Markov switching in variance, 505–511
 smooth transitions in variance, 511–517
 stochastic, 632
 with extraneously specified volatility changes, 496–505
- White noise, 20, 24
- Wold decomposition theorem, 27
- Wold moving average representation, 460
- X-12 ARIMA seasonal adjustment, 668