

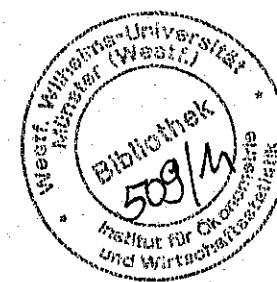
ÖW 5739.A

---

*Structural Macroeconomics*

SECOND EDITION

*David N. DeJong and Chetan Dave*



## Contents

---

<i>Preface</i>	xiii
<i>Preface to the First Edition</i>	xv
Part I Introduction	
1 Background and Overview	3
1.1 Background	3
1.2 Overview	4
2 Casting Models in Canonical Form	9
2.1 Notation	9
2.1.1 Log-Linear Model Representations	11
2.1.2 Nonlinear Model Representations	11
2.2 Linearization	12
2.2.1 Taylor Series Approximation	12
2.2.2 Log-Linear Approximations	14
2.2.3 Example Equations	15
3 DSGE Models: Three Examples	18
3.1 Model I: A Real Business Cycle Model	20
3.1.1 Environment	20
3.1.2 The Nonlinear System	23
3.1.3 Log-Linearization	26
3.2 Model II: Monopolistic Competition and Monetary Policy	28
3.2.1 Environment	28
3.2.2 The Nonlinear System	33
3.2.3 Log-Linearization	34
3.3 Model III: Asset Pricing	38
3.3.1 Single-Asset Environment	38
3.3.2 Multi-Asset Environment	39
3.3.3 Alternative Preference Specifications	40
Part II Model Solution Techniques	
4 Linear Solution Techniques	51
4.1 Homogeneous Systems	52
4.2 Example Models	54
4.2.1 The Optimal Consumption Model	54
4.2.2 Asset Pricing with Linear Utility	55
4.2.3 Ramsey's Optimal Growth Model	56

	Contents
4.3 Blanchard and Kahn's Method	57
4.4 Sims' Method	61
4.5 Klein's Method	64
4.6 An Undetermined Coefficients Approach	66
<b>5 Nonlinear Solution Techniques</b>	<b>69</b>
5.1 Projection Methods	71
5.1.1 Overview	71
5.1.2 Finite Element Methods	72
5.1.3 Orthogonal Polynomials	73
5.1.4 Implementation	74
5.1.5 Extension to the $l$ -dimensional Case	78
5.1.6 Application to the Optimal Growth Model	79
5.2 Iteration Techniques: Value-Function and Policy-Function Iterations	87
5.2.1 Dynamic Programming	87
5.2.2 Value-Function Iterations	89
5.2.3 Policy-Function Iterations	94
5.3 Perturbation Techniques	95
5.3.1 Notation	95
5.3.2 Overview	97
5.3.3 Application to DSGE Models	99
5.3.4 Application to an Asset-Pricing Model	105
<b>Part III Data Preparation and Representation</b>	<b>113</b>
<b>6 Removing Trends and Isolating Cycles</b>	<b>115</b>
6.1 Removing Trends	115
6.2 Isolating Cycles	120
6.2.1 Mathematical Background	120
6.2.2 Cramér Representations	124
6.2.3 Spectra	125
6.2.4 Using Filters to Isolate Cycles	126
6.2.5 The Hodrick-Prescott Filter	128
6.2.6 Seasonal Adjustment	130
6.2.7 Band Pass Filters	131
6.3 Spuriousness	134
<b>7 Summarizing Time Series Behavior When All Variables Are Observable</b>	<b>138</b>
7.1 Two Useful Reduced-Form Models	139
7.1.1 The ARMA Model	139
7.1.2 Allowing for Heteroskedastic Innovations	145
7.1.3 The VAR Model	147

	Contents
7.2 Summary Statistics	149
7.2.1 Determining Lag Lengths	157
7.2.2 Characterizing the Precision of Measurements	159
7.3 Obtaining Theoretical Predictions of Summary Statistics	162
<b>8 State-Space Representations</b>	<b>166</b>
8.1 Introduction	166
8.1.1 ARMA Models	167
8.2 DSGE Models as State-Space Representations	169
8.3 Overview of Likelihood Evaluation and Filtering	171
8.4 The Kalman Filter	173
8.4.1 Background	173
8.4.2 The Sequential Algorithm	175
8.4.3 Smoothing	178
8.4.4 Serially Correlated Measurement Errors	181
8.5 Examples of Reduced-Form State-Space Representations	182
8.5.1 Time-Varying Parameters	182
8.5.2 Stochastic Volatility	185
8.5.3 Regime Switching	186
8.5.4 Dynamic Factor Models	187
<b>Part IV Monte Carlo Methods</b>	<b>193</b>
<b>9 Monte Carlo Integration: The Basics</b>	<b>193</b>
9.1 Motivation and Overview	193
9.2 Direct Monte Carlo Integration	196
9.2.1 Model Simulation	198
9.2.2 Posterior Inference via Direct Monte Carlo Integration	201
9.3 Importance Sampling	202
9.3.1 Achieving Efficiency: A First Pass	206
9.4 Efficient Importance Sampling	211
9.5 Markov Chain Monte Carlo Integration	215
9.5.1 The Gibbs Sampler	216
9.5.2 Metropolis-Hastings Algorithms	218
<b>10 Likelihood Evaluation and Filtering in State-Space Representations Using Sequential Monte Carlo Methods</b>	<b>221</b>
10.1 Background	221
10.2 Unadapted Filters	224
10.3 Conditionally Optimal Filters	228
10.4 Unconditional Optimality: The EIS Filter	233
10.4.1 Degenerate Transitions	235
10.4.2 Initializing the Importance Sampler	236
10.4.3 Example	239



	<i>Contents</i>
10.5 Application to DSGE Models	241
10.5.1 Initializing the Importance Sampler	243
10.5.2 Initializing the Filtering Density	245
10.5.3 Application to the RBC Model	246
<b>Part V Empirical Methods</b>	
<b>11 Calibration</b>	253
11.1 Historical Origins and Philosophy	253
11.2 Implementation	258
11.3 The Welfare Cost of Business Cycles	261
11.4 Productivity Shocks and Business Cycle Fluctuations	268
11.5 The Equity Premium Puzzle	273
11.6 Critiques and Extensions	276
11.6.1 Critiques	276
11.6.2 Extensions	279
<b>12 Matching Moments</b>	285
12.1 Overview	285
12.2 Implementation	286
12.2.1 The Generalized Method of Moments	286
12.2.2 The Simulated Method of Moments	294
12.2.3 Indirect Inference	297
12.3 Implementation in DSGE Models	300
12.3.1 Analyzing Euler Equations	300
12.3.2 Analytical Calculations Based on Linearized Models	301
12.3.3 Simulations Involving Linearized Models	306
12.3.4 Simulations Involving Nonlinear Approximations	307
12.4 Empirical Application: Matching RBC Moments	308
<b>13 Maximum Likelihood</b>	314
13.1 Overview	314
13.2 Introduction and Historical Background	316
13.3 A Primer on Optimization Algorithms	318
13.3.1 Simplex Methods	319
13.3.2 Derivative-Based Methods	328
13.4 Ill-Behaved Likelihood Surfaces: Problems and Solutions	330
13.4.1 Problems	330
13.4.2 Solutions	331
13.5 Model Diagnostics and Parameter Stability	334
13.6 Empirical Application: Identifying Sources of Business Cycle Fluctuations	337
<b>14 Bayesian Methods</b>	351
14.1 Overview of Objectives	351
14.2 Preliminaries	352

14.3 Using Structural Models as Sources of Prior Information for Reduced-Form Analysis	355
14.4 Implementing Structural Models Directly	360
14.5 Model Comparison	361
14.6 Using an RBC Model as a Source of Prior Information for Forecasting	364
14.7 Estimating and Comparing Asset-Pricing Models	373
14.7.1 Estimates	380
14.7.2 Model Comparison	384
<b>References</b>	387
<b>Index</b>	401

## Preface

---

Scratch where it itches.

—Alice Roosevelt Longworth

If it ain't broke, don't fix it.

—Bert Lance

The usual excuse for crafting the second edition of a text is rapidly changing subject matter. We could certainly invoke this excuse here: the empirical analysis of dynamic stochastic general equilibrium (DSGE) models has evolved considerably over the past five years. Instead, we confess that this excuse is only second-order important in this instance. First-order important is our graduation from the steepest portion of a critical learning curve.

In particular, when we set out to characterize likelihood evaluation and filtering for nonlinear model representations, our comprehension of applicable numerical methods did not extend beyond the standard particle filter. Moreover, our understanding of the particle filter was limited to the mechanics of its execution (which is exceedingly simple), and did not extend to a sufficient appreciation of its shortcomings (which, alas, can be significant in empirically relevant situations). Indeed, as we were putting the finishing touches on the first edition, one of us (DeJong) was developing plans to implement the particle filter in pursuit of a promising empirical application. Shortly thereafter, the particle filter was found to be not up to the task, and despite the passage of time, the application remains on the back burner (thank goodness for tenure).

The good news is that, with necessity serving as the mother of invention, we have come to better appreciate the strengths and weaknesses of the particle filter, as well as alternative methods. Even better, our emergence from the learning curve has led to the development (by DeJong, in collaboration with Hariharan Dharmarajan, Roman Liesenfeld, Guilherme Moura, and Jean-François Richard) of an alternative filtering method: the efficient importance sampling (EIS) filter. As we shall demonstrate, implementation of the EIS filter is well suited to DSGE models, and the payoff to its adoption can be high.

The emphasis here on filtering reflects our appreciation of the importance that departures from linearity hold for the future of empirical work involving DSGE models. Five years ago, the use of nonlinear model

EIS Filter  
vs. Particle Filter

representations in empirical applications was exotic; five years from now, we expect this use to be routine. Accordingly, while in the first edition the treatment of nonlinear model representations was isolated to a pair of closing chapters, linear and nonlinear treatments are now characterized in tandem throughout the text.

Users of the first edition will notice additional changes as well: details are sketched in chapter 1. However, we remain satisfied with the bulk of the subjects covered in the first edition, and have heeded the wisdom of Bert Lance in their treatment. We hope we have done so judiciously.

Since the publication of the first edition, we have benefited from input on the text received from students we have taught at Carnegie-Mellon and Di Tella universities, the Kiel Institute for the World Economy, the University of Pittsburgh, and the University of Texas at Dallas. We have also been alerted to many errors of commission: we are particularly indebted to Roberto Chang, Andres Fernandez, Michael Hauser, Jan Christopher Ocampo, and Surach Tanboon for their careful reading. Certainly we committed many errors of omission in the first edition when it came to describing literatures related to the topics we covered. We have sought to update and expand our treatment of these literatures here, though in so doing we have surely left a great many excellent papers unmentioned: apologies in advance for this.

Dave DeJong and Chetan Dave

## *Preface to the First Edition*

---

Let me only say that what econometrics—aided by electronic computers—can do, is only to push forward by leaps and bounds the line of demarcation from where we have to rely on our intuition and sense of smell.

—Ragnar Frisch, Nobel Prize Lecture,  
June 1970.

It is a capital mistake to theorize before one has data. Insensibly one begins to twist facts to suit theories, instead of theories to suit facts.

—Sir Arthur Conan Doyle

A steady stream of conceptual and computational advances realized over the past three decades has helped to bridge the gulf that has historically separated theoretical from empirical research in macroeconomics. As a result, measurement is becoming increasingly aligned with theory. The purpose of this text is to provide guidance in bringing theoretical models to the forefront of macroeconomic analyses.

The text is suitable for use as a supplementary resource in introductory graduate courses in macroeconomics and econometrics, and as a primary textbook in advanced graduate courses devoted to the pursuit of applied research in macroeconomics. The lecture notes that ultimately gave rise to the text were designed for this latter purpose. The text's historical perspective, along with its unified presentation of alternative methodologies, should also make it a valuable resource for academic and professional researchers.

Readers of the text are assumed to have familiarity with multivariate calculus, matrix algebra, and difference equations, and cursory knowledge of basic econometric techniques. Familiarity with dynamic programming is also useful but not necessary. This is the tool used to map the class of models of interest here into the system of nonlinear expectational difference equations that serve as the point of departure for the empirical methodologies presented in the text. However, familiarity with dynamic programming is not needed to follow the text's presentation of empirical methodologies.

We decided to write this text because through our own teaching and research, we have sought to contribute towards the goal of aligning theory

with empirical analysis in macroeconomics; this text is a natural extension of these efforts. We set out to accomplish two objectives in writing it. First, we wished to provide a unified overview of this diverse yet interrelated area of active research. Second, we wanted to equip students with a set of tools that would expedite their own entry into the field.

The content of this text reflects much that we have learned over many years spent pursuing collaborative research with a fabulous group of coauthors: Sheryl Ball, Patty Beeson, Dan Berkowitz, Stefan Dodds, Scott Dressler, Catherine Eckel, Emilio Espino, Steve Husted, Beth Ingram, Roman Leisenfeld, John Nankervis, Jean-François Richard, Marla Ripoll, Gene Savin, Werner Troesken, and Chuck Whiteman. We are deeply indebted to them for their implicit and explicit contributions to this project. We have also benefited from input provided by Charlie Evans, Jim Feigenbaum, Jesús Fernández-Villaverde, Peter Ireland, Naryana Kocherlakota, Chris Otrok, Barbara Rossi, Juan Rubio-Ramírez, Tom Sargent, Thomas Steinberger, Yi Wen, and Tao Zha on various aspects of the text. Finally, Hariharan Dharmarajan provided valuable research assistance on the empirical application presented in chapter 11.

## *Part I*

---

### *Introduction*

# Chapter 1 - Structural Macroeconomics

## *Background and Overview*

Science is facts; just as houses are made of stones, so is science made of facts; but a pile of stones is not a house and a collection of facts is not necessarily science.

—Henri Poincaré

### **1.1 Background**

The seminal contribution of Kydland and Prescott (1982) marked a sea change in the way macroeconomists conduct empirical research. Under the empirical paradigm that remained predominant at the time, the focus was either on purely statistical (or reduced-form) characterizations of macroeconomic behavior, or on systems-of-equations models that ignored both general-equilibrium considerations and forward-looking behavior on the part of purposeful decision makers. But the powerful criticism of this approach set forth by Lucas (1976), and the methodological contributions of, for example, Sims (1972) and Hansen and Sargent (1980), sparked a transition to a new empirical paradigm. In this transitional stage, the formal imposition of theoretical discipline on reduced-form characterizations became established. The source of this discipline was a class of models that have come to be known as dynamic stochastic general equilibrium (DSGE) models. The imposition of discipline most typically took the form of “cross-equation restrictions,” under which the stochastic behavior of a set of exogenous variables, coupled with forward-looking behavior on the part of economic decision makers, yield implications for the endogenous stochastic behavior of variables determined by the decision makers. Nevertheless, the imposition of such restrictions was indirect, and reduced-form specifications continued to serve as the focal point of empirical research.

Kydland and Prescott turned this emphasis on its head. As a legacy of their work, DSGE models no longer serve as indirect sources of theoretical discipline to be imposed upon statistical specifications. Instead, they serve directly as the foundation upon which empirical work may be conducted. The methodologies used to implement DSGE models as foundational empirical models have evolved over time and vary considerably. The same

is true of the statistical formality with which this work is conducted. But despite the characteristic heterogeneity of methods used in pursuing contemporary empirical macroeconomic research, the influence of Kydland and Prescott remains evident today.

This book details the use of DSGE models as foundations upon which empirical work may be conducted. It is intended primarily as an instructional guide for graduate students and practitioners, and so contains a distinct how-to perspective throughout. The methodologies it presents are organized roughly following the chronological evolution of the empirical literature in macroeconomics that has emerged following the work of Kydland and Prescott; thus it also serves as a reference guide. Throughout, the methodologies are demonstrated using applications to three benchmark models: a real-business-cycle model (fashioned after King, Plosser, and Rebelo, 1988); a monetary model featuring monopolistically competitive firms (fashioned after Ireland, 2004a); and an asset-pricing model (fashioned after Lucas, 1978).

## 1.2 Overview

The empirical tools outlined in the text share a common foundation: a system of nonlinear expectational difference equations representing a DSGE model. The solution to a given system takes the form of a collection of policy functions that map a subset of the variables featured in the model—state variables—into choices for the remaining subset of variables—control variables; the choices represent optimal behavior on the part of the decision makers featured in the model, subject to constraints implied by the state variables. Policy functions, coupled with laws of motion for the state variables, collectively represent the solution to a given DSGE model; the solution is in the form of a state-space representation.

Policy functions can be calculated analytically for only a narrow set of models, and thus must typically be approximated numerically. The text presents several alternative methodologies for achieving both model approximation and the subsequent empirical implementation of the corresponding state-space representation. Empirical implementation is demonstrated in applications involving parameter estimation, assessments of fit and model comparison, forecasting, policy analysis, and measurement of unobservable facets of aggregate economic activity (e.g., measurement of productivity shocks).

This book is divided into five parts. Part I contains three chapters. Following this brief introduction, chapter 2 establishes the consistent set of notation used throughout the text, and presents an important preliminary

### 1.2 Overview

step—linearization—of the model solution stage. Chapter 3 then introduces the three benchmark models that serve as examples throughout the text.

Part II contains two chapters devoted to model-solution techniques. Chapter 4 describes solution techniques applicable to linear approximations of systems of expectational difference equations. Chapter 5 then describes three alternative classes of methods for obtaining nonlinear model approximations: projection, iteration, and perturbation. Although nonlinear approximations are necessarily more accurate than their linear counterparts, linear approximations are nevertheless valuable, as they often serve as the foundation upon which nonlinear representations are constructed.

Part III is devoted to data preparation and representation, and contains three chapters. Chapter 6 presents two important preliminary steps often needed for priming data for empirical analysis: removing trends and isolating cycles. The purpose of these steps is to align what is being measured in the data with what is being modeled by the theory. For example, the separation of trend from cycle is necessary in confronting trending data with models of business cycle activity.

Chapter 7 presents tools used to summarize properties of the data. First, two important reduced-form models are introduced: autoregressive moving average models for individual time series, and vector autoregressive models for sets of time series. These models provide flexible characterizations of the data that can be used as a means of calculating a wide range of important summary statistics. Next, a collection of popular summary statistics (along with algorithms available for calculating them) are introduced. These statistics often serve as targets for estimating the parameters of structural models, and as benchmarks for judging their empirical performance. Empirical analyses involving collections of summary statistics are broadly categorized as limited-information analyses.

Chapter 8 concludes Part III with an overview of state-space representations. Such representations characterize the evolution over time of two types of variables: those that are observed (possibly subject to measurement error) by the empirical analyst, and those that are not. Conditional on the parameterization of a given model, and the evolution of the observable variables through a given point in time, evolution of the unobservable variables up to the same point in time may be inferred through a process known as filtering. As a by-product of the filtering process, the likelihood function of the parameterized model can be derived. Chapter 8 first demonstrates the mapping of a given DSGE model into a generic state-space representation. It then provides a general overview of likelihood evaluation and filtering, and then characterizes the Kalman filter, which achieves likelihood evaluation and filtering analytically for the special case in which stochastic

Like sick  
Model  
parameters  
new

innovations and measurement errors are Normally distributed, and model representations are linear. Chapter 8 concludes by characterizing a special class of reduced-form state-space representations: dynamic factor models.

With the exception of the linear/Normal class of state-space representations, likelihood evaluation and filtering requires the calculation of integrals that cannot be computed analytically, but instead must be approximated numerically. Part IV contains two chapters that characterize numerical integration techniques, or Monte Carlo methods. Chapter 9 presents four methods, all in the context of generic integration problems: direct Monte Carlo integration; model simulation; importance sampling; and Markov Chain Monte Carlo. Chapter 10 then characterizes the application of numerical integration techniques to the specific objective of likelihood evaluation and filtering in applications involving state-space representations. The particle filter is first introduced as a foundational method for achieving these objectives. Next, concerns regarding numerical efficiency are introduced, with an eye towards achieving efficiency enhancements through the process of adaption. The EIS filter is then introduced generically as a specific means of achieving such enhancements. Finally, the specific application of the EIS filter to state-space representations associated with DSGE models is outlined in detail.

Taking as input the representation of a given DSGE model in a state-space framework, Part V concludes the text with four chapters devoted to alternative methods for achieving empirical implementation. Specifically, chapters 11 through 14 present the following empirical methodologies: calibration, methods of moments, maximum likelihood, and Bayesian inference. Each chapter contains a general presentation of the methodology, and then presents applications to the benchmark models in pursuit of alternative empirical objectives. (For an alternative textbook presentation of these methods, see Canova, 2007.)

Chapter 11 presents the most basic empirical methodology covered in the text: the calibration exercise, as pioneered by Kydland and Prescott (1982). Original applications of this exercise sought to determine whether models designed and parameterized to provide an empirically relevant account of long-term growth were also capable of accounting for the nature of short-term fluctuations that characterize business-cycle fluctuations, summarized using collections of sample statistics measured in the data. More generally, implementation begins with the identification of a set of empirical measurements that serve as constraints on the parameterization of the model under investigation: parameters are chosen to ensure that the model can successfully account for these measurements. (It is often the case that certain parameters must also satisfy additional *a priori* considerations.) Next, implications of the duly parameterized model for an additional set of statistical measurements are compared with their empirical

counterparts to judge whether the model is capable of providing a successful account of these additional features of the data. A challenge associated with this methodology arises in judging success, because this second-stage comparison is made in the absence of a formal statistical foundation.

The method-of-moments techniques presented in chapter 12 serve as one way to address problems arising from the statistical informality associated with calibration exercises. Motivation for their implementation stems from the fact that there is statistical uncertainty associated with the set of empirical measurements that serve as constraints in the parameterization stage of a calibration exercise. For example, a sample mean has an associated sample standard error. Thus there is also statistical uncertainty associated with model parameterizations derived from mappings onto empirical measurements (referred to generally as statistical moments). Method-of-moment estimation techniques account for this uncertainty formally: the parameterizations they yield are interpretable as estimates, featuring classical statistical characteristics. Moreover, if the number of empirical targets used in obtaining parameter estimates exceeds the number of parameters being estimated (i.e., if the model in question is overidentified), the estimation stage also yields objective goodness-of-fit measures that can be used to judge the model's empirical performance. Prominent examples of moment-matching techniques include the generalized and simulated methods of moments (GMM and SMM), and indirect-inference methods.

Moment-matching techniques share a common trait: they are based on a subset of information available in the data (the targeted measurements selected in the estimation stage). An attractive feature of these methodologies is that they may be implemented in the absence of explicit assumptions regarding the underlying distributions that govern the stochastic behavior of the variables featured in the model. A drawback is that decisions regarding the moments chosen in the estimation stage are often arbitrary, and results (e.g., regarding fit) can be sensitive to particular choices. Chapters 13 and 14 present full-information counterparts to these methodologies: likelihood-based analyses. Given a distributional assumption regarding sources of stochastic behavior in a given model, chapter 13 details how the full range of empirical implications of the model may be assessed via maximum-likelihood analysis, facilitated either by use of the Kalman filter or a numerical counterpart. Parameter estimates and model evaluation are facilitated in a straightforward way using maximum-likelihood techniques. Moreover, given model estimates, the implied behavior of unobservable variables present in the model (e.g., productivity shocks) may be inferred as a by-product of the estimation stage.

A distinct advantage in working directly with structural models is that, unlike their reduced-form counterparts, one often has clear *a priori* guidance concerning their parameterization. For example, specifications of

subjective annual discount rates that exceed 10% may be dismissed out of hand as implausible. This motivates the subject of chapter 14, which details the adoption of a Bayesian perspective in bringing full-information procedures to bear in working with structural models. From the Bayesian perspective, a priori views on model parameterization may be incorporated formally in the empirical analysis, in the form of a prior distribution. Coupled with the associated likelihood function via Bayes' Rule, the corresponding posterior distribution may be derived; this conveys information regarding the relative likelihood of alternative parameterizations of the model, conditional on the specified prior and observed data. In turn, conditional statements regarding the empirical performance of the model relative to competing alternatives, the implied behavior of unobservable variables present in the model, and likely future trajectories of model variables may also be derived.

In the spirit of reducing barriers to entry into the field, we have developed a textbook website that contains the data sets that serve as examples throughout the text, as well as computer code used to execute the methodologies we present. The code is in the form of procedures written in the GAUSS programming language. Instructions for executing the procedures are provided within the individual files. The website address is <http://www.pitt.edu/~dejong/text.htm>. References to procedures available via this site are provided throughout this book. In addition, a host of freeware is available throughout the Internet. In searching for code, good starting points include the collection housed by Christian Zimmerman in his Quantitative Macroeconomics web page, and the collection of programs that comprise DYNARE: <http://dge.repec.org/>, <http://www.dynare.org/>.

Much of the code provided at our website reflects the modification of code developed by others, and we have attempted to indicate this explicitly whenever possible. Beyond this attempt, we express our gratitude to the many generous programmers who have made their code available for public use. Finally, we thank Ryan Davis and Lauren Donahoe for assistance with updating the data sets reported in the first edition of the text.

## Chapter 2

### Casting Models in Canonical Form

We could, of course, use any notation we want; do not laugh at notations; invent them, they are powerful. In fact, mathematics is, to a large extent, invention of better notations.

—Richard P. Feynman

One of the cleverest features of the rational expectations revolution was the application of the term "rational." Thereby, the opponents of this approach were forced into the defensive position of either being irrational or of modelling others as irrational, neither of which are comfortable positions for an economist.

—R. J. Barro

### 2.1 Notation

A common set of notations is used throughout the text in presenting models and empirical methodologies. A summary of it follows. The complete set of variables associated with a particular model is represented by the  $n \times 1$  vector of stationary variables  $z_t$ , the  $i^{\text{th}}$  component of which is given by  $z_{it}$ . Among other things, stationarity implies that the unconditional expectation of the level of  $z_t$  does not depend on time, and thus the individual elements  $z_{it}$  show no tendency towards long-term growth. For a given model, this may necessitate the normalization of trending variables prior to analysis. Specific examples of normalization are presented in chapter 6 (specifically, in sections 1.2 and 2.1.4).

The environment and corresponding first-order conditions associated with any given DSGE model can be converted into a nonlinear first-order system of expectational difference equations. We represent such systems as

$$\Gamma(E_t z_{t+1}, z_t, v_{t+1}) = 0, \quad (2.1)$$

where 0 is an  $n \times 1$  vector of zeros,  $v_t$  is an  $m \times 1$  vector of structural shocks, and  $E_t z_{t+1}$  is the expectation of  $z_{t+1}$  formed by the model's

decision makers conditional on information available up to and including period  $t$ :

$$\begin{aligned} E_t z_{t+1} &= E(z_{t+1}|z_t, z_{t-1}, \dots, z_0) \\ &\equiv E(z_{t+1}|z^t). \end{aligned}$$

In this sense, expectations are said to be formed rationally. In addition to being conditional on  $z^t$ , it is typically assumed that expectations are formed given full information regarding the decision makers' environment (including the  $k \times 1$  vector  $\mu$ , denoting the collection of parameters associated with the model). For a textbook treatment of extensions wherein decision makers face uncertainty regarding aspects of their environment, see Hansen and Sargent (2008).

Rewriting forecasted variables as the composition of *ex post* realizations and forecast errors, we introduce expectations errors into the system, which becomes

$$\Gamma(z_{t+1}, z_t, v_{t+1}, \eta_{t+1}) = 0, \quad (2.2)$$

where  $\eta_t$  is an  $r \times 1$  vector of expectational errors associated with intertemporal optimality conditions. Note that  $\eta_t = f(v_t)$ ; that is, expectational errors arise from the realization of structural shocks. The deterministic steady state of the model is expressed as  $\bar{z}$ , defined in the context of (2.1) as satisfying

$$\Gamma(\bar{z}, \bar{z}, 0) = 0. \quad (2.3)$$

The components of  $z_t$  belong to one of three classifications: exogenous state variables, endogenous state variables, and control variables. The latter are denoted by the  $n_c \times 1$  vector  $c_t$ , and represent optimal choices by decision makers taking as given values of the state variables inherited in period  $t$ . Exogenous state variables evolve over time independently of decision makers' choices, while the evolution of endogenous state variables is influenced by these choices. Collectively, state variables are denoted by the  $n_s \times 1$  vector  $s_t$ , where  $n_c + n_s = n$ .

Often it will be useful to express variables as logged deviations from steady state values. This is done using tildes:

$$\tilde{z}_{it} = \ln \left( \frac{z_{it}}{\bar{z}_i} \right).$$

In turn, the vector  $x_t$  denotes the collection of model variables written (unless indicated otherwise) in terms of logged deviations from steady state values. So for a model consisting of output  $y_t$ , investment  $i_t$ , and labor hours  $n_t$ ,  $x_t$  is given by

$$x_t = [\tilde{y}_t \quad \tilde{i}_t \quad \tilde{n}_t]'.$$

## 2.1 Notation

### 2.1.1 Log-Linear Model Representations

Log-linear representations of structural models are expressed as

$$Ax_{t+1} = Bx_t + Cv_{t+1} + D\eta_{t+1}, \quad (2.4)$$

where the elements of  $A$ ,  $B$ ,  $C$ , and  $D$  are functions of the structural parameters  $\mu$ . Solutions of (2.4) are expressed as

$$x_{t+1} = F(\mu)x_t + G(\mu)v_{t+1}. \quad (2.5)$$

In (2.5), certain variables in the vector  $x_t$  are unobservable, whereas others (or linear combinations of variables) are observable. Thus filtering methods such as the Kalman filter must be used to evaluate the system empirically. The Kalman filter requires an observer equation linking observables to unobservables. Observable variables are denoted by  $X_t$ , where

$$X_t = H(\mu)'x_t + u_t, \quad (2.6)$$

with

$$E(u_t u_t') = \Sigma_u.$$

The presence of  $u_t$  in (2.6) reflects the possibility that the observations of  $X_t$  are associated with measurement error. Finally, defining

$$e_{t+1} = G(\mu)v_{t+1},$$

the covariance matrix of  $e_{t+1}$  is given by

$$Q(\mu) = E(e_{t+1} e_{t+1}'). \quad (2.7)$$

Given assumptions regarding the stochastic nature of measurement errors and the structural shocks, (2.5)–(2.7) yield a log-likelihood function  $\log L(X|\Lambda)$ , where  $\Lambda$  collects the parameters in  $F(\mu)$ ,  $H(\mu)$ ,  $\Sigma_u$ , and  $Q(\mu)$ . Often it will be convenient to take as granted mappings from  $\mu$  to  $F$ ,  $H$ ,  $\Sigma_u$ , and  $Q$ . In such cases the likelihood function will be written as  $L(X|\mu)$ .

### 2.1.2 Nonlinear Model Representations

Nonlinear approximations of structural models are represented using three sets of equations, written with variables expressed in terms of levels (again, possibly normalized to eliminate trend behavior). The first characterizes the evolution of the state variables included in the model:

$$s_t = f(s_{t-1}, v_t), \quad (2.8)$$

where once again  $v_t$  denotes the collection of structural shocks. The second set of equations are known as policy functions, which represent the optimal

specification of the control variables included in the model as a function of the state variables:

$$c_t = c(s_t). \quad (2.9)$$

The third set of equations map the full collection of model variables into the observables:

$$X_t = \tilde{g}(s_t, c_t, v_t, u_t) \quad (2.10)$$

$$\equiv g(s_t, u_t), \quad (2.11)$$

where once again  $u_t$  denotes measurement error. Parameters associated with  $f(s_{t-1}, v_t)$ ,  $c(s_t)$ , and  $g(s_t, u_t)$  are again obtained as mappings from  $\mu$ , thus their associated likelihood function is also written as  $L(X|\mu)$ .

Note that the solution to the log-linear system (2.5) can be converted to a form analogous to (2.9) by solving for the control variables contained in  $x_t$  as functions of the state variables contained in  $s_t$ . A straightforward means of executing this conversion is presented in chapter 5.

In the remainder of this chapter, we outline procedures for mapping nonlinear systems into (2.4). Then in chapter 3 we introduce the three benchmark models that serve as examples through the remainder of the text, and demonstrate their mapping into (2.4).

## 2.2 Linearization

### 2.2.1 Taylor Series Approximation

Consider the following simplification of (2.1) given by

$$\Psi(z_{t+1}, z_t) = 0, \quad (2.12)$$

where we have abstracted momentarily from the stochastic component of the model under investigation. We have done so because models are typically designed to incorporate stochastic behavior directly into the linearized system (including those introduced in chapter 3). Also, expectational terms need not be singled out at this point, because they receive no special treatment in the linearization stage.

Before proceeding, note that although (2.12) is written as a first-order system, higher-order specifications may be written as first-order systems by augmenting  $z_t$  to include variables observed at different points in time. For example, the  $p^{\text{th}}$ -order equation

$$\omega_{t+1} = \rho_1 \omega_t + \rho_2 \omega_{t-1} + \dots + \rho_p \omega_{t-p+1}$$

### 2.2 Linearization

can be written in first-order form as

$$\begin{bmatrix} \omega_{t+1} \\ \omega_t \\ \vdots \\ \omega_{t-p+2} \end{bmatrix} - \begin{bmatrix} \rho_1 & \rho_2 & \dots & \dots & \rho_p \\ 1 & 0 & \dots & \dots & 0 \\ \vdots & \vdots & \ddots & \dots & \vdots \\ 0 & 0 & \dots & 1 & 0 \end{bmatrix} \begin{bmatrix} \omega_t \\ \omega_{t-1} \\ \vdots \\ \omega_{t-p+1} \end{bmatrix} = 0,$$

or more compactly, as

$$z_{t+1} - \Pi z_t = 0, \quad z_{t+1} = [\omega_{t+1}, \omega_t, \dots, \omega_{t-p+2}]'$$

Thus (2.12) is sufficiently general to characterize a system of arbitrary order.

The goal of the linearization step is to convert (2.12) into a linear system, which can then be solved using any of the procedures outlined in chapter 4. The reason for taking this step is that explicit solutions to (2.12) are typically unavailable, rendering quantitative assessments of the system problematic. For textbook discussions of the analysis of nonlinear systems, see Azariadis (1993) and Sedaghat (2003).

In light of (2.4), the representation of the system we seek is given by

$$Ax_{t+1} = Bx_t, \quad (2.13)$$

where  $x_t$  represents a transformation of  $z_t$ . One way of establishing this representation is via linearization, which is accomplished via a first-order Taylor Series approximation of (2.12) around its steady state:

$$0 \approx \Psi(\bar{z}) + \frac{\partial \Psi}{\partial z_t}(\bar{z}) \times (z_t - \bar{z}) + \frac{\partial \Psi}{\partial z_{t+1}}(\bar{z}) \times (z_{t+1} - \bar{z}), \quad (2.14)$$

where  $(z_t - \bar{z})$  is  $n \times 1$ , and the  $n \times n$  matrix  $\frac{\partial \Psi}{\partial z_t}(\bar{z})$  denotes the Jacobian of  $\Psi(z_{t+1}, z_t)$  with respect to  $z_t$  evaluated at  $\bar{z}$ . That is, the  $(i, j)^{\text{th}}$  element of  $\frac{\partial \Psi}{\partial z_t}(\bar{z})$  is the derivative of the  $i^{\text{th}}$  equation in (2.12) with respect to the  $j^{\text{th}}$  element of  $z_t$ . Defining

$$A = \frac{\partial \Psi}{\partial z_{t+1}}(\bar{z}), \quad B = -\frac{\partial \Psi}{\partial z_t}(\bar{z}), \quad x_t = (z_t - \bar{z}),$$

we obtain (2.13), wherein variables are expressed as deviations from steady state values. (As we shall see in chapter 4, it is also possible to work with higher-order approximations of (2.12), e.g., following Schmitt-Grohé and Uribe, 2004.)

### 2.2.2 Log-Linear Approximations

It is often useful to work with log-linear approximations of (2.12), due in part to their ease of interpretation. For illustration, we begin with a simple example in which the system is  $1 \times 1$ , and can be written as

$$z_{t+1} = f(z_t).$$

Taking natural logs and noting that  $z_t = e^{\ln z_t}$ , the system becomes

$$\ln z_{t+1} = \ln[f(e^{\ln z_t})].$$

Then approximating,

$$\ln z_{t+1} \approx \ln[f(\bar{z})] + \frac{f'(\bar{z})\bar{z}}{f(\bar{z})}(\ln(z_t) - \ln(\bar{z})),$$

or because  $\ln[f(\bar{z})] = \ln \bar{z}$ ,

$$\ln\left(\frac{z_{t+1}}{\bar{z}}\right) \approx \frac{f'(\bar{z})\bar{z}}{f(\bar{z})}\left(\ln\left(\frac{z_t}{\bar{z}}\right)\right).$$

Note that  $\frac{f'(\bar{z})}{f(\bar{z})}$  is the elasticity of  $z_{t+1}$  with respect to  $z_t$ . Moreover, writing  $z_t$  as  $\bar{z} + \varepsilon_t$ , where  $\varepsilon_t$  denotes a small departure from steady state,

$$\ln\left(\frac{z_t}{\bar{z}}\right) = \ln\left(1 + \frac{\varepsilon_t}{\bar{z}}\right) \approx \frac{\varepsilon_t}{\bar{z}},$$

and thus  $\ln\left(\frac{z_t}{\bar{z}}\right)$  is seen as expressing  $z_t$  in terms of its percentage deviation from steady state.

Returning to the  $n \times 1$  case, rewrite (2.12) as

$$\Psi_1(z_{t+1}, z_t) = \Psi_2(z_{t+1}, z_t), \quad (2.15)$$

because it is not possible to take logs of both sides of (2.12). Again using  $z_t = e^{\ln z_t}$ , taking logs of (2.15) and rearranging yields

$$\ln \Psi_1(e^{\ln z_{t+1}}, e^{\ln z_t}) - \ln \Psi_2(e^{\ln z_{t+1}}, e^{\ln z_t}) = 0. \quad (2.16)$$

The first-order Taylor Series approximation of this converted system yields the log-linear approximation we seek. The approximation for the first term is

$$\begin{aligned} \ln \Psi_1(z_{t+1}, z_t) &\approx \ln[\Psi_1(\bar{z})] + \frac{\partial \ln[\Psi_1]}{\partial \ln(z_t)}(\bar{z}) \times \left[ \ln\left(\frac{z_t}{\bar{z}}\right) \right] \\ &+ \frac{\partial \ln[\Psi_1]}{\partial \ln(z_{t+1})}(\bar{z}) \times \left[ \ln\left(\frac{z_{t+1}}{\bar{z}}\right) \right], \end{aligned} \quad (2.17)$$

where  $\frac{\partial \ln[\Psi_1]}{\partial \ln(z_t)}(\bar{z})$  and  $\frac{\partial \ln[\Psi_1]}{\partial \ln(z_{t+1})}(\bar{z})$  are  $n \times n$  Jacobian matrices, and  $\left[ \ln\left(\frac{z_t}{\bar{z}}\right) \right]$  and  $\left[ \ln\left(\frac{z_{t+1}}{\bar{z}}\right) \right]$  are  $n \times 1$  vectors. The approximation of the second term in (2.16) is analogous. Then defining

### 2.2 Linearization

$$\begin{aligned} A &= \left[ \frac{\partial \ln[\Psi_1]}{\partial \ln(z_{t+1})}(\bar{z}) - \frac{\partial \ln[\Psi_2]}{\partial \ln(z_{t+1})}(\bar{z}) \right], \\ B &= - \left[ \frac{\partial \ln[\Psi_1]}{\partial \ln(z_t)}(\bar{z}) - \frac{\partial \ln[\Psi_2]}{\partial \ln(z_t)}(\bar{z}) \right], \\ x_t &= \left[ \ln\left(\frac{z_t}{\bar{z}}\right) \right], \end{aligned}$$

we once again obtain (2.13). The elements of  $A$  and  $B$  are now elasticities, and the variables of the system are expressed in terms of percentage deviations from steady state.

In Part V we will discuss several empirical applications that involve the need to approximate (2.12) or (2.16) repeatedly for alternative values of  $\mu$ . In such cases it is useful to automate the linearization stage via the use of a numerical gradient calculation procedure. We introduce this briefly here in the context of approximating (2.12); the approximation of (2.16) is analogous.

Gradient procedures are designed to construct the Jacobian matrices in (2.14) or (2.17) without analytical expressions for the required derivatives. Derivatives are instead calculated numerically, given the provision of three components by the user. The first two components are a specification of  $\mu$  and a corresponding specification of  $\bar{z}$ . The third component is a procedure designed to return the  $n \times 1$  vector of values  $\xi$  generated by (2.12) for two cases. In the first case  $z_{t+1}$  is treated as variable and  $z_t$  is fixed at  $\bar{z}$ ; in the second case  $z_t$  is treated as variable and  $z_{t+1}$  is fixed at  $\bar{z}$ . The gradient procedure delivers the Jacobian

$$\frac{\partial \Psi}{\partial z_{t+1}}(\bar{z}) = A$$

in the first case and

$$\frac{\partial \Psi}{\partial z_t}(\bar{z}) = -B$$

in the second case.

Before turning to the fully specified example models provided in chapter 3, we conclude here by demonstrating the linearization and log-linearization of two example equations.

### 2.2.3 Example Equations

Consider the simple resource constraint

$$y_t = c_t + i_t,$$

indicating that output ( $y_t$ ) can be either consumed ( $c_t$ ) or invested ( $i_t$ ). This equation is already linear. In the notation of (2.12) the equation appears as

$$y_t - c_t - i_t = 0;$$

and in terms of (2.14), with

$$z_t = [y_t \ c_t \ i_t]'$$

and the equation representing the  $i^{\text{th}}$  of the system, the  $i^{\text{th}}$  row of  $\frac{\partial \Psi}{\partial z_t}(\bar{z})$  is given by

$$\frac{\partial \Psi}{\partial z_i}(\bar{z}) = [1 \ -1 \ -1].$$

In the notation of (2.16), the equation appears as

$$\ln y_t - \ln[\exp(\ln c_t) - \exp(\ln i_t)] = 0,$$

and in terms of (2.13), the  $i^{\text{th}}$  row of the right-hand-side matrix  $B$  is

$$\left[ \frac{\partial \ln[\Psi_1]}{\partial \ln(z_{it})}(\bar{z}) - \frac{\partial \ln[\Psi_2]}{\partial \ln(z_i)}(\bar{z}) \right] = \left[ 1 \ \frac{-\bar{c}}{\bar{c} + \bar{i}} \ \frac{-\bar{i}}{\bar{c} + \bar{i}} \right]. \quad (2.18)$$

Finally, to use a gradient procedure to accomplish log-linear approximation, the  $i^{\text{th}}$  return of the system-evaluation procedure is

$$\varsigma_i = \ln y_t - \ln[\exp(\ln c_t) - \exp(\ln i_t)].$$

As an additional example consider the Cobb-Douglas production function

$$y_t = a_t k_t^\alpha n_t^{1-\alpha}, \quad \alpha \in (0, 1),$$

where output is produced by the use of capital ( $k_t$ ) and labor ( $n_t$ ) and is subject to a technology or productivity shock ( $a_t$ ). Linear approximation of this equation is left as an exercise. To accomplish log-linear approximation, take logs of the equation and rearrange to map into the notation of (2.16) as

$$\ln y_t - \ln a_t - \alpha \ln k_t - (1 - \alpha) \ln n_t = 0.$$

With

$$z_t = \left[ \ln \frac{y_t}{\bar{y}} \ \ln \frac{a_t}{\bar{a}} \ \ln \frac{k_t}{\bar{k}} \ \ln \frac{n_t}{\bar{n}} \right]',$$

## 2.2 Linearization

the  $i^{\text{th}}$  row of the right-hand-side matrix in (2.13) is

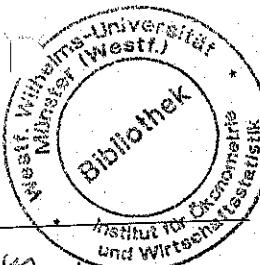
$$\left[ \frac{\partial \ln[\Psi_1]}{\partial \ln(z_t)}(\bar{z}) - \frac{\partial \ln[\Psi_2]}{\partial \ln(z_i)}(\bar{z}) \right] = [1 \ -1 \ -\alpha \ -(1 - \alpha)]. \quad (2.19)$$

And to use a gradient procedure to accomplish log-linear approximation, the  $i^{\text{th}}$  return of the system-evaluation procedure is

$$\varsigma_i = \ln y_t - \ln a_t - \alpha \ln k_t - (1 - \alpha) \ln n_t.$$

## Chapter 3 - Structural

### Macroeconomics DSGE Models: Three Examples



Example is the School of Mankind, and they will learn at no other.

—Edmund Burke, *Thoughts on the Cause of the Present Discontents*

Chapter 2 introduced the generic representation of a DSGE model as

$$\Gamma(E_t z_{t+1}, z_t, v_{t+1}) = 0,$$

and provided background for mapping the model into a log-linear representation of the form

$$Ax_{t+1} = Bx_t + Cv_{t+1} + D\eta_{t+1}.$$

This chapter demonstrates the representation and completion of this mapping for three prototypical model environments that serve as examples throughout the remainder of the text. This sets the stage for chapter 4, which details the derivation of linear model solutions of the form

$$x_{t+1} = Fx_t + Gv_{t+1},$$

and chapter 5, which details the derivation of nonlinear solutions of the form

$$c_t = c(s_t), \\ s_t = f(s_{t-1}, v_t).$$

For guidance regarding the completion of log-linear mappings for a broader range of models than those considered here, see Hansen and Sargent (2005).

The first environment is an example of a simple real business cycle (RBC) framework, patterned after that of Kydland and Prescott (1982). The foundation of models in the RBC tradition is a neoclassical growth environment, augmented with two key features: a labor-leisure trade-off that confronts decision makers, and uncertainty regarding the evolution of technological progress. The empirical question Kydland and Prescott

### 3 DSGE Models: Three Examples

(1982) sought to address was the extent to which such a model, bereft of market imperfections and featuring fully flexible prices, could account for observed patterns of business-cycle activity while capturing salient features of economic growth. This question continues to serve as a central focus of this active literature; overviews are available in the collection of papers presented in Barro (1989) and Cooley (1995).

Viewed through the lens of an RBC model, business cycle activity is interpretable as reflecting optimal responses to stochastic movements in the evolution of technological progress. Such interpretations are not without controversy. Alternative interpretations cite the existence of market imperfections, costs associated with the adjustment of prices, and other nominal and real frictions as potentially playing important roles in influencing business cycle behavior, and giving rise to additional sources of business cycle fluctuations. Initial skepticism of this nature was voiced by Summers (1986), and the papers contained in Mankiw and Romer (1991) provide an overview of DSGE models that highlight the role of, for example, market imperfections in influencing aggregate economic behavior.

As a complement to the RBC environment, the second environment presented here (that of Ireland, 2004a) provides an example of a model within this New Keynesian tradition. The empirical purpose of the model is to evaluate the role of cost, demand, and productivity shocks in driving business cycle fluctuations. Textbook references for models within this tradition are Benassy (2002) and Woodford (2003). In addition, the seminal papers of Christiano, Eichenbaum, and Evans (2005) and Smets and Wouters (2003, 2007) are must-reads for those interested in models of this class: based on the empirical performance of the models these papers describe, New Keynesian models have emerged as important tools for policymakers, particularly central bankers. Yet, in turn, this emergence is not without controversy: for a critique of the empirical performance of these models, see Chari, Kehoe, and McGrattan (2009).

The realm of empirical applications pursued through the use of DSGE models extends well beyond the study of business cycles. The third environment serves as an example of this point: it is a model of asset-pricing behavior adopted from Lucas (1978). The model represents financial assets as tools used by households to optimize intertemporal patterns of consumption in the face of exogenous stochastic movements in income and dividends earned from asset holdings. Viewed through the lens of this model, two particular features of asset-pricing behavior have proven exceptionally difficult to explain. First, LeRoy and Porter (1981) and Shiller (1981) used versions of the model to underscore the puzzling volatility of prices associated with broad indexes of assets (such as the Standard & Poor's 500), highlighting what has come to be known as the "volatility puzzle." Second, Mehra and Prescott (1985) and Weil (1989) used versions of the model to

highlight the puzzling dual phenomenon of a large gap observed between aggregate returns on risky and riskless assets, coupled with exceptionally low returns yielded by riskless assets. These features came to be known as the “equity premium” and “risk-free rate” puzzles. The texts of Shiller (1989), Campbell, Lo, and MacKinlay (1997), and Cochrane (2001) provide overviews of literatures devoted to analyses of these puzzles.

In addition to the references cited above, a host of introductory graduate-level textbooks serve as useful references for the specific example models considered here, and for a wide range of extensions. A partial listing includes Sargent (1987a, 1987b), Stokey and Lucas (1989), Blanchard and Fischer (1998), Romer (2006), Ljungqvist and Sargent (2004), and Wickens (2008).

### 3.1 Model I: A Real Business Cycle Model

#### 3.1.1 Environment

The economy consists of a large number of identical households; aggregate economic activity is analyzed by focusing on a representative household. The household’s objective is to maximize  $U$ , the expected discounted flow of utility arising from chosen streams of consumption and leisure:

$$\max_{c_t, l_t} U = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, l_t). \quad (3.1)$$

In (3.1),  $E_0$  is the expectations operator conditional on information available at time 0,  $\beta \in (0, 1)$  is the household’s subjective discount factor,  $u(\cdot)$  is an instantaneous utility function, and  $c_t$  and  $l_t$  denote levels of consumption and leisure chosen at time  $t$ .

The household is equipped with a production technology that can be used to produce a single good  $y_t$ . The production technology is represented by

$$y_t = z_t f(k_t, n_t), \quad (3.2)$$

where  $k_t$  and  $n_t$  denote quantities of physical capital and labor assigned by the household to the production process, and  $z_t$  denotes a random disturbance to the productivity of these inputs to production (that is, a productivity or technology shock).

Within a period, the household has one unit of time available for division between labor and leisure activities:

$$1 = n_t + l_t. \quad (3.3)$$

#### 3.1 Model I: A Real Business Cycle Model

In addition, output generated at time  $t$  can be either consumed or used to augment the stock of physical capital available for use in the production process in period  $t+1$ . That is, output can be either consumed or invested:

$$y_t = c_t + i_t, \quad (3.4)$$

where  $i_t$  denotes the quantity of investment. Finally, the stock of physical capital evolves according to

$$k_{t+1} = i_t + (1 - \delta)k_t, \quad (3.5)$$

where  $\delta \in (0, 1)$  denotes the depreciation rate. The household’s problem is to maximize (3.1) subject to (3.2)–(3.5), taking  $k_0$  and  $z_0$  as given.

Implicit in the specification of the household’s problem are two sets of trade-offs. One is a consumption/savings trade-off: from (3.4), higher consumption today implies lower investment (savings), and thus from (3.5), less capital available for production tomorrow. The other is a labor/leisure trade-off: from (3.3), higher leisure today implies lower labor today and thus lower output today.

In order to explore quantitative implications of the model, it is necessary to specify explicit functional forms for  $u(\cdot)$  and  $f(\cdot)$ , and to characterize the stochastic behavior of the productivity shock  $z_t$ . We pause before doing so to make some general comments. As noted, an explicit goal of the RBC literature is to begin with a model specified to capture important characteristics of economic growth, and then to judge the ability of the model to capture key components of business cycle activity. From the model builder’s perspective, the former requirement serves as a constraint on choices regarding the specifications for  $u(\cdot)$ ,  $f(\cdot)$ , and the stochastic process of  $z_t$ . Three key aspects of economic growth serve as constraints in this context: over long time horizons the growth rates of  $\{c_t, i_t, y_t, k_t\}$  are roughly equal (balanced growth), the marginal productivity of capital and labor (reflected by relative factor payments) are roughly constant over time, and  $\{l_t, n_t\}$  show no tendencies for long-term growth.

Beyond satisfying this constraint, functional forms chosen for  $u(\cdot)$  are typically strictly increasing in both arguments, twice continuously differentiable, strictly concave, and satisfy

$$\lim_{c \rightarrow 0} \frac{\partial u(c_t, l_t)}{\partial c_t} = \lim_{l \rightarrow 0} \frac{\partial u(c_t, l_t)}{\partial l_t} = \infty. \quad (3.6)$$

Functional forms chosen for  $f(\cdot)$  typically feature constant returns to scale and satisfy similar limit conditions.

Finally, we note that the inclusion of a single source of uncertainty in this framework, via the productivity shock  $z_t$ , implies that the model carries

nontrivial implications for the stochastic behavior of a single corresponding observable variable. For the purposes of this chapter, this limitation is not important; however, it will motivate the introduction of extensions of this basic model in Part V.

### FUNCTIONAL FORMS

The functional forms presented here enjoy prominent roles in the macroeconomics literature. Instantaneous utility is of the constant relative risk aversion (CRRA) form:

$$u(c_t, l_t) = \frac{(c_t^\phi l_t^{1-\phi})^{1-\phi}}{1-\phi}, \quad (3.7)$$

or when  $\phi = 1$ ,  $u(\cdot) = \ln(\cdot)$ . The parameter  $\phi > 0$  determines two attributes: it is the coefficient of relative risk aversion, and also determines the intertemporal elasticity of substitution, given by  $\frac{1}{\phi}$  (for textbook discussions, see e.g., Blanchard and Fischer, 1998; or Romer, 2006). Note that the larger is  $\phi$ , the more intense is the household's interest in maintaining a smooth consumption/leisure profile. Also,  $\phi \in (0, 1)$  indicates the importance of consumption relative to leisure in determining instantaneous utility.

Next, the production function is Cobb-Douglas:

$$y_t = z_t k_t^\alpha n_t^{1-\alpha}, \quad (3.8)$$

where  $\alpha \in (0, 1)$  represents capital's share of output. Finally, the log of the technology shock is assumed to follow a first-order autoregressive, or AR(1), process:

$$\ln z_t = (1 - \rho) \ln (\bar{z}) + \rho \ln z_{t-1} + \varepsilon_t \quad (3.9)$$

$$\varepsilon_t \sim NID(0, \sigma^2), \quad \rho \in (-1, 1). \quad (3.10)$$

The solution to the household's problem may be obtained via standard application of the theory of dynamic programming (e.g., as described in detail in Stokey and Lucas, 1989; and briefly in chapter 4 of this text). Necessary conditions associated with the household's problem expressed in general terms are given by

$$\frac{\partial u(c_t, l_t)}{\partial l_t} = \left\{ \frac{\partial u(c_t, l_t)}{\partial c_t} \right\} \times \left\{ \frac{\partial f(k_t, n_t)}{\partial n_t} \right\} \quad (3.11)$$

$$\frac{\partial u(c_t, l_t)}{\partial c_t} = \beta E_t \left\{ \frac{\partial u(c_{t+1}, l_{t+1})}{\partial c_{t+1}} \left[ \frac{\partial f(k_{t+1}, n_{t+1})}{\partial k_{t+1}} + (1 - \delta) \right] \right\}. \quad (3.12)$$

### 3.1 Model I: A Real Business Cycle Model

The intratemporal optimality condition (3.11) equates the marginal benefit of an additional unit of leisure time with its opportunity cost: the marginal value of the foregone output resulting from the corresponding reduction in labor time. The intertemporal optimality condition (3.12) equates the marginal benefit of an additional unit of consumption today with its opportunity cost: the discounted expected value of the additional utility tomorrow that the corresponding reduction in savings would have generated (higher output plus undepreciated capital).

Consider the qualitative implications of (3.11) and (3.12) for the impact of a positive productivity shock on the household's labor/leisure and consumption/savings decisions. From (3.11), higher labor productivity implies a higher opportunity cost of leisure, prompting a reduction in leisure time in favor of labor time. From (3.12), the curvature in the household's utility function carries with it a consumption-smoothing objective. A positive productivity shock serves to increase output, thus affording an increase in consumption; however, because the marginal utility of consumption is decreasing in consumption, this drives down the opportunity cost of savings. The greater is the curvature of  $u(\cdot)$ , the more intense is the consumption-smoothing objective, and thus the greater will be the intertemporal reallocation of resources in the face of a productivity shock.

Dividing (3.11) by the expression for the marginal utility of consumption, and using the functional forms introduced above, these conditions can be written as

$$\left( \frac{1 - \phi}{\phi} \right) \frac{c_t}{l_t} = (1 - \alpha) z_t \left( \frac{k_t}{n_t} \right)^\alpha \quad (3.13)$$

$$c_t^{\phi(1-\phi)-1} l_t^{(1-\phi)(1-\phi)} = \beta E_t \left\{ c_{t+1}^{\phi(1-\phi)-1} l_{t+1}^{(1-\phi)(1-\phi)} \left[ \alpha z_{t+1} \left( \frac{n_{t+1}}{k_{t+1}} \right)^{1-\alpha} + (1 - \delta) \right] \right\}. \quad (3.14)$$

#### 3.1.2 The Nonlinear System

Collecting components, the system of nonlinear stochastic difference equations that comprise the model is given by

$$\left( \frac{1 - \phi}{\phi} \right) \frac{c_t}{l_t} = (1 - \alpha) z_t \left( \frac{k_t}{n_t} \right)^\alpha \quad (3.15)$$

$$c_t^\kappa l_t^\lambda = \beta E_t \left\{ c_{t+1}^\kappa l_{t+1}^\lambda \left[ \alpha z_{t+1} \left( \frac{n_{t+1}}{k_{t+1}} \right)^{1-\alpha} + (1 - \delta) \right] \right\} \quad (3.16)$$

## 3 DSGE Models: Three Examples

$$y_t = z_t k_t^\alpha n_t^{1-\alpha} \quad (3.17)$$

$$y_t = c_t + i_t \quad (3.18)$$

$$k_{t+1} = i_t + (1 - \delta)k_t \quad (3.19)$$

$$l = n_t + l_t \quad (3.20)$$

$$\ln z_t = (1 - \rho) \ln(\bar{z}) + \rho \ln z_{t-1} + \varepsilon_t, \quad (3.21)$$

where

$$\kappa = \varphi(1 - \phi) - 1$$

and

$$\lambda = (1 - \varphi)(1 - \phi).$$

This is the specific form of the generic nonlinear model representation

$$\Gamma(E_t z_{t+1}, z_t, v_{t+1}) = 0$$

introduced in chapter 2, with the slight abuse of notation involving the interpretation of  $z_t$ . This representation serves as the point of departure for deriving solutions of the form

$$\begin{aligned} c_t &= c(s_t), \\ s_t &= f(s_{t-1}, v_t), \end{aligned}$$

using the nonlinear solution methods described in chapter 5.

Steady states of the variables  $\{y_t, c_t, i_t, n_t, l_t, k_t, z_t\}$  may be computed analytically from this system. These are derived by holding  $z_t$  to its steady state value  $\bar{z}$ , which we set to 1:

$$\begin{aligned} \frac{\bar{y}}{\bar{n}} &= \eta, \\ \frac{\bar{c}}{\bar{n}} &= \eta - \delta\theta, \\ \frac{\bar{i}}{\bar{n}} &= \delta\theta, \\ \bar{n} &= \frac{1}{1 + \left(\frac{1}{1-\alpha}\right) \left(\frac{1-\varphi}{\varphi}\right) [1 - \delta\theta^{1-\alpha}]}, \quad (3.22) \\ \bar{l} &= 1 - \bar{n}, \end{aligned}$$

$$\frac{\bar{k}}{\bar{n}} = \theta,$$

## 3.1 Model I: A Real Business Cycle Model

where

$$\begin{aligned} \theta &= \left(\frac{\alpha}{1/\beta - 1 + \delta}\right)^{\frac{1}{1-\alpha}} \\ \eta &= \theta^\alpha. \end{aligned}$$

Note that in steady state the variables  $\{y_t, c_t, i_t, k_t\}$  do not grow over time. Implicitly, these variables are represented in the model in terms of deviations from trend, and steady state values indicate the relative heights of trend lines. To incorporate growth explicitly, consider an alternative specification of  $z_t$ :

$$z_t = z_0 (1 + g)^t e^{\omega_t}, \quad (3.23)$$

$$\omega_t = \rho \omega_{t-1} + \varepsilon_t. \quad (3.24)$$

Note that, absent shocks, the growth rate of  $z_t$  is given by  $g$ , and that removal of the trend component  $(1 + g)^t$  from  $z_t$  yields the specification for  $\ln z_t$  given by (3.21). Further, the reader is invited to verify that under this specification for  $z_t$ ,  $\{c_t, i_t, y_t, k_t\}$  will have a common growth rate given by  $\frac{g}{1-\alpha}$ . Thus the model is consistent with the balanced-growth requirement, and as specified, all variables are interpreted as being measured in terms of deviations from their common trend.

One subtlety is associated with the issue of trend removal that arises in dealing with the dynamic equations of the system. Consider the law of motion for capital (3.19). Trend removal here involves division of both sides of (3.19) by  $(1 + \frac{g}{1-\alpha})^t$ ; however, the trend component associated with  $k_{t+1}$  is  $(1 + \frac{g}{1-\alpha})^{t+1}$ , so the specification in terms of detrended variables is

$$\left(1 + \frac{g}{1-\alpha}\right) k_{t+1} = i_t + (1 - \delta)k_t. \quad (3.25)$$

Likewise, a residual trend factor will be associated with  $c_{t+1}$  in the intertemporal optimality condition (3.16). Because  $c_{t+1}$  is raised to the power

$$\kappa = \varphi(1 - \phi) - 1,$$

the residual factor is given by  $\left(1 + \frac{g}{1-\alpha}\right)^\kappa$ :

$$\begin{aligned} c_t^\kappa l_t^\lambda &= \beta E_t \\ &\times \left\{ \left(1 + \frac{g}{1-\alpha}\right)^\kappa c_{t+1}^\kappa l_{t+1}^\lambda \left[ \alpha z_{t+1} \left( \frac{n_{t+1}}{k_{t+1}} \right)^{1-\alpha} + (1 - \delta) \right] \right\}. \quad (3.26) \end{aligned}$$

With  $\kappa$  negative (ensured by  $\frac{1}{\phi} < 1$ , i.e., an inelastic intertemporal elasticity of substitution specification), the presence of  $g$  provides an incentive to shift resources away from  $(t+1)$  towards  $t$ .

### Exercise 3.1

Rederive the steady state expressions (3.22) by replacing (3.19) with (3.25), and (3.16) with (3.26). Interpret the intuition behind the impact of  $g$  on the expressions you derive.

#### 3.1.3 Log-Linearization

This step involves taking a log-linear approximation of the model at steady state values. In this case, the objective is to map (3.15)–(3.21) into the linearized system

$$Ax_{t+1} = Bx_t + Cv_{t+1} + D\eta_{t+1}$$

for eventual empirical evaluation. Regarding  $D$ , dropping  $E_t$  from the Euler equation (3.16) introduces an expectations error in the model's second equation, therefore

$$D = [0 \ 1 \ 0 \ 0 \ 0 \ 0 \ 0]'$$

Likewise, the presence of the productivity shock in the model's seventh equation (3.21) implies

$$C = [0 \ 0 \ 0 \ 0 \ 0 \ 0 \ 1]'$$

Regarding  $A$  and  $B$ , these can be constructed by introducing the following system of equations into a gradient procedure (where time subscripts are dropped so that, for example,  $y = y_t$  and  $y' = y_{t+1}$ ):

$$0 = \ln \left( \frac{1-\varphi}{\varphi} \right) + \ln c' - \ln l' - \ln(1-\alpha) - \ln z' \\ - \alpha \ln k' + \alpha \ln n' \quad (3.27)$$

$$0 = \kappa \ln c + \lambda \ln l - \ln \beta - \kappa \ln c' - \lambda \ln l' \quad (3.28)$$

$$- \ln \left[ \alpha \exp(\ln z') \frac{\exp[(1-\alpha)\ln n']}{\exp[(1-\alpha)\ln k']} + (1-\delta) \right] \quad (3.29)$$

$$0 = \ln y' - \ln z' - \alpha \ln k' - (1-\alpha) \ln n' \quad (3.30)$$

$$0 = \ln k' - \ln \{ \exp[\ln(c')] + \exp[\ln(i')] \} \quad (3.31)$$

$$0 = -\ln \{ \exp[\ln(n')] + \exp[\ln(l')] \} \quad (3.32)$$

$$\hat{\gamma} = \dots \quad (3.33)$$

### 3.1 Model I: A Real Business Cycle Model

The mapping from (3.15)–(3.21) to (3.27)–(3.33) involves four steps. First, logs of both sides of each equation are taken; second, all variables not converted into logs in the first step are converted using the fact, for example, that  $y = \exp(\ln(y))$ ; third, all terms are collected on the right-hand side of each equation; fourth, all equations are multiplied by  $-1$ . Derivatives taken with respect to  $\ln y'$ , and so on evaluated at steady state values yield  $A$ , and derivatives taken with respect to  $\ln y$ , and so on yield  $-B$ .

Having obtained  $A$ ,  $B$ ,  $C$ , and  $D$ , the system can be solved using any of the solution methods to be presented in chapter 4, yielding a system of the form

$$x_{t+1} = F(\mu)x_t + e_{t+1}.$$

### Exercise 3.2

With  $x_t$  given by

$$x_t = \left[ \ln \frac{y_t}{\bar{y}}, \ln \frac{c_t}{\bar{c}}, \ln \frac{i_t}{\bar{i}}, \ln \frac{n_t}{\bar{n}}, \ln \frac{l_t}{\bar{l}}, \ln \frac{k_t}{\bar{k}}, \ln \frac{z_t}{\bar{z}} \right]'$$

and

$$\mu = [\alpha \ \beta \ \varphi \ \delta \ \rho \ \sigma]' = [0.33 \ 0.975 \ 2 \ 0.5 \ 0.06 \ 0.9 \ 0.01]',$$

show that the steady state values of the model are  $\bar{y} = 0.90$ ,  $\bar{c} = 0.70$ ,  $\bar{i} = 0.21$ ,  $\bar{n} = 0.47$ ,  $\bar{l} = 0.53$ , and  $\bar{k} = 3.49$  (and take as granted  $\bar{z} = 1$ ). Next, use a numerical gradient procedure to derive

$$A = \begin{bmatrix} 0 & 1 & 0 & 0.33 & -1 & -0.33 & -1 \\ 0 & 1.5 & 0 & -0.06 & 0.5 & 0.06 & -0.08 \\ 1 & 0 & 0 & -0.67 & 0 & -0.33 & -1 \\ 1 & -0.77 & -0.23 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -0.47 & -0.53 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix},$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0.33 & 0 \\ 0 & 1.5 & 0 & 0 & 0.5 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.06 & 0 & 0 & 0.94 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.9 \end{bmatrix}.$$

### Exercise 3.3

Rederive the matrices  $A$  and  $B$  given the explicit incorporation of growth in the model. That is, derive  $A$  and  $B$  using the steady state expressions obtained in Exercise 3.1, and using (3.25) and (3.26) in place of (3.19) and (3.16).

## 3.2 Model II: Monopolistic Competition and Monetary Policy

This section outlines a model of imperfect competition featuring “sticky” prices. The model includes three sources of aggregate uncertainty: shocks to demand, technology, and the competitive structure of the economy. The model is due to Ireland (2004a), who designed it to determine how the apparent role of technology shocks in driving business cycle fluctuations is influenced by the inclusion of these additional sources of uncertainty.

From a pedagogical perspective, the model differs in two interesting ways relative to the RBC model outlined above. Whereas the linearized RBC model is a first-order system of difference equations, the linearized version of this model is a second-order system. However, recall that it is possible to represent a system of arbitrary order using the first-order form

$$Ax_{t+1} = Bx_t + Cv_{t+1} + D\eta_{t+1},$$

given appropriate specification of the elements of  $x_t$ . Second, the problem of mapping implications carried by a stationary model into the behavior of nonstationary data is revisited from a perspective different from that adopted in the discussion of the RBC model. Specifically, rather than assuming the actual data follow stationary deviations around deterministic trends, here the data are modeled as following drifting random walks; stationarity is induced via differencing rather than detrending.

### 3.2.1 Environment

The economy once again consists of a continuum of identical households. Here, there are two distinct production sectors: an intermediate-goods sector and a final-goods sector. The former is imperfectly competitive: it consists of a continuum of firms that produce differentiated products that serve as factors of production in the final-goods sector. Although firms in this sector have the ability to set prices, they face a friction in doing so. Finally, there is a central bank.

## 3.2 Model II: Monopolistic Competition and Monetary Policy

### HOUSEHOLDS

The representative household maximizes lifetime utility defined over consumption, money holdings, and labor:

$$\max_{c_t, m_t, n_t} \quad U = E_0 \sum_{t=0}^{\infty} \beta^t \left\{ a_t \ln c_t + \ln \frac{m_t}{p_t} - \frac{n_t^\xi}{\xi} \right\} \quad (3.34)$$

$$\text{s.t.} \quad p_t c_t + \frac{b_t}{r_t} + m_t = m_{t-1} + b_{t-1} + \tau_t + w_t n_t + d_t, \quad (3.35)$$

where  $\beta \in (0, 1)$  and  $\xi \geq 1$ . According to the budget constraint (3.35), the household divides its wealth between holdings of bonds  $b_t$  and money  $m_t$ ; bonds mature at the gross nominal rate  $r_t$  between time periods. The household also receives transfers  $\tau_t$  from the monetary authority and works  $n_t$  hours in order to earn wages  $w_t$  to finance its expenditures. Finally, the household owns an intermediate-goods firm, from which it receives a dividend payment  $d_t$ . Note from (3.34) that the household is subject to an exogenous demand shock  $a_t$  that affects its consumption decision.

Recognizing that the instantaneous marginal utility derived from consumption is given by  $\frac{a_t}{c_t}$ , the first-order conditions associated with the household's choices of labor, bond holdings, and money holdings are given by

$$\left( \frac{w_t}{p_t} \right) \left( \frac{a_t}{c_t} \right) = n_t^{\xi-1} \quad (3.36)$$

$$\beta E_t \left\{ \left( \frac{1}{p_{t+1}} \right) \left( \frac{a_{t+1}}{c_{t+1}} \right) \right\} = \left( \frac{1}{r_t p_t} \right) \left( \frac{a_t}{c_t} \right) \quad (3.37)$$

$$\left( \frac{m_t}{p_t} \right)^{-1} + \beta E_t \left\{ \left( \frac{1}{p_{t+1}} \right) \left( \frac{a_{t+1}}{c_{t+1}} \right) \right\} = \left( \frac{1}{p_t} \right) \left( \frac{a_t}{c_t} \right). \quad (3.38)$$

### Exercise 3.4

Interpret how (3.36)–(3.38) represent the optimal balancing of trade-offs associated with the household's choices of  $n$ ,  $b$ , and  $m$ .

### FIRMS

There are two types of firms: one produces a final consumption good  $y_t$ , which sells at price  $p_t$ ; the other is a continuum of intermediate-goods firms that supply inputs to the final-good firm. The output of the  $i^{\text{th}}$  intermediate good is given by  $y_{it}$ , which sells at price  $p_{it}$ . The intermediate goods combine to produce the final good via a constant elasticity of substitution

(CES) production function. The final-good firm operates in a competitive environment and pursues the following objective:

$$\max_{y_{it}} \Pi_t^F = p_t y_t - \int_0^1 p_{it} y_{it} di \quad (3.39)$$

$$s.t. \quad y_t = \left\{ \int_0^1 y_{it}^{\frac{\theta_t-1}{\theta_t}} di \right\}^{\frac{\theta_t}{\theta_t-1}}. \quad (3.40)$$

The solution to this problem yields a standard demand for intermediate inputs and a price aggregator:

$$y_{it} = y_t \left\{ \frac{p_{it}}{p_t} \right\}^{-\theta_t} \quad (3.41)$$

$$p_t = \left\{ \int_0^1 p_{it}^{1-\theta_t} di \right\}^{\frac{1}{1-\theta_t}}. \quad (3.42)$$

Notice that  $\theta_t$  is the markup of price above marginal cost; randomness in  $\theta_t$  provides the notion of a cost-push shock in this environment.

Intermediate-goods firms are monopolistically competitive. Because the output of each firm enters the final-good production function symmetrically, the focus is on a representative firm. The firm is owned by the representative household; thus its objectives are aligned with the household's. It manipulates the sales price of its good in pursuit of these objectives, subject to a quadratic adjustment cost:

$$\max_{p_{it}} \Pi_{it}^I = E_0 \sum_{t=0}^{\infty} \beta^t \left( \frac{a_t}{c_t} \right) \left( \frac{d_t}{p_t} \right), \quad (3.43)$$

$$s.t. \quad y_{it} = z_t n_{it} \quad (3.44)$$

$$y_{it} = y_t \left\{ \frac{p_{it}}{p_t} \right\}^{-\theta_t} \quad (3.45)$$

$$\chi(p_{it}, p_{it-1}) = \frac{\phi}{2} \left[ \frac{p_{it}}{\bar{\pi} p_{it-1}} - 1 \right]^2 y_t, \quad \phi > 0, \quad (3.46)$$

where  $\bar{\pi}$  is the gross inflation rate targeted by the monetary authority (described below), and the real value of dividends in (3.43) is given by

$$\frac{d_t}{p_t} = \left\{ \frac{p_{it} y_{it} - w_t n_{it}}{p_t} - \chi(p_{it}, p_{it-1}) \right\}. \quad (3.47)$$

The associated first-order condition can be written as

$$\begin{aligned} & (\theta_t - 1) \left( \frac{p_{it}}{p_t} \right)^{-\theta_t} \frac{y_t}{p_t} \\ &= \theta_t \left( \frac{p_{it}}{p_t} \right)^{-\theta_t-1} \frac{w_t y_t}{p_t z_t p_t} - \left\{ \phi \left[ \frac{p_{it}}{\bar{\pi} p_{it-1}} - 1 \right] \frac{y_t}{\bar{\pi} p_{it-1}} - \beta \phi \right. \\ & \quad \times E_t \left. \left( \frac{a_{t+1}}{a_t} \frac{c_t}{c_{t+1}} \left( \frac{p_{it+1}}{\bar{\pi} p_{it}} - 1 \right) \frac{y_{t+1} p_{it+1}}{\bar{\pi} p_{it}^2} \right) \right\}. \end{aligned} \quad (3.48)$$

The left-hand side of (3.48) reflects the marginal revenue to the firm generated by an increase in price; the right-hand side reflects associated marginal costs. Under perfect price flexibility ( $\phi = 0$ ) there is no dynamic component to the firm's problem; the price-setting rule reduces to

$$p_{it} = \frac{\theta_t}{\theta_t - 1} \frac{w_t}{z_t},$$

which is a standard markup of price over marginal cost  $\frac{w_t}{z_t}$ . Under "sticky prices" ( $\phi > 0$ ) the marginal cost of an increase in price has two additional components: the direct cost of a price adjustment, and an expected discounted cost of a price change adjusted by the marginal utility to the households of conducting such a change. Empirically, the estimation of  $\phi$  is of particular interest: this parameter plays a central role in distinguishing this model from its counterparts in the RBC literature.

### THE MONETARY AUTHORITY

The monetary authority chooses the nominal interest rate according to a Taylor Rule. With all variables expressed in terms of logged deviations from steady state values, the rule is given by

$$\tilde{r}_t = \rho_r \tilde{r}_{t-1} + \rho_\pi \tilde{\pi}_t + \rho_g \tilde{g}_t + \rho_o \tilde{o}_t + \varepsilon_{rt}, \quad \varepsilon_{rt} \sim iid N(0, \sigma_r^2), \quad (3.49)$$

where  $\tilde{\pi}_t$  is the gross inflation rate,  $\tilde{g}_t$  is the gross growth rate of output, and  $\tilde{o}_t$  is the output gap (defined below). The  $\rho_i$  parameters denote elasticities. The inclusion of  $\tilde{r}_{t-1}$  as an input into the Taylor Rule allows for the gradual adjustment of policy to demand and technology shocks, for example, as in Clarida, Gali, and Gertler (2000).

The output gap is the logarithm of the ratio of actual output  $y_t$  to capacity output  $\hat{y}_t$ . Capacity output is defined to be the "efficient" level of

output, which is equivalent to the level of output chosen by a benevolent social planner who solves:

$$\max_{\hat{y}_t, n_{it}} U^S = E_0 \sum_0^\infty \beta^t \left\{ a_t \ln \hat{y}_t - \frac{1}{\xi} \left( \int_0^1 n_{it} di \right)^\xi \right\} \quad (3.50)$$

$$s.t. \quad \hat{y}_t = z_t \left( \int_0^1 n_{it}^{\frac{\theta_t-1}{\theta_t}} di \right)^{\frac{\theta_t}{\theta_t-1}} \quad (3.51)$$

The symmetric solution to this problem is simply

$$\hat{y}_t = a_t^{\frac{1}{\xi}} z_t. \quad (3.52)$$

### STOCHASTIC SPECIFICATION

In addition to the monetary policy shock  $\varepsilon_{rt}$  introduced in (3.49), the model features a demand shock  $a_t$ , a technology shock  $z_t$ , and a cost-push shock  $\theta_t$ . The former is *iid*; the latter three evolve according to

$$\ln(a_t) = (1 - \rho_a) \ln(\bar{a}) + \rho_a \ln(a_{t-1}) + \varepsilon_{at}, \quad \bar{a} > 1 \quad (3.53)$$

$$\ln(z_t) = \ln(\bar{z}) + \ln(z_{t-1}) + \varepsilon_{zt}, \quad \bar{z} > 1 \quad (3.54)$$

$$\ln(\theta_t) = (1 - \rho_\theta) \ln(\bar{\theta}) + \rho_\theta \ln(\theta_{t-1}) + \varepsilon_{\theta t}, \quad \bar{\theta} > 1, \quad (3.55)$$

with  $|\rho_i| < 1$ ,  $i = a, \theta$ . Note that the technology shock is nonstationary: it evolves as a drifting random walk. This induces similar behavior in the model's endogenous variables, and necessitates the use of an alternative to the detrending method discussed above in the context of the RBC model. Here, stationarity is induced by normalizing model variables by  $z_t$ . For the corresponding observable variables, stationarity is induced by differencing rather than detrending: the observables are measured as deviations of growth rates (logged differences of levels) from sample averages. Details are provided in the linearization step discussed below.

The model is closed through two additional steps. The first is the imposition of symmetry among the intermediate-goods firms. Given that the number of firms is normalized to one, symmetry implies:

$$y_{it} = y_t, \quad n_{it} = n_t, \quad p_{it} = p_t, \quad d_{it} = d_t. \quad (3.56)$$

The second is the requirement that the money and bond markets clear:

$$m_t = m_{t-1} + \tau_t \quad (3.57)$$

$$b_t = b_{t-1} = 0. \quad (3.58)$$

### 3.2.2 The Nonlinear System

In its current form, the model consists of 12 equations: the household's first-order conditions and budget constraint, the aggregate production function, the aggregate real dividends paid to the household by its intermediate-goods firm, the intermediate-goods firm's first-order condition, the stochastic specifications for the structural shocks, and the expression for capacity output. Following Ireland's (2004a) empirical implementation the focus is on a linearized reduction to an eight-equation system consisting of an IS curve, a Phillips curve, the Taylor Rule (specified in linearized form in (3.49)), the three exogenous shock specifications, and definitions for the growth rate of output and the output gap.

The reduced system is recast in terms of the following normalized variables:

$$\begin{aligned} \ddot{y}_t &= \frac{y_t}{z_t}, & \ddot{c}_t &= \frac{c_t}{z_t}, & \ddot{\hat{y}}_t &= \frac{\hat{y}_t}{z_t}, & \pi_t &= \frac{p_t}{p_{t-1}}, \\ \ddot{d}_t &= \frac{(d_t/p_t)}{z_t}, & \ddot{w}_t &= \frac{(w_t/p_t)}{z_t}, & \ddot{m}_t &= \frac{(m_t/p_t)}{z_t}, & \ddot{z}_t &= \frac{z_t}{z_{t-1}}. \end{aligned}$$

Using the expression for real dividends given by (3.47), the household's budget constraint in equilibrium is rewritten as

$$\ddot{y}_t = \ddot{c}_t + \frac{\phi}{2} \left( \frac{\pi_t}{\bar{\pi}} - 1 \right)^2 \ddot{y}_t. \quad (3.59)$$

Next, the household's first-order condition (3.37) is written in normalized form as

$$\frac{a_t}{\ddot{c}_t} = \beta r_t E_t \left\{ \frac{a_{t+1}}{\ddot{c}_{t+1}} \times \frac{1}{\ddot{z}_{t+1}} \times \frac{1}{\pi_{t+1}} \right\}. \quad (3.60)$$

Next, the household's remaining first-order conditions, the expression for the real dividend payment (3.47) it receives, and the aggregate production function can be combined to eliminate wages, money, labor, dividends, and capacity output from the system. Having done this, we then introduce the expression for the output gap into the system:

$$o_t \equiv \frac{y_t}{\ddot{y}_t} = \frac{\ddot{y}_t}{a_t^{\frac{1}{\xi}}}. \quad (3.61)$$

Finally, normalizing the first-order condition of the intermediate-goods firm and the stochastic specifications leads to the following nonlinear system:

$$\ddot{y}_t = \ddot{c}_t + \frac{\phi}{2} \left( \frac{\pi_t}{\bar{\pi}} - 1 \right)^2 \ddot{y}_t \quad (3.62)$$

$$\frac{a_t}{\tilde{c}_t} = \beta r_t E_t \left\{ \frac{a_{t+1}}{\tilde{z}_{t+1}} \times \frac{1}{\tilde{y}_{t+1}} \times \frac{1}{\pi_{t+1}} \right\} \quad (3.63)$$

$$0 = 1 - \theta_t + \theta_t \frac{\tilde{c}_t}{a_t} \tilde{y}_t^{\xi-1} - \phi \left( \frac{\pi_t}{\bar{\pi}} - 1 \right) \frac{\pi_t}{\bar{\pi}} + \beta \phi E_t \left\{ \frac{\tilde{c}_t a_{t+1}}{\tilde{z}_{t+1} a_t} \left( \frac{\pi_{t+1}}{\bar{\pi}} - 1 \right) \frac{\pi_{t+1}}{\bar{\pi}} \frac{\tilde{y}_{t+1}}{\tilde{y}_t} \right\} \quad (3.64)$$

$$g_t = \frac{\tilde{z}_t \tilde{y}_t}{\tilde{y}_{t-1}} \quad (3.65)$$

$$\sigma_t = \frac{y_t}{\tilde{y}_t} = \frac{\tilde{y}_t}{a_t^{\xi}} \quad (3.66)$$

$$\ln(a_t) = (1 - \rho_a) \ln(\bar{a}) + \rho_a \ln(a_{t-1}) + \varepsilon_{at} \quad (3.67)$$

$$\ln(\theta_t) = (1 - \rho_\theta) \ln(\bar{\theta}) + \rho_\theta \ln(\theta_{t-1}) + \varepsilon_{\theta t} \quad (3.68)$$

$$\ln(\tilde{z}_t) = \ln(\bar{z}) + \varepsilon_{zt}. \quad (3.69)$$

Along with the Taylor Rule, this is the system to be log-linearized.

### 3.2.3 Log-Linearization

Log-linearization proceeds with the calculation of steady state values of the endogenous variables:

$$\begin{aligned} \bar{r} &= \frac{\bar{z}}{\beta}, \\ \bar{c} &= \bar{y} = \left( \bar{a} \frac{\bar{\theta} - 1}{\bar{\theta}} \right)^{\frac{1}{\xi}}, \\ \bar{\sigma} &= \left( \bar{\theta} - 1 \right)^{\frac{1}{\xi}}; \end{aligned} \quad (3.70)$$

(3.62)–(3.69) are then log-linearized around these values. As with the RBC model, this can be accomplished through the use of a numerical gradient procedure. However, as an alternative to this approach, here we follow Ireland (2004a) and demonstrate the use of a more analytically oriented procedure. In the process, it helps to be mindful of the reconfiguration Ireland worked with: an IS curve, a Phillips curve, the Taylor Rule, the shock processes, and definitions of the growth rate of output and the output gap.

### 3.2 Model II: Monopolistic Competition and Monetary Policy

As a first step, the variables appearing in (3.62)–(3.69) are written in logged form. Log-linearization of (3.62) then yields

$$\tilde{y}_t \equiv \ln \left( \frac{\tilde{y}_t}{\bar{y}} \right) = \tilde{c}_t,$$

because the partial derivative of  $\tilde{y}_t$  with respect to  $\tilde{z}_t$  (evaluated at steady state) is zero. (Recall our notational convention: tildes denote logged deviations of variables from steady state values.) Hence upon linearization, this equation is eliminated from the system, and  $\tilde{c}_t$  is replaced by  $\tilde{y}_t$  in the remaining equations.

Next, recalling that  $E_t(\tilde{z}_{t+1}) = 0$ , log-linearization of (3.63) yields

$$0 = \tilde{r}_t - E_t \tilde{z}_{t+1} - (E_t \tilde{y}_{t+1} - \tilde{y}_t) + E_t \tilde{\theta}_{t+1} - \tilde{\theta}_t. \quad (3.71)$$

Relating output and the output gap via the log-linearization of (3.66),

$$\tilde{y}_t = \frac{1}{\xi} \tilde{a}_t + \tilde{\sigma}_t, \quad (3.72)$$

the term  $E_t(\tilde{y}_{t+1}) - \tilde{y}_t$  can be substituted out of (3.71), yielding the IS curve:

$$\tilde{\sigma}_t = E_t \tilde{\sigma}_{t+1} - (\tilde{r}_t - E_t \tilde{z}_{t+1}) + (1 - \xi^{-1})(1 - \rho_a)\tilde{a}_t. \quad (3.73)$$

Similarly, log-linearizing (3.64) and eliminating  $\tilde{y}_t$  using (3.72) yields the Phillips curve:

$$\tilde{\pi}_t = \beta E_t \tilde{\pi}_{t+1} + \psi \tilde{\sigma}_t - \tilde{e}_t, \quad (3.74)$$

where  $\psi = \frac{\xi(\theta-1)}{\phi}$  and  $\tilde{e}_t = \frac{1}{\phi} \tilde{\theta}_t$ . This latter equality is a normalization of the cost-push shock; like the cost-push shock itself, the normalized shock follows an AR(1) process with persistence parameter  $\rho_\theta = \rho_e$ , and innovation standard deviation  $\sigma_e = \frac{1}{\phi} \sigma_\theta$ .

The resulting IS and Phillips curves are forward looking: they include the one-step-ahead expectations operator. However, prior to empirical implementation, Ireland augmented these equations to include lagged variables of the output gap and inflation in order to enhance the empirical coherence of the model. This final step yields the system he analyzed. Dropping time subscripts and denoting, for example,  $\tilde{\sigma}_{t-1}$  as  $\tilde{\sigma}^-$ , the system is given by

$$\tilde{\sigma} = \alpha_o \tilde{\sigma}^- + (1 - \alpha_o) E_t \tilde{\sigma}' - (\tilde{r} - E_t \tilde{\pi}') + (1 - \xi^{-1})(1 - \rho_a)\tilde{a} \quad (3.75)$$

$$\tilde{\pi} = \beta \alpha_\pi \tilde{\pi}^- + \beta (1 - \alpha_\pi) E_t \tilde{\pi}' + \psi \tilde{\sigma} - \tilde{e} \quad (3.76)$$

$$\tilde{g}' = \tilde{y}' - \tilde{y} + \tilde{z}' \quad (3.77)$$

$$\tilde{o}' = \tilde{y}' - \xi^{-1} \tilde{a}' \quad (3.78)$$

$$\tilde{r}' = \rho_r \tilde{r} + \rho_\pi \tilde{\pi}' + \rho_g \tilde{g}' + \rho_a \tilde{o}' + \varepsilon'_r \quad (3.79)$$

$$\tilde{a}' = \rho_a \tilde{a} + \varepsilon'_a \quad (3.80)$$

$$\tilde{e}' = \rho_e \tilde{e} + \varepsilon'_e \quad (3.81)$$

$$\tilde{z}' = \varepsilon'_z \quad (3.82)$$

where the structural shocks

$$v_t = \{\varepsilon_{rt}, \varepsilon_{at}, \varepsilon_{et}, \varepsilon_{zt}\}$$

are  $iidN$  with diagonal covariance matrix  $\Sigma$ . The additional parameters introduced are  $\alpha_0 \in [0, 1]$  and  $\alpha_\pi \in [0, 1]$ ; setting  $\alpha_0 = \alpha_\pi = 0$  yields the original microfoundations.

The augmentation of the IS and Phillips curves with lagged values of the output gap and inflation converts the model from a first- to a second-order system. Thus a final step is required in mapping this system into the first-order specification

$$Ax_{t+1} = Bx_t + Cv_{t+1} + D\eta_{t+1}.$$

This is accomplished by augmenting the vector  $x_t$  to include not only contemporaneous observations of the variables of the system, but also lagged values of the output gap and inflation:

$$x_t = [\tilde{o}_t \ \tilde{o}_{t-1} \ \tilde{\pi}_t \ \tilde{\pi}_{t-1} \ \tilde{y}_t \ \tilde{r}_t \ \tilde{g}_t \ \tilde{a}_t \ \tilde{e}_t \ \tilde{z}_t]'$$

This also requires the introduction of two additional equations into the system:  $\tilde{\pi}' = \tilde{\pi}'$  and  $\tilde{o}' = \tilde{o}'$ . Specifying these as the final two equations of the system, the corresponding matrices  $A$  and  $B$  are given by

$$A = \begin{bmatrix} -(1-\alpha_0) & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & -\psi & -\beta(1-\alpha_\pi) & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & -1 \\ 1 & 0 & 0 & 0 & -1 & 0 & 0 & \xi^{-1} & 0 & 0 \\ -\rho_0 & 0 & -\rho_\pi & 0 & 0 & 1 & -\rho_g & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \quad (3.83)$$

$$B = \begin{bmatrix} 0 & \alpha_0 & 0 & 0 & 0 & -1 & 0 & (1-\xi^{-1})(1-\rho_a) & 0 & 0 \\ 0 & 0 & 0 & \beta\alpha_\pi & 0 & 0 & 0 & 0 & -1 & 0 \\ 0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & \rho_r & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho_a & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho_e & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \end{bmatrix}. \quad (3.84)$$

Further, defining  $\eta_t = [\eta_{1t} \ \eta_{2t}]'$ , where

$$\eta_{1t+1} = E_t \tilde{o}_{t+1} - \tilde{o}_{t+1}$$

and

$$\eta_{2t+1} = E_t \tilde{\pi}_{t+1} - \tilde{\pi}_{t+1},$$

the matrices  $C$  and  $D$  are given by

$$C = \begin{bmatrix} 0_{4 \times 4} \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad D = \begin{bmatrix} -(1-\alpha_0) & -1 \\ 0 & -\beta(1-\alpha_\pi) \\ 0_{2 \times 8} \\ 0_{2 \times 4} \end{bmatrix}. \quad (3.85)$$

The final step needed for empirical implementation is to identify the observable variables of the system. For Ireland, these are the gross growth rate of output  $g_t$ , the gross inflation rate  $\pi_t$ , and the nominal interest rate  $r_t$  (all measured as logged ratios of sample averages). Under the assumption that output and aggregate prices follow drifting random walks,  $g_t$  and  $\pi_t$  are stationary; the additional assumption of stationarity for  $r_t$  is all that is necessary to proceed with the analysis.

### Exercise 3.5

Consider the following CRRA form for the instantaneous utility function for model II:

$$u(c_t, \frac{m_t}{n}, n_t) = a_t \frac{c_t^\lambda}{\lambda} + \log \frac{m_t}{n} - \frac{n_t^\xi}{\varepsilon}.$$

1. Derive the nonlinear system under this specification.
2. Sketch the linearization of the system via a numerical gradient procedure.

### 3.3 Model III: Asset Pricing

The final model is an adaptation of Lucas' (1978) one-tree model of asset-pricing behavior. Alternative versions of the model have played a prominent role in two important strands of the empirical finance literature. The first, launched by LeRoy and Porter (1981) and Shiller (1981) in the context of single-asset versions of the model, concerns the puzzling degree of volatility exhibited by prices associated with aggregate stock indexes. The second, launched by Mehra and Prescott (1985) in the context of a multi-asset version of the model, and subsequently underscored by Weil (1989), concerns the puzzling coincidence of a large gap observed between the returns of risky and risk-free assets, and a low average risk-free return. Resolutions to both puzzles have been investigated using alternate preference specifications. After outlining single- and multi-asset versions of the model given a generic specification of preferences, we introduce alternative functional forms. Overviews of the role of preferences in the equity-premium literature are provided by Kocherlakota (1996); Campbell, Lo, and MacKinlay (1997); and Cochrane (2001); and in the stock-price volatility literature by Shiller (1989) and DeJong and Ripoll (2004).

#### 3.3.1 Single-Asset Environment

The model features a continuum of identical households and a single risky asset. Shares held during period  $(t - 1)$ ,  $s_{t-1}$ , yield a dividend payment  $d_t$  at time  $t$ ; time- $t$  share prices are  $p_t$ . Households maximize expected lifetime utility by financing consumption  $c_t$  from an exogenous stochastic dividend stream, proceeds from sales of shares, and an exogenous stochastic endowment  $q_t$ . The utility maximization problem of the representative household is given by

$$\max_{c_t} \quad U = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t), \quad (3.86)$$

where  $\beta \in (0, 1)$  again denotes the discount factor, and optimization is subject to

$$c_t + p_t(s_t - s_{t-1}) = d_t s_{t-1} + q_t. \quad (3.87)$$

#### 3.3 Model III: Asset Pricing

Because households are identical, equilibrium requires  $s_t = s_{t-1}$  for all  $t$ , and thus

$$c_t = d_t s_t + q_t = d_t + q_t$$

(hereafter,  $s_t$  is normalized to 1). Combining this equilibrium condition with the household's necessary condition for a maximum yields the pricing equation

$$p_t = \beta E_t \left[ \frac{u'(d_{t+1} + q_{t+1})}{u'(d_t + q_t)} (d_{t+1} + p_{t+1}) \right]. \quad (3.88)$$

From (3.88), following a shock to either  $d_t$  or  $q_t$ , the response of  $p_t$  depends in part upon the variation of the marginal rate of substitution between  $t$  and  $t + 1$ . This in turn depends upon the instantaneous utility function  $u(\cdot)$ . The puzzle identified by LeRoy and Porter (1981) and Shiller (1981) is that  $p_t$  is far more volatile than what (3.88) would imply, given the observed volatility of  $d_t$ .

The model is closed by specifying stochastic processes for  $(d_t, q_t)$ . These are given by

$$\ln d_t = (1 - \rho_d) \ln (\bar{d}) + \rho_d \ln (d_{t-1}) + \varepsilon_{dt} \quad (3.89)$$

$$\ln q_t = (1 - \rho_q) \ln (\bar{q}) + \rho_q \ln (q_{t-1}) + \varepsilon_{qt}, \quad (3.90)$$

with  $|\rho_i| < 1$ ,  $i = d, q$ , and

$$\begin{bmatrix} \varepsilon_{dt} \\ \varepsilon_{qt} \end{bmatrix} \sim iidN(0, \Sigma). \quad (3.91)$$

#### 3.3.2 Multi-Asset Environment

An  $n$ -asset extension of the environment leaves the household's objective function intact, but modifies its budget constraint to incorporate the potential for holding  $n$  assets. As a special case, Mehra and Prescott (1985) studied a two-asset specification, including a risk-free asset (ownership of government bonds) and risky asset (ownership of equity). In this case, the household's budget constraint is given by

$$c_t + p_t^e(s_t^e - s_{t-1}^e) + p_t^f s_t^f = d_t s_{t-1}^e + s_{t-1}^f + q_t, \quad (3.92)$$

where  $p_t^e$  denotes the price of the risky asset,  $s_t^e$  represents the number of shares held in the asset during period  $t - 1$ , and  $p_t^f$  and  $s_t^f$  are analogous for the risk-free asset. The risk-free asset pays one unit of the consumption

good at time  $t$  if held at time  $t - 1$  (hence the loading factor of 1 associated with  $s_{t-1}^f$  on the right-hand side of the budget constraint).

First-order conditions associated with the choice of the assets are analogous to the pricing equation (3.88) established in the single-asset specification. Rearranging slightly:

$$\beta E_t \left\{ \frac{u'(c_{t+1})}{u'(c_t)} \times \frac{p_{t+1}^e + d_t}{p_t^e} \right\} = 1 \quad (3.93)$$

$$\beta E_t \left\{ \frac{u'(c_{t+1})}{u'(c_t)} \times \frac{1}{p_t^f} \right\} = 1. \quad (3.94)$$

Defining gross returns associated with the assets as

$$r_{t+1}^e = \frac{p_{t+1}^e + d_t}{p_t^e}$$

$$r_{t+1}^f = \frac{1}{p_t^f},$$

Mehra and Prescott's identification of the equity premium puzzle centers on

$$\beta E_t \left\{ \frac{u'(c_{t+1})}{u'(c_t)} r_{t+1}^f \right\} = 1 \quad (3.95)$$

$$E_t \left\{ \frac{u'(c_{t+1})}{u'(c_t)} [r_{t+1}^e - r_{t+1}^f] \right\} = 0, \quad (3.96)$$

where (3.96) is derived by subtracting (3.94) from (3.93). The equity premium puzzle has two components. First, taking  $\{c_t\}$  as given, the average value of  $r^e - r^f$  is quite large: given CRRA preferences, implausibly large values of the risk-aversion parameter are needed to account for the average difference observed in returns. Second, given a specification of  $u(c)$  that accounts for (3.96), and again taking  $\{c_t\}$  as given, the average value observed for  $r^f$  is far too low to reconcile with (3.95). This second component, as emphasized by Weil (1989), is the risk-free rate puzzle.

### 3.3.3 Alternative Preference Specifications

As noted, alternative preference specifications have been considered for their potential in resolving both puzzles. Here, in the context of the single-asset environment, three forms for the instantaneous utility function are presented in anticipation of the empirical applications to be presented in Part II: CRRA preferences, habit/durability preferences, and self-control preferences. The presentation follows that of DeJong and Ripoll

(2004), who sought to evaluate empirically the ability of these preference specifications to make headway in resolving the stock-price volatility puzzle.

#### CRRA

Once again, CRRA preferences are parameterized as

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}, \quad (3.97)$$

thus  $\gamma > 0$  measures the degree of relative risk aversion, and  $1/\gamma$  the intertemporal elasticity of substitution. The equilibrium pricing equation is given by

$$p_t = \beta E_t \left[ \frac{(d_{t+1} + q_{t+1})^{-\gamma}}{(d_t + q_t)^{-\gamma}} (d_{t+1} + p_{t+1}) \right]. \quad (3.98)$$

Notice that, *ceteris paribus*, a relatively large value of  $\gamma$  will increase the volatility of price responses to exogenous shocks, at the cost of decreasing the correlation between  $p_t$  and  $d_t$  (due to the heightened role assigned to  $q_t$  in driving price fluctuations). Because  $\{d_t\}$  and  $\{q_t\}$  are exogenous, their steady states  $\bar{d}$  and  $\bar{q}$  are simply parameters. Normalizing  $\bar{d}$  to 1 and defining  $\eta = \frac{\bar{q}}{\bar{d}}$ , so that  $\eta = \bar{q}$ , the steady state value of consumption (derived from the budget constraint) is  $\bar{c} = 1 + \eta$ . And from the pricing equation,

$$\begin{aligned} \bar{p} &= \frac{\beta}{1-\beta} \bar{d} \\ &= \frac{\beta}{1-\beta}. \end{aligned} \quad (3.99)$$

Letting  $\beta = 1/(1+r)$ , where  $r$  denotes the household's discount rate, (3.99) implies  $\bar{p}/\bar{d} = 1/r$ . Thus as the household's discount rate increases, its asset demand decreases, driving down the steady state price level. Empirically, the average price/dividend ratio observed in the data serves to pin down  $\beta$  under this specification of preferences.

#### Exercise 3.6

*Log-linearize the pricing equation (3.98) around the model's steady state values.*

#### HABIT/DURABILITY

Following Persion and Constantinides (1991) and Heaton (1995), an alternative specification of preferences that introduces habit and durability into the specification of preferences is parameterized as

$$u(b_t) = \frac{b_t^{1-\gamma}}{1-\gamma}, \quad (3.100)$$

with

$$b_t = b_t^d - \alpha b_t^h, \quad (3.101)$$

where  $\alpha \in (0, 1)$ ,  $b_t^d$  is the household's durability stock, and  $b_t^h$  its habit stock. The stocks are defined by

$$b_t^d = \sum_{j=0}^{\infty} \delta^j c_{t-j} \quad (3.102)$$

$$\begin{aligned} b_t^h &= (1-\theta) \sum_{j=0}^{\infty} \theta^j b_{t-1-j}^d \\ &= (1-\theta) \sum_{j=0}^{\infty} \theta^j \sum_{i=0}^{\infty} \delta^i c_{t-1-i}, \end{aligned} \quad (3.103)$$

where  $\delta \in (0, 1)$  and  $\theta \in (0, 1)$ . Thus the durability stock represents the flow of services from past consumption, which depreciates at rate  $\delta$ . This parameter also represents the degree of intertemporal substitutability of consumption. The habit stock can be interpreted as a weighted average of the durability stock, where the weights sum to one. Notice that more recent durability stocks, or more recent flows of consumption, are weighted relatively heavily; thus the presence of habit captures intertemporal complementarity. The variable  $b_t$  represents the current level of durable services net of the average of past services; the parameter  $\alpha$  measures the fraction of the average of past services that is netted out. Notice that if  $\delta = 0$ , there would be only habit persistence, whereas if  $\alpha = 0$ , only durability survives. Finally, when  $\theta = 0$ , the habit stock includes only one lag. Thus estimates of these parameters are of particular interest empirically.

Using the definitions of durability and habit stocks,  $b_t$  becomes

$$\begin{aligned} b_t &= c_t + \sum_{j=1}^{\infty} \left[ \delta^j - \alpha(1-\theta) \sum_{i=0}^{j-1} \delta^i \theta^{j-i-1} \right] c_{t-j} \\ &\equiv \sum_{j=0}^{\infty} \Phi_j c_{t-j}, \end{aligned} \quad (3.104)$$

### 3.3 Model III: Asset Pricing

where  $\Phi_0 \equiv 1$ . Thus for these preferences, the pricing equation is given by

$$p_t = \beta E_t \frac{\sum_{j=0}^{\infty} \beta^j \Phi_j (\sum_{i=0}^{\infty} \Phi_i c_{t+1+j-i})^{-\gamma}}{\sum_{j=0}^{\infty} \beta^j \Phi_j (\sum_{i=0}^{\infty} \Phi_i c_{t+j-i})^{-\gamma}} (d_{t+1} + p_{t+1}), \quad (3.105)$$

where as before  $c_t = d_t + q_t$  in equilibrium.

To see how the presence of habit and durability can potentially influence the volatility of the prices, rewrite the pricing equation as

$$\begin{aligned} p_t &= \beta E_t \frac{(c_{t+1} + \Phi_1 c_t + \Phi_2 c_{t-1} + \dots)^{-\gamma}}{(c_t + \Phi_1 c_{t-1} + \Phi_2 c_{t-2} + \dots)^{-\gamma}} \\ &\quad + \beta \Phi_1 (c_{t+2} + \Phi_1 c_{t+1} + \Phi_2 c_t + \dots)^{-\gamma} + \dots (d_{t+1} + p_{t+1}). \\ &\quad + \beta \Phi_1 (c_{t+1} + \Phi_1 c_t + \Phi_2 c_{t-1} + \dots)^{-\gamma} + \dots \end{aligned} \quad (3.106)$$

When there is a positive shock to, say,  $q_t$ ,  $c_t$  increases by the amount of the shock, say  $\sigma_q$ . Given (3.89)–(3.90),  $c_{t+1}$  would increase by  $\rho_q \sigma_q$ ,  $c_{t+2}$  would increase by  $\rho_q^2 \sigma_q$ , and so on. Now, examine the first term in parentheses in both the numerator and the denominator. First, in the denominator  $c_t$  will grow by  $\sigma_q$ . Second, in the numerator  $c_{t+1} + \Phi_1 c_t$  goes up by  $(\rho_q + \Phi_1) \sigma_q \leq \sigma_q$ . Thus, whether the share price  $p_t$  increases by more than in the standard CRRA case depends ultimately on whether  $\rho_q + \Phi_1 \leq 1$ . Notice that if  $\Phi_j = 0$  for  $j > 0$ , (3.106) reduces to the standard CRRA utility case. If we had only habit and not durability, then  $\Phi_1 < 0$ , and thus the response of prices would be greater than in the CRRA case. This result is intuitive: habit captures intertemporal complementarity in consumption, which strengthens the smoothing motive relative to the time-separable CRRA case.

Alternatively, if there was only durability and not habit, then  $0 < \Phi_1 < 1$ , but one still would not know whether  $\rho + \Phi_1 \leq 1$ . Thus with only durability, we cannot judge how the volatility of  $p_t$  would be affected: this will depend upon the sizes of  $\rho$  and  $\Phi_1$ . Finally, we also face indeterminacy under a combination of both durability and habit: if  $\alpha$  is large and  $\delta$  is small enough to make  $\rho + \Phi_1 < 1$ , then we would get increased price volatility. Thus this issue is fundamentally quantitative. Finally, with respect to the steady state price, note from (3.106) that it is identical to the CRRA case.

### Exercise 3.7

Given that the pricing equation under Habit/Durability involves an infinite number of lags, truncate the lags to 3 and log-linearize the pricing equation (3.106) around its steady state.

## SELF-CONTROL PREFERENCES

Consider next a household that in every period faces a temptation to consume all of its wealth. Resisting this temptation imposes a self-control utility cost. To model these preferences we follow Gul and Pesendorfer (2004), who identified a class of dynamic self-control preferences. In this case, the problem of the household can be formulated recursively as

$$W(s, P) = \max_{s'} \{u(c) + v(c) + \beta E W(s', P')\} - \max_{\tilde{c}} v(\tilde{c}), \quad (3.107)$$

where  $P = (p, d, q)$ ,  $u(\cdot)$  and  $v(\cdot)$  are von Neumann–Morgenstern utility functions,  $\beta \in (0, 1)$ ,  $\tilde{c}$  represents temptation consumption, and  $s'$  denotes share holdings in the next period. Whereas  $u(\cdot)$  is the momentary utility function,  $v(\cdot)$  represents temptation. The problem is subject to the following budget constraints:

$$c = ds + q - p(s' - s) \quad (3.108)$$

$$\tilde{c} = ds + q - p(\tilde{s}' - s). \quad (3.109)$$

In (3.107),  $v(c) - \max_{\tilde{c}} v(\tilde{c}) \leq 0$  represents the disutility of self-control given that the agent has chosen  $c$ . With  $v(c)$  specified as strictly increasing, the solution for  $\max_{\tilde{c}} v(\tilde{c})$  is simply to drive  $\tilde{c}$  to the maximum allowed by the constraint

$$\tilde{c} = ds + q - p(\tilde{s}' - s),$$

which is attained by setting  $\tilde{s}' = 0$ . Thus the problem is written as

$$W(s, P) = \max_{s'} \{u(c) + v(c) + \beta E W(s', P')\} - v(ds + q + ps) \quad (3.110)$$

subject to

$$c = ds + q - p(s' - s). \quad (3.111)$$

The optimality condition reads

$$[u'(c) + v'(c)]p = \beta E W'(s', P'), \quad (3.112)$$

and since

$$W'(s, P) = [u'(c) + v'(c)](d + p) - v'(ds + q + ps)(d + p), \quad (3.113)$$

the optimality condition becomes

$$[u'(c) + v'(c)]p = \beta E [u'(c') + v'(c') - v'(d's' + q' + p's')](d' + p'). \quad (3.114)$$

## 3.3 Model III: Asset Pricing

Combining this expression with the equilibrium conditions  $s = s' = 1$  and  $c = d + q$  yields

$$p = \beta E \left\{ (d' + p') \left[ \frac{u'(d' + q') + v'(d' + q') - v'(d' + q' + p')}{u'(d + q) + v'(d + q)} \right] \right\}. \quad (3.115)$$

Notice that when  $v(\cdot) = 0$ , there is no temptation, and the pricing equation reduces to the standard case. Otherwise, the term  $u'(d' + q') + v'(d' + q') - v'(d' + q' + p')$  represents tomorrow's utility benefit from saving today. This corresponds to the standard marginal utility of wealth tomorrow  $u'(d' + q')$ , plus the term  $v'(d' + q') - v'(d' + q' + p')$ , which represents the derivative of the utility cost of self-control with respect to wealth.

DeJong and Ripoll (2004) assume the following functional forms for the momentary and temptation utility functions:

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma} \quad (3.116)$$

$$v(c) = \lambda \frac{c^\phi}{\phi}, \quad (3.117)$$

with  $\lambda > 0$ , which imply the following pricing equation:

$$p = \beta E \left\{ (d' + p') \left[ \frac{(d' + q')^{-\gamma} + \lambda(d' + q')^{\phi-1} - \lambda(d' + q' + p')^{\phi-1}}{(d + q)^{-\gamma} + \lambda(d + q)^{\phi-1}} \right] \right\}. \quad (3.118)$$

The concavity/convexity of  $v(\cdot)$  plays an important role in determining implications of this preference specification for the stock-price volatility issue. To understand why, rewrite (3.118) as

$$p = \beta E \left\{ (d' + p') \times \left[ \frac{\frac{(d' + q')^{-\gamma}}{(d + q)^{-\gamma}} + \lambda(d + q)^\gamma [(d' + q')^{\phi-1} - (d' + q' + p')^{\phi-1}]}{1 + \lambda(d + q)^{\phi-1+\gamma}} \right] \right\}. \quad (3.119)$$

Suppose  $\phi > 1$ , so that  $v(\cdot)$  is convex, and consider the impact on  $p$  of a positive endowment shock. This increases the denominator, while decreasing the term

$$\lambda(d + q)^\gamma [(d' + q')^{\phi-1} - (d' + q' + p')^{\phi-1}]$$

in the numerator. Both effects imply that relative to the CRRA case, in which  $\lambda = 0$ , this specification *reduces* price volatility in the face of an endowment shock, which is precisely the opposite of what one would like to achieve in seeking to resolve the stock-price volatility puzzle.

The mechanism behind this reduction in price volatility is as follows: a positive shock to  $d$  or  $q$  increases the household's wealth today, which has three effects. The first ("smoothing") captures the standard intertemporal motive: the household would like to increase saving, which drives up the share price. Second, there is a "temptation" effect: with more wealth today, the feasible budget set for the household increases, which represents more temptation to consume, and less willingness to save. This effect works opposite to the first, and reduces price volatility with respect to the standard case. Third, there is the "self-control" effect: due to the assumed convexity of  $v(\cdot)$ , marginal self-control costs also increase, which reinforces the second effect. As shown above, the last two effects dominate the first, and thus under convexity of  $v(\cdot)$  the volatility is reduced relative to the CRRA case.

In contrast, price volatility would not necessarily be reduced if  $v(\cdot)$  is concave, and thus  $0 < \phi < 1$ . In this case, when  $d$  or  $q$  increases, the term

$$\lambda(d + q)^\gamma [(d' + q')^{\phi-1} - (d' + q' + p')^{\phi-1}]$$

increases. On the other hand, if  $\phi - 1 + \gamma > 0$ , that is, if the risk-aversion parameter  $\gamma > 1$ , the denominator also increases. If the increase in the numerator dominates that in the denominator, then higher price volatility can be observed than in the CRRA case.

To understand this effect, note that the derivative of the utility cost of self-control with respect to wealth is positive if  $v(\cdot)$  is concave:

$$v'(d' + q') - v'(d' + q' + p') > 0.$$

This means that as agents get wealthier, self-control costs become lower. This explains why it might be possible to get higher price volatility in this case. The mechanism behind this result still involves the three effects discussed above: smoothing, temptation, and self-control. The difference is on the latter effect: under concavity, self-control costs are *decreasing* in wealth. This gives the agent an incentive to save *more* rather than less. If this self-control effect dominates the temptation effect, then these preferences will produce higher price volatility.

Notice that when  $v(\cdot)$  is concave, conditions need to be imposed to guarantee that  $W(\cdot)$  is strictly concave, so that the solution corresponds to a maximum (e.g., see Stokey and Lucas, 1989). In particular, the second derivative of  $W(\cdot)$  must be negative:

$$-\gamma(d + q)^{-\gamma-1} + \lambda(\phi - 1)[(d + q)^{\phi-2} - (d + q + p)^{\phi-2}] < 0, \quad (3.120)$$

which holds for any  $d$ ,  $q$ , and  $p > 0$ , and for  $\gamma > 0$ ,  $\lambda > 0$ , and  $0 < \phi < 1$ . The empirical implementation in Part V of this book proceeds under this set of parameter restrictions.

Finally, from the optimality conditions under self-control preferences, steady state temptation consumption is

$$\tilde{c} = 1 + \eta + \bar{p}.$$

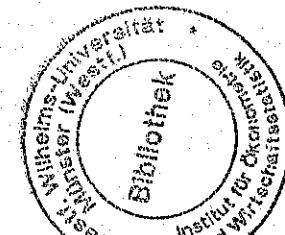
From (3.118), the steady state price in this case is given by

$$\bar{p} = \beta(1 + \bar{p}) \times \left[ \frac{(1 + \eta)^{-\gamma} + \lambda(1 + \eta)^{\phi-1} - \lambda(1 + \eta + \bar{p})^{\phi-1}}{(1 + \eta)^{-\gamma} + \lambda(1 + \eta)^{\phi-1}} \right]. \quad (3.121)$$

Regarding (3.121), the left-hand side is a 45-degree line. The right-hand side is strictly concave in  $\bar{p}$ , has a positive intercept, and a positive slope that is less than one at the intercept. Thus (3.121) yields a unique positive solution for  $\bar{p}$  for any admissible parameterization of the model. (In practice, (3.121) can be solved numerically, e.g., using GAUSS's quasi-Newton algorithm NLSYS; see Judd, 1998, for a presentation of alternative solution algorithms.) An increase in  $\lambda$  causes the function of  $\bar{p}$  on the right-hand side of (3.121) to shift down and flatten; thus  $\bar{p}$  is decreasing in  $\lambda$ . The intuition for this is again straightforward: an increase in  $\lambda$  represents an intensification of the household's temptation to liquidate its asset holdings. This drives down its demand for asset shares, and thus  $\bar{p}$ . Note the parallel between this effect and that generated by an increase in  $r$ , or a decrease in  $\beta$ , which operates analogously in both (3.99) and (3.121).

### Exercise 3.8

Solve for  $\bar{p}$  in (3.121) using  $\beta = 0.96$ ,  $\gamma = 2$ ,  $\lambda = 0.01$ ,  $\eta = 10$ ,  $\phi = 0.4$ . Log-linearize the asset-pricing equation (3.119) using the steady state values for  $(\bar{p}, \bar{d}, \bar{q})$  implied by these parameter values.



## Chapter 4 - Structural Macroeconomics

### Linear Solution Techniques

Chapters 2 and 3 demonstrated the mapping of DSGE models from nonlinear first-order systems of expectational difference equations of the form

$$\Gamma(E_t z_{t+1}, z_t, v_{t+1}) = 0 \quad (4.1)$$

into linear or log-linear representations of the form

$$Ax_{t+1} = Bx_t + Cv_{t+1} + D\eta_{t+1}. \quad (4.2)$$

The purpose of this chapter is to characterize the derivation of solutions of (4.2), expressed as

$$x_{t+1} = F(\mu)x_t + G(\mu)v_{t+1}. \quad (4.3)$$

Absent stochastic innovations, linear systems admit three types of dynamic behavior in general. Systems are said to be asymptotically stable if for any given initial value of the state variables  $s_0$ , the system will converge to  $\bar{x}$  given any choice of control variables  $c_0$  satisfying feasibility constraints implicit in (4.2). Systems are said to be unstable if any perturbation from  $\bar{x}$  leads to divergence from  $\bar{x}$  regardless of the subsequent choice of controls. Finally, systems are said to be saddle-path stable if for a given value  $s_0$ , there exists a unique choice of  $c_0$  that ensures asymptotic convergence to  $\bar{x}$ . These systems are of primary interest here. For an excellent general description of dynamic systems presented in the context of macroeconomic models, see Azariadis (1993).

The unique mapping from states to controls that ensures asymptotic convergence under saddle-path systems is given by

$$c_t = c(s_t); \quad (4.4)$$

these are the policy functions we seek through the application of alternative solution methods. Given this mapping, the behavior of  $x_t$  under (4.3) takes the form of stationary fluctuations around steady state values in response to the stochastic innovations  $v_t$ . Moreover, under this mapping the expectational errors  $\eta_t$  will depend linearly on these stochastic innovations.

Here we present four popular approaches to the derivation of (4.3). Each approach builds on the latter in some way (4.1) and (4.2) and (4.3)

specialized notation. To provide intuition behind the mechanics of the approaches, we begin by presenting mathematical background regarding homogeneous systems. Next, we demonstrate the solution of two simple models, and then introduce a more complicated model that we use to map into the notation employed under each solution procedure to aid with our exposition. The first example demonstrates the imposition of saddle-path constraints in a simple deterministic system. The second example demonstrates the mapping between expectational errors and stochastic innovations in a simple stochastic system. Finally, we present the solution procedures.

#### 4.1 Homogeneous Systems

An  $n$ -dimensional first-order homogeneous system is given by

$$\begin{bmatrix} y_{1,t+1} \\ y_{2,t+2} \\ \vdots \\ y_{n,t+1} \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & & & \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} y_{1,t} \\ y_{2,t} \\ \vdots \\ y_{n,t} \end{bmatrix}. \quad (4.5)$$

To begin, consider the case in which  $A$  is diagonal, so that (4.5) represents a system of  $n$  independent equations. Given the initialization  $y_0$ , the dynamic trajectory of the  $i^{\text{th}}$  variable is given by

$$y_{i,t} = a_{ii}^t y_{i,0}, \quad (4.6)$$

which is obtained via forward recursion on

$$y_{i,t+1} = a_{ii} y_{i,t}. \quad (4.7)$$

Note that the relationship of  $|a_{ii}|$  relative to 1 is critical for determining the behavior of  $\{y_{i,t}\}$ . If  $|a_{ii}| < 1$ ,  $y_{i,t}$  will converge asymptotically to 0 for any  $y_{i,0}$ ; if  $|a_{ii}| > 1$ ,  $y_{i,t}$  will diverge towards infinity for any  $y_{i,0}$ ; and if  $|a_{ii}| = 1$ ,  $y_{i,t} = y_{i,0} \forall t$ .

Consider now the case in which  $A$  is diagonalizable; a sufficient condition for this case is that the eigenvalues of  $A$  are distinct. Let  $\Lambda$  denote a diagonal matrix with the eigenvalues of  $A$  on the main diagonal, and  $E$  the matrix of corresponding eigenvectors. Specifically, the columns of  $E$  are linearly independent, and the  $i^{\text{th}}$  column  $e_i = (e_{i1}, e_{i2}, \dots, e_{in})'$  corresponds with the  $i^{\text{th}}$  eigenvalue  $\lambda_i$ . Then we have

$$E^{-1}AE = \Lambda, \quad EE^{-1} = I,$$

which can be used to transform (4.5) into a diagonal system.

#### 4.1 Homogeneous Systems

To achieve the desired transformation, premultiply (4.5) by  $E^{-1}$ , which yields

$$\begin{aligned} E^{-1}y_{t+1} &\equiv \hat{y}_{t+1} = E^{-1}A(EE^{-1})y_t \\ &= \Lambda E^{-1}y_t \\ &= \Lambda \hat{y}_t. \end{aligned} \quad (4.8)$$

Once again, we have a system of  $n$  independent equations; given the initialization  $\hat{y}_0$ , the dynamic trajectory of  $\hat{y}_{i,t}$  is given by

$$\hat{y}_{i,t} = \lambda_i^t \hat{y}_{i,0}. \quad (4.9)$$

And again, the relationship of  $|\lambda_{ii}|$  to 1 is critical for determining the dynamic behavior of  $\{\hat{y}_{i,t}\}$ . Hereafter, we assume without loss of generality that the diagonal elements of  $\Lambda$  are arranged in ascending order.

Now consider the context appropriate for DSGE models under which a subset  $n_c$  of the elements of  $y_t$  consists of endogenous control variables, and the remaining elements  $n_s = n - n_c$  are state variables. Then there are three possibilities regarding the dynamic properties of the system, given that  $A$  is diagonalizable.

First, if exactly  $n_c$  of the eigenvalues of  $A$  are greater than 1, then exactly  $n_c$  elements of  $\hat{y}_t$  will potentially exhibit explosive behavior. Note carefully our use of the word "potentially": explosiveness can be avoided given judicious choice of the control variables. Specifically, given  $s_0$ , if  $c_0$  is chosen such that

$$\hat{y}_{i,0} = 0, \quad i = n_s + 1, \dots, n, \quad (4.10)$$

then the influence of the explosive eigenvalues on the dynamics of the system is eliminated. Given  $s_0$ , note that (4.10) amounts to a system of  $n_c$  equations in the  $n_c$  control variables; these are saddle-path conditions, the solution to which yields the policy functions (4.4) we seek. Thus under this scenario the system is said to be saddle-path stable. Once (4.10) has been imposed, the dynamics of the original variables can be recovered by inverting the transformation  $E^{-1}y_t$  used to obtain (4.8):

$$E\hat{y}_t = EE^{-1}y_t = y_t. \quad (4.11)$$

If more than  $n_c$  of the eigenvalues of  $A$  are greater than 1, it will not be possible in general to specify the  $n_c$  controls such that (4.10) is satisfied for arbitrary  $s_0$ . Under this scenario, the system is said to be unstable, or a "source." And if less than  $n_c$  of the eigenvalues of  $A$  are greater than 1, there will exist many possible choices for the controls such that (4.10) is satisfied for arbitrary  $s_0$ . Under this scenario, the system is said to be stable, or a "sink."

In the presentation that follows, we focus attention on saddle-path systems. In the context of DSGE models, such systems yield unique mappings from states to controls, which when imposed yield stationary systems of the form (4.3).

## 4.2 Example Models

### 4.2.1 The Optimal Consumption Model

Consider a simplification of the RBC model introduced in section 1 of chapter 3 wherein  $z_t = 1 \forall t$ , and  $\varphi = 1$ , implying  $n_t = 1 \forall t$ . The model can be collapsed into the following two equations in  $(k_t, c_t)$ :

$$\begin{aligned} u'(c_t) &= \beta u'(c_{t+1})(f'(k_t) + (1 - \delta)) \\ k_{t+1} &= f(k_t) - c_t + (1 - \delta)k_t. \end{aligned}$$

Using the functional forms

$$u(c_t) = \frac{c_t^{1-\phi}}{1-\phi}, \quad f(k_t) = k_t^\alpha, \quad \beta = \frac{1}{1+r},$$

the linearized model can be written as

$$\begin{bmatrix} \tilde{k}_{t+1} \\ \tilde{c}_{t+1} \end{bmatrix} = \begin{bmatrix} 1+r & -1 \\ -\mu & 1 \end{bmatrix} \begin{bmatrix} \tilde{k}_t \\ \tilde{c}_t \end{bmatrix}, \quad (4.12)$$

where

$$\mu = \frac{(1-\alpha)(r+\delta)}{\phi(1+r)} \frac{\bar{k}}{\bar{k}} > 0,$$

$$\bar{k} = \left( \frac{\alpha}{r+\delta} \right)^{1/(1-\alpha)}, \quad \frac{\bar{c}}{\bar{k}} = \frac{r+\delta}{\alpha} - \delta,$$

and tildes denote deviations from steady state.

Note that (4.12) is a specialization of (4.2), with

$$A = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}, \quad B = \begin{bmatrix} 1+r & -1 \\ -\mu & 1 \end{bmatrix}.$$

Note further that it is a homogeneous system of the form (4.5). The eigenvalues of the system are given by

$$\lambda_i = 1 + \frac{r \pm \sqrt{r^2 + 4\mu}}{2}, \quad \lambda_1 < \lambda_2. \quad (4.13)$$

### 4.2 Example Models

Since  $\mu > 0$ ,  $\lambda_1 < 1$ , and  $\lambda_2 > 1$ , the system is saddle-path stable. Denoting  $E^{-1}$  as

$$E^{-1} = \begin{bmatrix} a & c \\ b & d \end{bmatrix},$$

the decoupled system is given by

$$\begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} \tilde{k}_{t+1} \\ \tilde{c}_{t+1} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ 0 & \lambda_2 \end{bmatrix} \begin{bmatrix} a & c \\ b & d \end{bmatrix} \begin{bmatrix} \tilde{k}_t \\ \tilde{c}_t \end{bmatrix}. \quad (4.14)$$

The associated saddle-path condition is given by

$$b\tilde{k}_t + d\tilde{c}_t = 0,$$

which yields the policy function

$$\tilde{c}_t = -\frac{b}{d}\tilde{k}_t. \quad (4.15)$$

Imposing (4.15), (4.14) can be written as

$$\begin{bmatrix} \tilde{k}_{t+1} \\ \tilde{c}_{t+1} \end{bmatrix} = \begin{bmatrix} \lambda_1 & 0 \\ -\lambda_1 \frac{b}{d} & 0 \end{bmatrix} \begin{bmatrix} \tilde{k}_t \\ \tilde{c}_t \end{bmatrix}, \quad (4.16)$$

which is the specific version of (4.3) we seek.

### 4.2.2 Asset Pricing with Linear Utility

Consider now a simple version of the single-asset pricing model introduced in section 3 of chapter 3. Here, utility is linear and  $q_t = 0 \forall t$ , so that the model is given by

$$p_t = \beta E_t(p_{t+1} + d_{t+1}) \quad (4.17)$$

$$d_{t+1} = \rho d_t + v_{t+1}. \quad (4.18)$$

Note that we have eliminated a nonzero steady state term for  $d_t$  (and thus also  $p_t$ ), which merely serves to ease notation. Dropping the expectations operator from the pricing equation yields

$$p_t = \beta(p_{t+1} + d_{t+1} + \eta_{t+1}). \quad (4.19)$$

Thus written in the form of (4.2), the model is given by

$$\begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} p_{t+1} \\ d_{t+1} \end{bmatrix} = \begin{bmatrix} \beta^{-1} & 0 \\ 0 & \rho \end{bmatrix} \begin{bmatrix} p_t \\ d_t \end{bmatrix} + \begin{bmatrix} 0 \\ 1 \end{bmatrix} v_{t+1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \eta_t.$$

Here we use a very simple guess-and-verify approach to solving for the policy function  $p_t = c(d_t)$ ; for a far more rigorous and general approach

of this nature, see Whiteman (1983). Specifically, we guess that prices are proportional to dividends:

$$p_{t+1} = \theta d_{t+1}. \quad (4.20)$$

Substituting for  $p_t$  in (4.17), we find that the guess is consistent with the model, and in addition, obtain the following expression for the factor of proportionality:

$$\theta = \frac{\rho\beta}{1 - \rho\beta},$$

thus the guess is verified. Combining (4.18) and (4.20) yields

$$p_{t+1} = \rho p_t + \theta v_{t+1}. \quad (4.21)$$

Comparing (4.18) and (4.21), we see a very simple example of the cross-equation restrictions that Thomas Sargent has characterized as "the hallmark of rational expectations models."

To obtain the mapping from stochastic innovations to expectational errors, substitute for  $(d_{t+1}, p_{t+1})$  in (4.19) using (4.18) and (4.21) to obtain

$$p_t = \beta(\rho p_t + \theta v_{t+1} + \rho d_t + v_{t+1} + \eta_{t+1}).$$

Next, substitute the one-step-ahead forecasts  $(\rho d_t, \rho p_t)$  for  $(E_t d_{t+1}, E_t p_{t+1})$  in (4.17) to obtain

$$p_t = \beta(\rho p_t + \rho d_t).$$

Comparing the resulting expressions, we obtain

$$\begin{aligned} \eta_{t+1} &= -(1 + \theta)v_{t+1} \\ &= -\frac{1}{1 - \rho\beta}v_{t+1}. \end{aligned} \quad (4.22)$$

Thus stochastic innovations are seen to map directly into expectational errors.

### 4.2.3 Ramsey's Optimal Growth Model

The foregoing examples were intended to help set intuition. Here we offer a slightly richer example that we will map into the notation employed under each solution approach to aid with their exposition.

The example is a linearized stochastic version of Ramsey's (1928) optimal growth model:

$$\tilde{y}_{t+1} - \tilde{a}_{t+1} - \alpha \tilde{k}_{t+1} = 0 \quad (4.23)$$

$$\tilde{y}_{t+1} - \gamma_c \tilde{c}_{t+1} - \gamma_i \tilde{i}_{t+1} = 0 \quad (4.24)$$

### 4.3 Blanchard and Kahn's Method

$$\theta_{1c} E_t(\tilde{c}_{t+1}) + \theta_a E_t(\tilde{a}_{t+1}) + \theta_k E_t(\tilde{k}_{t+1}) + \theta_{2c} \tilde{c}_t = 0 \quad (4.25)$$

$$\tilde{k}_{t+1} - \delta_k \tilde{k}_t - \delta_i \tilde{i}_t = 0 \quad (4.26)$$

$$\tilde{a}_{t+1} - \rho \tilde{a}_t = \varepsilon_{t+1}. \quad (4.27)$$

The variables  $\{\tilde{y}_t, \tilde{c}_t, \tilde{i}_t, \tilde{k}_t, \tilde{a}_t\}$  represent output, consumption, investment, physical capital, and a productivity shock, all expressed as logged deviations from steady state values. The variable  $\varepsilon_t$  is a serially uncorrelated stochastic process. The vector

$$\mu = [\alpha \ \gamma_c \ \gamma_i \ \theta_{1c} \ \theta_a \ \theta_k \ \theta_{2c} \ \delta_k \ \delta_i \ \rho]'$$

contains the "deep" parameters of the model.

Dropping the expectations operator  $E_t(\cdot)$  from (4.25) introduces an expectational error into the modified equation; denote this error as  $\eta_{ct+1}$ . Then the specific form of (4.2) in this example is given by

$$\begin{aligned} &\left[ \begin{array}{cccccc} 1 & 0 & 0 & -\alpha & -1 & \\ 1 & -\gamma_c & -\gamma_i & 0 & 0 & \\ 0 & \theta_{1c} & 0 & \theta_k & \theta_a & \\ 0 & 0 & 0 & 1 & 0 & \\ 0 & 0 & 0 & 0 & 1 & \end{array} \right] \left[ \begin{array}{c} \tilde{y}_{t+1} \\ \tilde{c}_{t+1} \\ \tilde{i}_{t+1} \\ \tilde{k}_{t+1} \\ \tilde{a}_{t+1} \end{array} \right] \\ &= \underbrace{\left[ \begin{array}{ccccc} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -\theta_{2c} & 0 & 0 & 0 \\ 0 & 0 & \delta_i & \delta_k & 0 \\ 0 & 0 & 0 & 0 & 0 \end{array} \right]}_B \underbrace{\left[ \begin{array}{c} \tilde{y}_t \\ \tilde{c}_t \\ \tilde{i}_t \\ \tilde{k}_t \\ \tilde{a}_t \end{array} \right]}_{x_t} + \underbrace{\left[ \begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 1 \end{array} \right]}_{v_{t+1}} \underbrace{\varepsilon_t + \left[ \begin{array}{c} 0 \\ 0 \\ 0 \\ 0 \\ 0 \end{array} \right]}_{\eta_{ct+1}} + \underbrace{\left[ \begin{array}{c} 0 \\ 0 \\ 1 \\ 0 \\ 0 \end{array} \right]}_{\eta_{t+1}}. \end{aligned} \quad (4.28)$$

### 4.3 Blanchard and Kahn's Method

The first solution method we present was developed by Blanchard and Kahn (1980), and is applied to models written as

$$\begin{bmatrix} x_{1t+1} \\ E_t(x_{2t+1}) \end{bmatrix} = \tilde{A} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} + Ef_t, \quad (4.29)$$

where the model variables have been divided into an  $n_1 \times 1$  vector of endogenous predetermined variables  $x_{1t}$  (defined as variables for which

$E_t x_{1t+1} = x_{1t+1}$ ), and an  $n_2 \times 1$  vector of endogenous nonpredetermined variables  $x_{2t}$ . The  $k \times 1$  vector  $f_t$  contains exogenous forcing variables.

Following the approach of King and Watson (2002), a preliminary step is taken before casting a given model into the form (4.29). The step is referred to as a system reduction: it involves writing the model in terms of a subset of variables that are uniquely determined. In terms of the example, note that observations on  $\tilde{a}_t$  and  $\tilde{k}_t$  are sufficient for determining  $\tilde{y}_t$  using (4.23), and that given  $\tilde{y}_t$ , the observation of either  $\tilde{c}_t$  or  $\tilde{i}_t$  is sufficient for determining both variables using (4.24). Thus we proceed in working directly with  $\{\tilde{c}_t, \tilde{k}_t, \tilde{a}_t\}$  using (4.25)–(4.27), and recover  $\{\tilde{y}_t, \tilde{i}_t\}$  as functions of  $\{\tilde{c}_t, \tilde{k}_t, \tilde{a}_t\}$  using (4.23) and (4.24). Among  $\{\tilde{c}_t, \tilde{k}_t, \tilde{a}_t\}$ ,  $\tilde{k}_t$  is predetermined (given  $k_t$  and  $i_t$ ,  $k_{t+1}$  is determined as in (4.26));  $\tilde{c}_t$  is endogenous but not predetermined (as indicated in (4.25), its time- $(t+1)$  realization is associated with an expectations error); and  $\tilde{a}_t$  is an exogenous forcing variable. Thus in the notation of (4.29), we seek a specification of the model in the form

$$\begin{bmatrix} \tilde{k}_{t+1} \\ E_t(\tilde{c}_{t+1}) \end{bmatrix} = \tilde{A} \begin{bmatrix} \tilde{k}_t \\ \tilde{c}_t \end{bmatrix} + E \tilde{a}_t. \quad (4.30)$$

To obtain this expression, let  $\xi_t = [\tilde{y}_t \ \tilde{i}_t]'$ ,  $\zeta_t = [\tilde{k}_t \ \tilde{c}_t]'$ , and note that  $E_t(\tilde{a}_{t+1}) = \rho \tilde{a}_t$ . In terms of these variables, the model may be written as

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & -\gamma_i \end{bmatrix}}_{\Psi_0} \xi_t = \underbrace{\begin{bmatrix} \alpha & 0 \\ 0 & \gamma_c \end{bmatrix}}_{\Psi_1} \zeta_t + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\Psi_2} \tilde{a}_t \quad (4.31)$$

$$\underbrace{\begin{bmatrix} \theta_k & \theta_{1c} \\ 1 & 0 \end{bmatrix}}_{\Psi_3} E_t(\xi_{t+1}) = \underbrace{\begin{bmatrix} 0 & -\theta_{2c} \\ \delta_k & 0 \end{bmatrix}}_{\Psi_4} \xi_t + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & \delta_i \end{bmatrix}}_{\Psi_5} \zeta_t + \underbrace{\begin{bmatrix} -\theta_{k\rho} \\ 0 \end{bmatrix}}_{\Psi_6} \tilde{a}_t. \quad (4.32)$$

Next, substituting (4.31) into (4.32), which requires inversion of  $\Psi_0$ , we obtain

$$\Psi_3 E_t(\xi_{t+1}) = [\Psi_4 + \Psi_5 \Psi_0^{-1} \Psi_1] \xi_t + [\Psi_6 + \Psi_5 \Psi_0^{-1} \Psi_2] \tilde{a}_t. \quad (4.33)$$

Finally, premultiplying (4.33) by  $\Psi_3^{-1}$  yields a specification in the form of (4.30); Blanchard and Kahn's solution method may now be implemented. Hereafter, we describe its implementation in terms of the notation employed in (4.29).

### 4.3 Blanchard and Kahn's Method

The method begins with a Jordan decomposition of  $\tilde{A}$ , yielding

$$\tilde{A} = \Lambda^{-1} J \Lambda, \quad (4.34)$$

where the diagonal elements of  $J$ , consisting of the eigenvalues of  $\tilde{A}$ , are ordered in increasing absolute value in moving from left to right.<sup>1</sup> Thus  $J$  may be written as

$$J = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}, \quad (4.35)$$

where the eigenvalues in  $J_1$  lie on or within the unit circle, and those in  $J_2$  lie outside of the unit circle.  $J_2$  is said to be unstable or explosive, since  $J_2^n$  diverges as  $n$  increases. The matrices  $\Lambda$  and  $E$  are partitioned conformably as

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}, \quad E = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}, \quad (4.36)$$

where  $\Lambda_{11}$  is conformable with  $J_1$ , and so forth. As noted, if the number of explosive eigenvalues is equal to the number of nonpredetermined variables, the system is said to be saddle-path stable and a unique solution to the model exists. If the number of explosive eigenvalues exceeds the number of nonpredetermined variables, no solution exists (and the system is said to be a source); and in the opposite case an infinity of solutions exists (and the system is said to be a sink).

Proceeding under the case of saddle-path stability, substitution for  $\tilde{A}$  in (4.30) yields

$$\begin{bmatrix} x_{1t+1} \\ E_t(x_{2t+1}) \end{bmatrix} = \Lambda^{-1} J \Lambda \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} + \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} f_t. \quad (4.37)$$

Next, the system is premultiplied by  $\Lambda$ , yielding

$$\begin{bmatrix} \hat{x}_{1t+1} \\ E_t(\hat{x}_{2t+1}) \end{bmatrix} = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} \begin{bmatrix} \hat{x}_{1t} \\ \hat{x}_{2t} \end{bmatrix} + \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} f_t, \quad (4.38)$$

where

$$\begin{bmatrix} \hat{x}_{1t} \\ \hat{x}_{2t} \end{bmatrix} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \quad (4.39)$$

$$\begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}. \quad (4.40)$$

This transformation effectively "decouples" the system, so that the non-predetermined variables depend only upon the unstable eigenvalues of  $\tilde{A}$  contained in  $J_2$ , as expressed in the lower part of (4.38).

<sup>1</sup> Eigenvalues of a matrix  $\Theta$  are obtained from the solution of equations of the form  $\Theta e = \lambda e$ , where  $e$  is an eigenvector and  $\lambda$  the associated eigenvalue. The GAUSS command eigv performs this decomposition.

Having decoupled the system, we derive a solution for the nonpredetermined variables by performing a forward iteration on the lower portion of (4.38). Using  $f_{2t}$  to denote the portion of  $f_t$  conformable with  $D_2$ , this is accomplished as follows. First, reexpress the lower portion of (4.38) as

$$\dot{x}_{2t} = J_2^{-1} E_t(\dot{x}_{2t+1}) - J_2^{-1} D_2 f_{2t}. \quad (4.41)$$

This implies an expression for  $x_{2t+1}$  of the form

$$\dot{x}_{2t+1} = J_2^{-1} E_{t+1}(\dot{x}_{2t+2}) - J_2^{-1} D_2 f_{2t+1}, \quad (4.42)$$

which can be substituted into (4.41) to obtain

$$\dot{x}_{2t} = J_2^{-2} E_t(\dot{x}_{2t+1}) - J_2^{-2} D_2 E_t(f_{2t+1}) - J_2^{-1} D_2 f_{2t}. \quad (4.43)$$

In writing (4.43) we have exploited the Law of Iterated Expectations, which holds that  $E_t[E_{t+1}(x_t)] = E_t(x_t)$  for any  $x_t$  (e.g., see Ljungqvist and Sargent, 2004). Since  $J_2$  contains explosive eigenvalues,  $J_2^{-n}$  disappears as  $n$  approaches infinity, thus continuation of the iteration process yields

$$\dot{x}_{2t} = - \sum_{i=0}^{\infty} J_2^{-(i+1)} D_2 E_t(f_{2t+i}). \quad (4.44)$$

Mapping this back into an expression for  $x_{2t}$  using (4.39), we obtain

$$x_{2t} = -\Lambda_{22}^{-1} \Lambda_{21} x_{1t} - \Lambda_{22}^{-1} \sum_{i=0}^{\infty} J_2^{-(i+1)} D_2 E_t(f_{2t+i}). \quad (4.45)$$

In the case of the example model presented above,  $E_t(f_{2t+i}) = \rho^i \tilde{a}_t$ , and thus (4.45) becomes

$$x_{2t} = -\Lambda_{22}^{-1} \Lambda_{21} x_{1t} - \Lambda_{22}^{-1} J_2^{-1} (I - \rho J_2^{-1} D_2)^{-1} \tilde{a}_t. \quad (4.46)$$

Finally, to solve the nonexplosive portion of the system, begin by expanding the upper portion of (4.37):

$$x_{1t+1} = \tilde{A}_{11} x_{1t} + \tilde{A}_{22} x_{2t} + E_1 f_t, \quad (4.47)$$

where  $\tilde{A}_{11}$  and  $\tilde{A}_{22}$  are partitions of  $\Lambda^{-1} J \Lambda$  conformable with  $x_{1t}$  and  $x_{2t}$ . Then substituting for  $x_{2t}$  using (4.45) yields a solution for  $x_{1t}$  of the form given by (4.3).

We conclude this subsection by highlighting two requirements of this solution method. First, a model-specific system reduction is employed to obtain an expression of the model that consists of a subset of its variables. The variables in the subset are distinguished as being either predetermined or nonpredetermined. Second, invertibility of the lead matrices  $\Psi_0$  and  $\Psi_3$  is required in order to obtain a specification of the model amenable for solution.

#### 4.4 Sims' Method

##### Exercise 4.1

Write computer code for mapping the example model expressed in (4.23)–(4.27) into the form of the representation given in (4.29).

#### 4.4 Sims' Method

Sims (2001) proposes a solution method applied to models expressed as

$$Ax_{t+1} = Bx_t + E + Cv_{t+1} + D\eta_{t+1}, \quad (4.48)$$

where  $E$  is a matrix of constants.<sup>2</sup> Relative to the notation we have employed above,  $E$  is unnecessary because the variables in  $x_t$  are expressed in terms of deviations from steady state values. Like Blanchard and Kahn's (1980) method, Sims' method involves a decoupling of the system into explosive and nonexplosive portions. However, rather than expressing variables in terms of expected values, expectations operators have been dropped, giving rise to the expectations errors contained in  $\eta_{t+1}$ . Also, while Blanchard and Kahn's method entails isolation of the forcing variables from  $x_{t+1}$ , these are included in  $x_{t+1}$  under Sims' method; thus the appearance in the system of the vector of shocks to these variables  $v_{t+1}$ . Third, Sims' method does not require an initial system-reduction step. Finally, it does not entail a distinction between predetermined and nonpredetermined variables.

Note from (4.28) that the example model has already been cast in the form of (4.48); thus we proceed directly to a characterization of the solution method. The first step employs a “*QZ factorization*” to decompose  $A$  and  $B$  into unitary upper triangular matrices:

$$A = Q' \Lambda Z' \quad (4.49)$$

$$B = Q' \Omega Z', \quad (4.50)$$

where  $(Q, Z)$  are unitary, and  $(\Lambda, \Omega)$  are upper triangular.<sup>3</sup> Next,  $(Q, Z, \Lambda, \Omega)$  are ordered such that, in absolute value, the generalized eigenvalues of  $A$  and  $B$  are organized in  $\Lambda$  and  $\Omega$  in increasing order moving from left to right, just as in Blanchard and Kahn's Jordan decomposition

<sup>2</sup> The programs available on Sims' website perform all of the steps of this procedure. The web address is <http://www.princeton.edu/~sims/>. The programs are written in Matlab; analogous code written in GAUSS is available at the textbook website.

<sup>3</sup> A unitary matrix  $\Theta$  satisfies  $\Theta' \Theta = \Theta \Theta' = I$ . If  $Q$  and/or  $Z$  contains complex values, the transpositions reflect complex conjugation, that is, each complex entry is replaced by its conjugate and then transposed.

procedure.<sup>4</sup> Having obtained the factorization, the original system is then premultiplied by  $Q$ , yielding the transformed system expressed in terms of  $z_{t+1} = Z'x_{t+1}$ :

$$\Lambda z_t = \Omega z_{t-1} + \underbrace{QE + QCv_t + QD\eta_t}_{\omega_t}, \quad (4.51)$$

where we have lagged the system by one period in order to match the notation (and code) of Sims.

Next, as with Blanchard and Kahn's (1980) method, (4.51) is partitioned into explosive and nonexplosive blocks:

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ 0 & \Omega_{22} \end{bmatrix} \begin{bmatrix} z_{1t-1} \\ z_{2t-1} \end{bmatrix} + \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} [E + Cv_t + D\eta_t]. \quad (4.52)$$

The explosive block (the lower equations) is solved as follows. Letting  $w_t = Q(E + Cv_t + D\eta_t)$  (partitioned conformably as  $w_{1t}$  and  $w_{2t}$ ), the lower block of (4.52) is given by

$$\Lambda_{22} z_{2t} = \Omega_{22} z_{2t-1} + w_{2t}. \quad (4.53)$$

Leading (4.53) by one period and solving for  $z_{2t}$  yields

$$z_{2t} = Mz_{2t+1} - \Omega_{22}^{-1} w_{2t+1}, \quad (4.54)$$

where  $M = \Omega_{22}^{-1} \Lambda_{22}$ . Then recursive substitution for  $z_{2t+1}, z_{2t+2}, \dots$  yields

$$z_{2t} = -\sum_{i=0}^{\infty} M^i \Omega_{22}^{-1} w_{2t+1+i}, \quad (4.55)$$

since  $\lim_{t \rightarrow \infty} M^t z_{2t} = 0$ . Recalling that  $w_t$  is defined as  $w_t = Q(E + Cv_t + D\eta_t)$ , note that (4.55) expresses  $z_{2t}$  as a function of future values of structural and expectational errors. But  $z_{2t}$  is known at time  $t$ , and  $E_t(\eta_{t+s}) = E_t(v_{t+s}) = 0$  for  $s > 0$ ; thus (4.55) may be written as

$$z_{2t} = -\sum_{i=0}^{\infty} M^i \Omega_{22}^{-1} Q_2 E_i, \quad (4.56)$$

$$z_{2t} = Q_2 \cdot E \cdot (I - M \cdot Q_2)^{-1} \Lambda_{22}$$

<sup>4</sup> Generalized eigenvalues of  $\Theta$  are obtained as the solution to  $\Theta e = \lambda \Xi e$ , where  $\Xi$  is a symmetric matrix.

#### 4.4 Sims' Method

where  $Q_2 E_2$  are the lower portions of  $QE$  conformable with  $z_2$ .<sup>5</sup> Postmultiplying (4.56) by  $\Omega_{22}^{-1} Q_2 E_2$  and noting that  $-\sum_{i=0}^{\infty} M^i = -(I - M)^{-1}$ , we obtain the solution of  $z_{2t}$  as

$$z_{2t} = (\Lambda_{22} - \Omega_{22})^{-1} Q_2 E. \quad (4.57)$$

Having solved for  $z_{2t}$ , the final step is to solve for  $z_{1t}$  in (4.52). Note that the solution of  $z_{1t}$  requires a solution for the expectational errors that appear in (4.52). Recall that when a unique solution for the model exists, it will be the case that a systematic relationship exists between the expectational errors associated with  $z_{1t}$  and  $z_{2t}$ ; exploiting this relationship yields a straightforward means of solving for  $z_{1t}$ . The necessary and sufficient condition for uniqueness is given by the existence of a  $k \times (n - k)$  matrix  $\Phi$  that satisfies

$$Q_1 D = \Phi Q_2 D, \quad (4.58)$$

which represents the systematic relationship between the expectational errors associated with  $z_{1t}$  and  $z_{2t}$  noted above. Given uniqueness, and thus the ability to calculate  $\Phi$  as in (4.58), the solution of  $z_{1t}$  proceeds with the premultiplication of (4.51) by  $[I - \Phi]$ . With  $z_1$  being  $n_s \times 1$  and  $z_2$  being  $n_c \times 1$ , such that  $n_s + n_c = n$ , then in this case  $I$  is  $n_s \times n_s$  and  $\Phi$  is  $n_s \times n_c$ . Premultiplication yields

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} - \Phi \Lambda_{22} \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} - \Phi \Omega_{22} \end{bmatrix} \begin{bmatrix} z_{1t-1} \\ z_{2t-1} \end{bmatrix} + [Q_1 - \Phi Q_2] [E + Cv_t + D\eta_t]. \quad (4.59)$$

Then due to (4.58), the loading factor for the expectational errors in (4.59) is zero, and thus the system may be written in the form

$$x_t = \Theta_E + \Theta_0 x_{t-1} + \Theta_1 v_t, \quad (4.60)$$

where

$$H = Z \begin{bmatrix} \Lambda_{11}^{-1} & -\Lambda_{11}^{-1}(\Lambda_{12} - \Phi \Lambda_{22}) \\ 0 & I \end{bmatrix} \quad (4.61)$$

$$\Theta_E = H \begin{bmatrix} Q_1 - \Phi Q_2 \\ (\Omega_{22} - \Lambda_{22})^{-1} Q_2 \end{bmatrix} E \quad (4.62)$$

<sup>5</sup> Sims also considers the case in which the structural innovations  $v_t$  are serially correlated, which leads to a generalization of (4.56).

$$\Theta_0 = Z_{\cdot 1} \Lambda_{11}^{-1} [\Omega_{11} - (\Omega_{12} - \Phi \Omega_{22})] Z \quad (4.63)$$

$$\Theta_1 = H \begin{bmatrix} Q_1 - \Phi Q_2 \\ 0 \end{bmatrix} \cancel{P}. \quad (4.64)$$

In (4.63),  $Z_{\cdot 1}$  is defined as the  $n \times n_s$  matrix containing the first through  $n_s^{\text{th}}$  columns of  $Z$ .

### Exercise 4.2

Using the code cited for this method, compute the solution for (4.23)–(4.27) for given values of  $\mu$ .

### 4.5 Klein's Method

Klein (2000) proposes a solution method that is a hybrid of those of Blanchard and Kahn (1980) and Sims (2001).<sup>6</sup> The method is applied to systems written as

$$\tilde{A}E_t(x_{t+1}) = \tilde{B}x_t + Ef_t, \quad (4.65)$$

where the vector  $f_t$  (of length  $n_z$ ) has a zero-mean vector autoregressive (VAR) specification with autocorrelation matrix  $\Phi$ ; additionally  $\tilde{A}$  may be singular.<sup>7</sup>

Like Blanchard and Kahn, Klein distinguishes between the predetermined and nonpredetermined variables of the model. The former are contained in  $x_{1,t+1}$ , the latter in  $x_{2,t+1}$ :  $E_t(x_{t+1}) = [x_{1,t+1} \ E_t(x_{2,t+1})]'$ . The solution approach once again involves decoupling the system into non-explosive and explosive components, and solving the two components in turn.

Returning to the example model expressed in (4.23)–(4.27), the form of the model amenable to the implementation of Klein's method is given by (4.33), repeated here for convenience:

$$\Psi_3 E_t(\zeta_{t+1}) = [\Psi_4 + \Psi_5 \Psi_0^{-1} \Psi_1] \zeta_t + [\Psi_6 + \Psi_5 \Psi_0^{-1} \Psi_2] \tilde{a}_t. \quad (4.66)$$

The main advantage of Klein's approach relative to Blanchard and Kahn's is that  $\Psi_3$  may be singular. To proceed with the description of Klein's approach, we revert to the notation employed in (4.65).

<sup>6</sup> GAUSS and Matlab codes that implement this solution method are available at <http://www.ssc.uwo.ca/economics/faculty/klein/>.

<sup>7</sup> See chapter 7 for a description of VAR models.

Klein's approach overcomes the potential noninvertibility of  $\tilde{A}$  by implementing a complex generalized Schur decomposition to decompose  $\tilde{A}$  and  $\tilde{B}$ . This is in place of the  $QZ$  decomposition employed by Sims. In short, the Schur decomposition is a generalization of the  $QZ$  decomposition that allows for complex eigenvalues associated with  $\tilde{A}$  and  $\tilde{B}$ . Given the decomposition of  $\tilde{A}$  and  $\tilde{B}$ , Klein's method closely follows that of Blanchard and Kahn.

The Schur decompositions of  $\tilde{A}$  and  $\tilde{B}$  are given by

$$Q \tilde{A} Z = S \quad (4.67)$$

$$Q \tilde{B} Z = T, \quad (4.68)$$

where  $(Q, Z)$  are unitary and  $(S, T)$  are upper triangular matrices with diagonal elements containing the generalized eigenvalues of  $\tilde{A}$  and  $\tilde{B}$ . Once again the eigenvalues are ordered in increasing value in moving from left to right. Partitioning  $Z$  as

$$Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}, \quad (4.69)$$

$Z_{11}$  is  $n_1 \times n_1$  and corresponds to the nonexplosive eigenvalues of the system. Given saddle-path stability, this conforms with  $x_1$ , which contains the predetermined variables of the model.

Having obtained this decomposition, the next step in solving the system is to triangularize (4.65) as was done in working with the  $QZ$  decomposition. Begin by defining

$$z_t = Z^H x_t, \quad (4.70)$$

where  $Z^H$  refers to a Hermitian transpose.<sup>8</sup> This transformed vector is divided into  $n_1 \times 1$  stable ( $s_t$ ) and  $n_2 \times 1$  unstable ( $u_t$ ) components. Then since  $\tilde{A} = Q' S Z^H$  and  $\tilde{B} = Q' S Z^H$ , (4.65) may be written as

$$\begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} E_t \begin{bmatrix} s_{t+1} \\ u_{t+1} \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \begin{bmatrix} s_t \\ u_t \end{bmatrix} + \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} E_t f_t; \quad (4.71)$$

once again, the linear portion of (4.71) contains the unstable components of the system. Solving this component via forward iteration, we obtain

$$u_t = M f_t \quad (4.72)$$

$$\text{vec}(M) = [(\Phi^T \otimes S_{22}) - I_{n_z} \otimes T_{22}]^{-1} \text{vec}(Q_2 E). \quad (4.73)$$

<sup>8</sup> Given a matrix  $\Theta$ , if the lower triangular portion of  $\Theta$  is the complex conjugate transpose of the upper triangle portion of  $\Theta$ , then  $\Theta$  is denoted as Hermitian.

The appearance of the *vec* operator accommodates the VAR specification for  $f_t$ . In the context of the example model,  $\Phi^T$  is replaced by the scalar  $\rho^T$ , and (4.73) becomes  $M = [\rho^T S_{22} - T_{22}]^{-1} Q_2 E$ .

This solution for the unstable component is then used to solve the stable component, yielding

$$s_{t+1} = S_{11}^{-1} T_{11} s_t + S_{11}^{-1} \{T_{12} M - S_{12} M \Phi + Q_1 E\} f_t - Z_{11}^{-1} Z_{12} M v_{t+1}, \quad (4.74)$$

where  $v_{t+1}$  is a serially uncorrelated stochastic process representing the innovations in the VAR specification for  $f_{t+1}$ . In the context of our example model,  $f_t$  corresponds to  $\tilde{\epsilon}_t$ , the innovation to which is  $\epsilon_t$ . In terms of the original variables the solution is expressed as

$$x_{2t} = Z_{21} Z_{11}^{-1} x_{1t} + N f_t \quad (4.75)$$

$$x_{1t+1} = Z_{11} S_{11}^{-1} T_{11} Z_{11}^{-1} x_{1t} + L f_t \quad (4.76)$$

$$N = (Z_{22} - Z_{21} Z_{11}^{-1} Z_{12}) M \quad (4.77)$$

$$\begin{aligned} L = & -Z_{11} S_{11}^{-1} T_{11} Z_{11}^{-1} Z_{12} M + Z_{11} S_{11}^{-1} \{T_{12} M - S_{12} M \Phi + Q_1 E\} \\ & + Z_{12} M \Phi. \end{aligned} \quad (4.78)$$

This solution can be cast into the form of (4.3) as

$$x_{1t+1} = [Z_{11} S_{11}^{-1} T_{11} Z_{11}^{-1} Z_{21} Z_{11}^{-1}] x_{1t} + [Z_{11} S_{11}^{-1} T_{11} Z_{11}^{-1} N + L] f_t. \quad (4.79)$$

### Exercise 4.3

Apply Klein's code to the example model presented in (4.23)–(4.27).

## 4.6 An Undetermined Coefficients Approach

Uhlig (1999) proposes a solution method based on the method of undetermined coefficients.<sup>9</sup> The method is applied to systems written as

$$0 = E_t [F x_{t+1} + G x_t + H x_{t-1} + L f_{t+1} + M f_t] \quad (4.80)$$

$$f_{t+1} = N f_t + v_{t+1}, \quad E_t (v_{t+1}) = 0. \quad (4.81)$$

With respect to the example model in (4.23)–(4.26) let  $x_t = [y_t \ c_t \ i_t \ k_t]'$ . Then lagging the first two equations, which are subject neither to structural shocks nor expectations errors, the matrices in (4.80) and (4.81) are given by

<sup>9</sup> Matlab code for implementing this solution method is available at <http://www.wiwi.hu-berlin.de/wpol/html/toolkit.htm>.

$$F = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \theta_{1c} & 0 & \theta_k \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 & 0 & -\alpha \\ 1 & -\gamma_c & -\gamma_i & 0 \\ 0 & \theta_{2c} & 0 & 0 \\ 0 & 0 & -\delta_i & -\delta_k \end{bmatrix}, \quad (4.82)$$

$H = 0$ ,  $L = [0 \ 0 \ \theta_a \ 0]'$ ,  $M = [-1 \ 0 \ 0 \ 0]'$ , and  $N = \rho$ .

Solutions to (4.80)–(4.81) take the form

$$x_t = P x_{t-1} + Q f_t. \quad (4.83)$$

In deriving (4.83), we will confront the problem of solving matrix quadratic equations of the form

$$\Psi P^2 - \Gamma P - \Theta = 0 \quad (4.84)$$

for the  $m \times m$  matrix  $P$ . Thus we first describe the solution of such equations.

To begin, define

$$\Xi_{2m \times 2m} = \begin{bmatrix} \Gamma & \Theta \\ I_m & 0_{m \times m} \end{bmatrix}, \quad \Delta_{2m \times 2m} = \begin{bmatrix} \Psi & 0_{m \times m} \\ 0_{m \times m} & I_m \end{bmatrix}. \quad (4.85)$$

Given these matrices, let  $s$  and  $\lambda$  denote the generalized eigenvector and eigenvalue of  $\Xi$  with respect to  $\Delta$ , and note that  $s' = [\lambda \ s' \ x']$  for some  $x \in \mathbb{R}^m$ . Then the solution to the matrix quadratic is given by

$$P = \Omega \Lambda \Omega^{-1}, \quad \Omega = [x_1, \dots, x_m], \quad \Lambda = \text{diag}(\lambda_1, \dots, \lambda_m), \quad (4.86)$$

so long as the  $m$  eigenvalues contained in  $\Lambda$  and  $(x_1, \dots, x_m)$  are linearly independent. The solution is stable if the generalized eigenvalues are all less than one in absolute value.

Returning to the solution of the system in (4.80)–(4.81), the first step towards obtaining (4.83) is to combine these three equations into a single equation. This is accomplished in two steps. First, write  $x_t$  in (4.80) in terms of its relationship with  $x_{t-1}$  given by (4.83), and do the same for  $x_{t+1}$ , where the relationship is given by

$$x_{t+1} = P^2 x_{t-1} + P Q f_t + Q f_{t+1}. \quad (4.87)$$

Next, write  $f_{t+1}$  in terms with its relationship with  $f_t$  given by (4.81). Taking expectations of the resulting equation yields

$$0 = [FP^2 + GP + H] x_{t-1} + [(FP + G)Q + M + (FQ + L)N] f_t. \quad (4.88)$$

Note that in order for (4.88) to hold, the coefficients on  $x_{t-1}$  and  $f_t$  must be zero. The first restriction implies that  $P$  must satisfy the matrix quadratic equation

$$0 = FP^2 + GP + H, \quad (4.89)$$

the solution of which is obtained as indicated in (4.85) and (4.86). The second restriction requires the derivation of  $Q$ , which satisfies

$$(FP + G)Q + M + (FQ + L)N = 0. \quad (4.90)$$

The required  $Q$  can be shown to be given by

$$Q = V^{-1}[-\text{vec}(LN + M)], \quad (4.91)$$

where  $V$  is defined as

$$V = N' \otimes F + I_k \otimes (FP + G). \quad (4.92)$$

The solutions for  $P$  and  $Q$  will be unique so long as the matrix  $P$  has stable eigenvalues.

As noted by Christiano (2002), this solution method is particularly convenient for working with models involving endogenous variables that have differing associated information sets. Such models can be cast in the form of (4.80)–(4.81), with the expectations operator  $\hat{E}_t$  replacing  $E_t$ . In terms of calculating the expectation of an  $n \times 1$  vector  $X_t$ ,  $\hat{E}_t$  is defined as

$$\hat{E}_t(X_t) = \begin{bmatrix} E(X_{1t} | \Xi_{1t}) \\ \vdots \\ E(X_{nt} | \Xi_{nt}) \end{bmatrix}, \quad (4.93)$$

where  $\Xi_{it}$  represents the information set available for formulating expectations over the  $i^{\text{th}}$  element of  $X_t$ . Thus systems involving this form of heterogeneity may be accommodated using an expansion of the system (4.80)–(4.81) specified for a representative agent. The solution of the expanded system proceeds as indicated above; for details and extensions, see Christiano (2002).

#### Exercise 4.4

Apply Uhlig's code to the example model presented in (4.23)–(4.27).

## Chapter 5 – Structural Macroeconomics

### Nonlinear Solution Techniques

Close only counts in horseshoes and handgrenades.

—Anonymous

There are two generic reasons for moving beyond linear or log-linear model representations and solutions. First, for the purpose of the task at hand, the enhanced accuracy associated with nonlinear model representations may be important. For example, Fernandez-Villaverde and Rubio-Ramirez (2005) have emphasized the importance of preserving model nonlinearities in conducting likelihood-based analyses. Second, it is often the case that departures from linearity are of direct interest in light of the question under investigation. This is the case, for example, in examining the role of precautionary motives in driving consumption/savings decisions (Hugget and Ospina, 2001); or in measuring the relative roles of stochastic volatility versus changes in monetary policy in accounting for time-varying volatility in aggregate U.S. data (Fernandez-Villaverde, Guerrier-Quintana, and Rubio-Ramirez, 2010).

Here we present three general methods for obtaining solutions to nonlinear model representations, including several specific examples of each: projection methods, iteration procedures, and perturbation methods. For textbook discussions of these and other solution methods, see Judd (1998), Adda and Cooper (2003), Ljungqvist and Sargent (2004), and Heer and Maussner (2005). And for a comparison of the performance of these methods for a standard RBC model, see Arouba, Fernandez-Villaverde, and Rubio-Ramirez (2006).

Recall that our generic representation of a DSGE model is given by

$$F(E_t z_{t+1}, z_t, v_{t+1}) = 0,$$

and that the solution we seek is of the form

$$c_t = c(s_t), \quad (5.1)$$

$$s_t = f(s_{t-1}, v_t). \quad (5.2)$$

The policy function  $c(s_t)$  is derived as the solution to

$$F(c(s)) = 0, \quad (5.3)$$

Schmitt-Grohé

where  $F(\cdot)$  is an operator defined over function spaces. It is not possible in general to derive  $c(s)$  analytically; thus we rely instead upon numerical approximation methods, which yield as output  $\hat{c}(s)$ .

To fix ideas, in sections 1 and 2 we work with a simple version of the optimal growth model, which is a modification of the RBC model introduced in chapter 3, section 1. The modification is that labor supply is fixed exogenously (and normalized to unity). Then in section 3 we work with a simple version of the single-asset pricing model introduced in chapter 3, section 3.

Briefly, the optimal growth model is given by

$$y = zk^\alpha \quad (5.4)$$

$$y = c + i \quad (5.5)$$

$$k' = i + (1 - \delta)k \quad (5.6)$$

$$\ln z' = \rho \ln z + \varepsilon', \quad (5.7)$$

where  $(y, k, c, i, z)$  denote time- $t$  values of output, capital, consumption, investment, and total factor productivity (TFP), and  $k'$  denotes the time- $(t+1)$  value of  $k$ , and so on. Taking  $(k_0, z_0)$  as given, and with  $\{\varepsilon_t\}$  following a known stochastic process, the representative household in this model seeks to maximize the expected value of its discounted stream of lifetime utility:

$$U = \max_{\{c_t\}} E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\phi}}{1-\phi}. \quad (5.8)$$

The collection of parameters  $\mu = (\alpha, \delta, \rho, \beta, \phi, \sigma_\varepsilon)$  represents capital's share of output, depreciation, the persistence of innovations to  $\ln z$ , the household's subjective discount factor, the coefficient of relative risk aversion, and the standard deviation of innovations to  $\ln z$ .<sup>1</sup>

The model can be collapsed into laws of motion for  $s = (z' k')$  and a first-order condition for choosing  $c$ :

$$k' = zk^\alpha + (1 - \delta)k - c \quad (5.9)$$

$$\ln z' = \rho \ln z + \varepsilon' \quad (5.10)$$

$$c^{-\phi} = \max \left\{ \frac{\beta E_t [(\varepsilon'^{-\phi})(\alpha z' k'^{\alpha-1} + 1 - \delta)]}{(zk^\alpha)^{-\phi}} \right\}. \quad (5.11)$$

In terms of the general notation introduced above, (5.9) and (5.10) combine to comprise (5.2);  $\varepsilon$  comprises  $v$ ; and (5.11) represents the functional equation to be solved in deriving  $c(s)$ . The  $\max\{\}$  operator is included in

<sup>1</sup> In section 2.1 of chapter 4, we introduced the optimal consumption model as a simplification of the RBC model under which  $z_t = 1 \forall t$ . In contrast, under the optimal growth

(5.11) to account for the possibility of a corner solution under which the household consumes all of its available time- $t$  output:  $c = y = zk^\alpha$ .

## 5.1 Projection Methods

### 5.1.1 Overview

Projection methods are a collection of approximation techniques for constructing  $\hat{c}(s)$  differentiated along two dimensions: the functional form over  $s$  chosen as the basis upon which  $\hat{c}(s)$  is constructed, and the criterion used to judge whether the approximation to (5.3)

$$F(\hat{c}(s)) = 0 \quad (5.12)$$

is satisfactory. Our presentation of these methods begins with the case in which  $s$  is one-dimensional; we then extend the discussion to the multi-dimensional case.<sup>2</sup>

Regarding the construction of  $\hat{c}(s)$ , the point of departure is the Weierstrass theorem, which states that any continuous function  $c(\cdot)$  can be approximated uniformly well over the range  $s \in [\underline{s}, \bar{s}]$  by a sequence of  $r$  polynomials  $p_r(s)$  (for details, e.g., see Judd, 1998). That is,

$$\lim_{r \rightarrow \infty} \max_{s \in [\underline{s}, \bar{s}]} |c(s) - p_r(s)| = 0. \quad (5.13)$$

A simple example of  $p_r(s)$  for the case in which  $s$  is one-dimensional is a linear combination of monomials in  $s$ ,

$$\begin{aligned} &\{s^0, s^1, \dots, s^r\}: \\ &p_r(s) = \sum_{i=0}^r \chi_i s^i. \end{aligned} \quad (5.14)$$

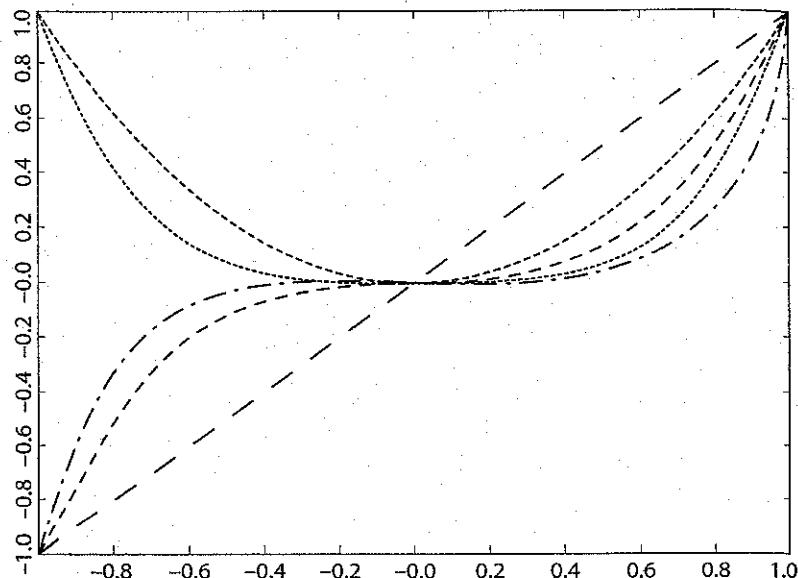
Let  $\chi = [\chi_0 \chi_1 \dots \chi_r]'$ . In using  $p_r(s)$  to construct  $\hat{c}(s, \chi)$ , the goal is to choose the parameters collected in the vector  $\chi$  to provide an optimal characterization of  $c(s)$ .

Although this example is illustrative, it turns out that approximations constructed using (5.14) as a functional form are not particularly effective

<sup>2</sup> GAUSS code for implementing projection methods that accompanies the text of Heer and Maussner (2005) is available at <http://www.wiwi.uni-augsburg.de/vwl/maussner/dgebook/download.html>.

Fortran code for implementing projection methods that accompanies the text of Judd (1998) is available at <http://bucky.stanford.edu/MWRJET/Judd.txt>.

Similar GAUSS, Matlab, and Fortran code developed by Wouter denHaan and Christian Haefke is available at <http://weber.ucsd.edu/~wdenhaan/peanew.html>.

Figure 5.1 Monomials in  $s$ 

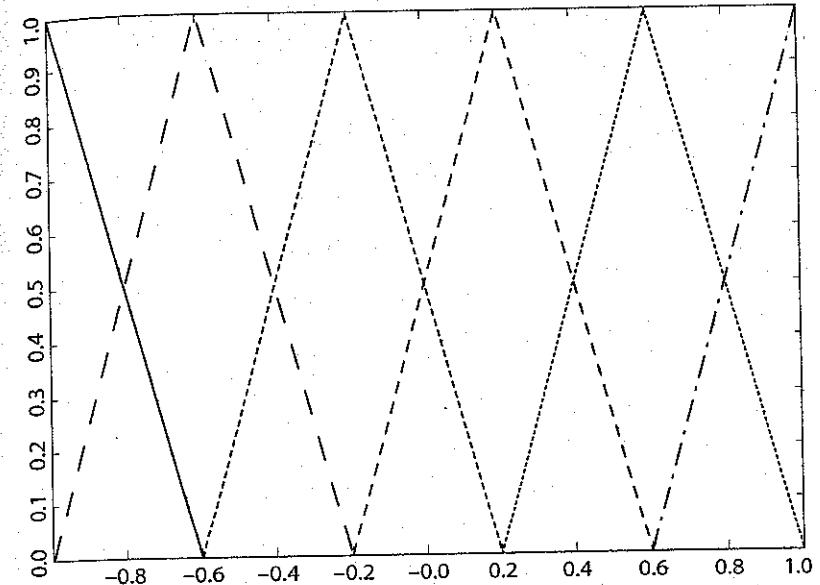
in practice, due in part to colinearity between elements of the resulting polynomial. This problem is illustrated in figure 5.1, which graphs the behavior of  $\{s^0, s^1, \dots, s^r\}$  over  $[-1, 1]$ ; adding terms beyond the power of 3 yields little value-added.

Two leading alternatives to the use of powers of  $s$  as a basis for constructing  $\hat{c}(s)$  involve the use of finite element methods and orthogonal polynomials. Details regarding the former are available, for example, from McGrattan (1996, 1999) and Aruoba, Fernandez-Villaverde, and Rubio-Ramirez (2006); details regarding the latter are available, for example, from Judd (1992).

### 5.1.2 Finite Element Methods

Finite element methods begin by establishing a grid over the sample space. In the one-dimensional case, the grid is in the form of nonoverlapping line segments. Denoting grid points as  $\{s^i\}_{i=1}^r$ , the  $i^{\text{th}}$  component of  $p_r(s)$ ,  $p_r^i(s)$ , is specified over a single element of  $s$  as

$$p_r^i(s) = \begin{cases} \frac{s-s^{i-1}}{s^i-s^{i-1}}, & s \in [s^{i-1}, s^i] \\ \frac{s^{i+1}-s}{s^{i+1}-s^i}, & s \in [s^i, s^{i+1}] \end{cases}. \quad (5.15)$$

Figure 5.2  $p_r^i(s)$ 

This forms a sequence of overlapping tent functions over the grid. Centered at the  $i^{\text{th}}$  grid point,  $p_r^i(s)$  increases linearly from zero to 1 as  $s$  ranges from  $s^{i-1}$  to  $s^i$ , and then decreases linearly towards zero as  $s$  ranges from  $s^i$  to  $s^{i+1}$ . The approximation  $\hat{c}(s, \chi)$  is then constructed using

$$\hat{c}(s, \chi) = \sum_{i=0}^r \chi_i p_r^i(s). \quad (5.16)$$

The behavior of  $p_r^i(s)$  over  $[-1, 1]$  is illustrated in figure 5.2.

### 5.1.3 Orthogonal Polynomials

The definition of an orthogonal polynomial is based on the inner product between two functions  $f_1$  and  $f_2$ , given the weighting function  $w$ :

$$\langle f_1, f_2 \rangle = \int_{\underline{s}}^{\bar{s}} f_1(s) f_2(s) w(s) ds. \quad (5.17)$$

The family of polynomials  $\{\varphi_r\}$  is defined to be mutually orthogonal with respect to  $w(s)$  if and only if  $\langle \varphi_r, \varphi_q \rangle = 0$  for  $r \neq q$ . One such family is the Chebyshev family, which is defined over  $s \in [-1, 1]$  and the weighting function  $w(s) = \sqrt{1 - s^2}$ .

polynomial for constructing  $\hat{c}(s, \chi)$ , the resulting approximation takes the form

$$\hat{c}(s, \chi) = \sum_{i=0}^r \chi_i T_i(s), \quad (5.18)$$

where

$$\begin{aligned} T_i(s) &= \cos(i \cos^{-1}(s)) \\ &= \cos(i\theta), \quad s = \cos(\theta). \end{aligned} \quad (5.19)$$

For a given value of  $s$ , the corresponding value of  $\hat{c}(s, \chi)$  can be calculated using the following algorithm due to Judd (1998). First, set

$$T_0(s) = 1, \quad T_1(s) = s. \quad (5.20)$$

Next, perform the recursion

$$T_{i+1}(s) = 2sT_i(s) - T_{i-1}(s), \quad i = 1, \dots, r. \quad (5.21)$$

Finally, collecting the  $T_i(s)$  terms in the  $r + 1 \times 1$  vector  $T(s)$ , calculate

$$\hat{c}(s, \chi) = T(s)' \chi. \quad (5.22)$$

The behavior of  $T_i(s)$  over  $[-1, 1]$  is illustrated in figure 5.3.

Although  $s$  must be constrained between  $[-1, 1]$  in working with Chebyshev polynomials, note that a simple transformation can be used to map a state variable defined over a general range  $[s, \bar{s}]$  into a variable defined over the range  $[-1, 1]$ . For example, for an element of  $s$  ranging  $\pm\omega$  units above and below the steady state value  $s^*$ , the transformation

$$\tilde{s} = \frac{s - s^*}{\omega} \quad (5.23)$$

yields the desired range. Hereafter, we take this transformation as granted in working with Chebyshev polynomials.

#### 5.1.4 Implementation

Given the use of either a finite element method or orthogonal polynomials for constructing  $\hat{c}(s, \chi)$ , the approximation of  $c(s)$  boils down to the task of choosing  $\chi$  to "nearly" satisfy

$$F(\hat{c}(s, \chi)) = 0. \quad (5.24)$$

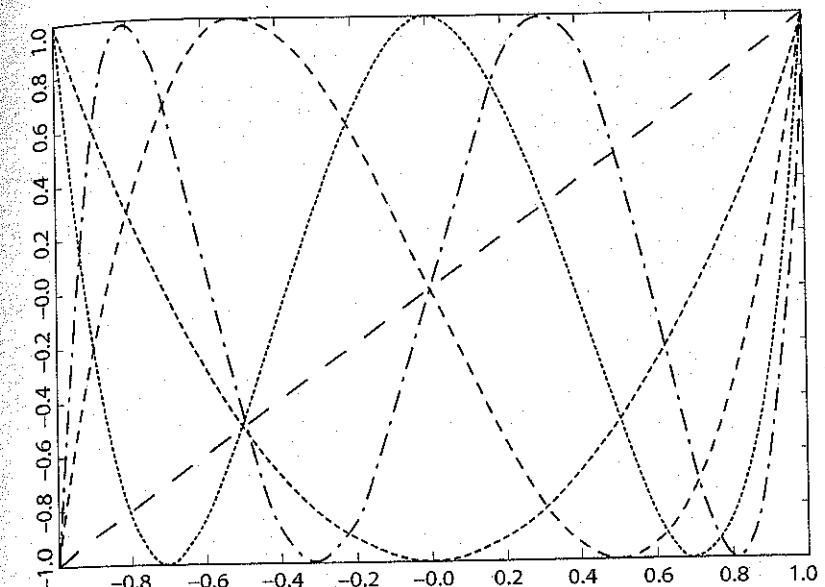


Figure 5.3  $T_i(s)$

In the case of the optimal growth model (with  $a \equiv 1$  in the one-dimensional case),  $F(\hat{c}(s, \chi))$  is given by

$$\begin{aligned} F(\hat{c}(k, \chi)) &= \hat{c}(k)^{-\phi} \\ &- \max \left\{ \beta E_t \left[ \frac{(\hat{c}(zk^\alpha + (1-\delta)k - \hat{c}(k))^{-\phi})}{(\alpha z'(zk^\alpha + (1-\delta)k - \hat{c}(k))^{\alpha-1} + 1 - \delta)} \right] \right\}, \end{aligned} \quad (5.25)$$

which emerges from the replacement of  $k'$  with  $zk^\alpha + (1-\delta)k - \hat{c}(k)$  in (5.11).

Alternative choices of the criterion upon which judgments regarding the quality of approximation are based establish the second dimension along which projection methods are differentiated. Returning to the use of inner products, the approximation we seek is defined by the parameter vector  $\chi$  that minimizes

$$\langle F(\hat{c}(s, \chi)), f(s) \rangle = \int_{\underline{s}}^{\bar{s}} F(\hat{c}(s, \chi)) f(s) w(s) ds. \quad (5.26)$$

Alternative specifications for  $f(s)$  and  $w(s)$  differentiate projection methods along this dimension.

The choice of  $F(\hat{c}(s, \chi))$  for  $f(s)$  casts (5.26) in the form of a weighted-least-squares problem. Choosing  $N$  distinct values for  $s$  ranging over  $[\underline{s}, \bar{s}]$ , (5.26) can be approximated using

$$\sum_{i=1}^N F(\hat{c}(s_i, \chi))^2 w(s_i), \quad (5.27)$$

which is a single-valued objective function that can be optimized using, for example, the derivative-based methods discussed in chapter 13. Regarding the specification of  $w(s)$ , a tent function specified over the range  $[\underline{s}, \bar{s}]$ , and peaking at the steady state,  $s^*$  could be used to assign relatively heavy weights to values of  $s$  in the neighborhood of the model's steady state. An alternative specification that serves this purpose is given by

$$w(s) = 1 - \left( \left| \frac{s - s^*}{s^*} \right| \right)^k. \quad (5.28)$$

Under the finite element method, a common choice for  $f(s)$  (along with  $w(s) = 1$ ) is the sequence of basis functions  $p_r^i(s)$ ,  $i = 1, \dots, r$ , as defined in (5.15). In this case, known as the Galerkin method, a single-valued objective function is replaced by a system of  $r$  equations, which are to be solved by the choice of the  $r$ -dimensional vector of coefficients  $\chi$ :

$$(F(\hat{c}(s, \chi)), p_r^i(s)) = \int_{\underline{s}}^{\bar{s}} F(\hat{c}(s, \chi)) p_r^i(s) ds = 0, \quad i = 1, \dots, r. \quad (5.29)$$

The integrals in (5.29) can again be approximated using a sum over a range of distinct values chosen for  $s$ . Also, derivative-based methods akin to those available for solving single-valued objective functions are available for solving nonlinear systems of the form (5.29).<sup>3</sup>

Finally, an additional choice for  $f(s)$  (once again, coupled with  $w(s) = 1$ ) is the indicator function  $\delta(s - s_i)$ ,  $i = 1, \dots, r$ , with

$$\delta(s - s_i) = \begin{cases} 1, & s = s_i \\ 0, & s \neq s_i \end{cases}. \quad (5.30)$$

Under  $\delta(s - s_i)$ , the functional equation  $F(\cdot) = 0$  is restricted to hold exactly at  $r$  fixed points. Approximations based on  $\delta(s - s_i)$  are referenced as having been obtained using the collocation method. Once again, the resulting system can be solved using derivative-based methods.

As Judd (1998) notes, the collocation method is particularly attractive in using Chebyshev polynomials for constructing  $\hat{c}(s, \chi)$ . This is due to the

### 5.1 Projection Methods

Chebyshev interpolation theorem, which provides a means of optimizing the  $r$  choices of  $s$  used to construct  $\delta(s - s_i)$ . Specifically, let  $\{\tilde{s}_i\}_{i=1}^r$  denote the roots of the  $r^{\text{th}}$ -order component  $T_r(s)$ . Then if  $F(\hat{c}(\tilde{s}_i, \chi)) = 0$  for  $i = 1, \dots, r$  and  $F(\hat{c}(s, \chi))$  is continuous,  $F(\hat{c}(s, \chi))$  will be close to 0 over the entire range  $[-1, 1]$ .

Regarding the roots of  $T_r(s)$ , recall that

$$\cos(\hat{\theta}_j) = 0, \quad \hat{\theta}_j = (2j - 1) \frac{\pi}{2}, \quad j = 1, 2, \dots, r. \quad (5.31)$$

(i.e.,  $\cos(\hat{\theta}) = 0$  for all odd multiples of  $\frac{\pi}{2}$ ). Thus with  $s$  defined as  $s = \cos(\theta)$  in (5.19), the  $r$  roots of  $T_r(s)$  are given by

$$\tilde{s}_j = \cos\left(\frac{(2j - 1)\pi}{r}\right), \quad j = 1, 2, \dots, r. \quad (5.32)$$

A simple means of assessing the quality of a given approximation  $\hat{c}(s, \chi)$  is to assess the resulting residual function  $F(\hat{c}(s, \chi))$  over a range of values chosen for  $s$ . Plots of the residual function, sums of squared residual errors, and so on, are effective diagnostic tools. In using collocation methods, choices for  $s$  other than those used to construct  $\hat{c}(s, \chi)$  are advisable.

We conclude this general discussion with a note regarding starting values. Depending upon the particular problem and approximation method, the construction of accurate approximations to  $c(s)$  can depend critically upon good choices of starting values for the vector of parameters  $\chi$ . The identification of good starting values is particularly important in using these solution methods as a basis for conducting empirical work, because empirical analyses of the model in question typically require that the model be solved repeatedly for alternative specifications of the structural parameters  $\mu$ . The ability to obtain accurate approximations quickly and reliably is therefore critical in this context.

Unfortunately, the identification of effective starting values is highly problem-specific. However, a good general approach is to select starting values that mimic the behavior of  $c(s)$  estimated using a linearized approximation of the model in the neighborhood of the steady state. For example, the first two elements of an  $r^{\text{th}}$ -order Chebyshev polynomial are  $[1, s]$ ; thus a good candidate choice of starting values can be obtained by setting initial values of  $\chi_0$  and  $\chi_1$  in a way that captures the estimated linear relationship between  $c$  and  $s$  in the neighborhood of  $c^*$  and  $s^*$ , and setting all other initial values of  $\chi_i$  to zero. Further elaboration of this example is provided in section 1.6.

<sup>3</sup> GAUSS's n1sys package is available for this purpose.

### 5.1.5 Extension to the $\ell$ -dimensional Case

The extension of projection methods to the  $\ell$ -dimensional case is accomplished through the construction of multidimensional functions specified over  $s = (s_1 \ s_2 \ \dots \ s_\ell)'$ . One approach to this construction involves the use of tensor products of univariate functions specified over the individual elements of  $s$ . Let  $p_{r_j}(s_j)$  denote an  $r_j^{\text{th}}$ -order sequence of polynomials specified for the  $j^{\text{th}}$  element of  $s$ . Then the  $\ell$ -dimensional tensor product of  $p_{r_j}(s_j)$ ,  $j = 1, \dots, \ell$ , is given by

$$P = \prod_{j=1}^{\ell} p_{r_j}(s_j). \quad (5.33)$$

For example, with  $\ell = 2$ ,  $r_1 = 2$ ,  $r_2 = 3$ , and  $p_{r_j}(s_j)$  representing Chebychev polynomials, we have

$$P = (1 + s_1 + T_2(s_1))(1 + s_2 + T_2(s_2) + T_3(s_2)), \quad (5.34)$$

where  $T_i(s_j)$  is the  $i^{\text{th}}$ -order term corresponding with the  $j^{\text{th}}$  element of  $s$ , and is defined as in (5.21).

Using a given tensor product, the approximation  $\hat{c}(s, \chi)$  is constructed using

$$\hat{c}(s, \chi) = \sum_{i_1=1}^{r_1} \sum_{i_2=1}^{r_2} \dots \sum_{i_\ell=1}^{r_\ell} \chi_{i_1 i_2 \dots i_\ell} P_{i_1 i_2 \dots i_\ell}(s_1, s_2, \dots, s_\ell), \quad (5.35)$$

where

$$P_{i_1 i_2 \dots i_\ell}(s_1, s_2, \dots, s_\ell) = p_{i_1}(s_1) p_{i_2}(s_2) \dots p_{i_\ell}(s_\ell). \quad (5.36)$$

In the example of (5.34), the elements  $P_{i_1 i_2 \dots i_\ell}(s_1, s_2, \dots, s_\ell)$  can be constructed by forming the vectors

$$\begin{aligned} T^1 &= (1 \ s_1 \ T_2(s_1))', \\ T^2 &= (1 \ s_2 \ T_2(s_2) \ T_3(s_2))', \end{aligned}$$

and retrieving the elements of the  $3 \times 4$  matrix  $T = T^1 T^2'$ .

Regarding the estimation of  $\chi_{i_1 i_2 \dots i_\ell}$ , the procedures described above for the univariate case extend naturally. For example, the collocation method remains an attractive means of implementing Chebyshev polynomials for constructing  $\hat{c}(s, \chi)$ , because the Chebyshev interpolation theorem continues to apply. Specifically, letting  $\{\tilde{s}_{j,i}\}_{i=1}^{r_j}$  denote the roots of the  $r_j^{\text{th}}$ -order component  $T_{r_j}(s_j)$ , if  $F(\hat{c}(\tilde{s}_{j,i}, \chi)) = 0$  for  $i = 1, \dots, r_j$ ,  $j = 1, \dots, \ell$ , and

### 5.1 Projection Methods

$F(\hat{c}(s, \chi))$  is continuous,  $F(\hat{c}(s, \chi))$  will be close to 0 over the entire  $\ell$ -dimensional hypercube ranging over  $[-1, 1]$  in each dimension.

Although conceptually straightforward, extensions to the  $\ell$ -dimensional case involving tensor products can be complicated by a computational challenge when  $\ell$  is large: as  $\ell$  increases, the number of elements  $\chi_{i_1 i_2 \dots i_\ell}$  that must be estimated to construct the approximation  $\hat{c}(s, \chi)$  increases exponentially. This phenomenon is a manifestation of the curse of dimensionality. One means of combating it in this context is to keep the polynomial orders  $r_1, r_2, \dots, r_\ell$  relatively modest. An alternative means is to replace tensor products with complete polynomials.

The complete set of polynomials of degree  $k$  is given by the collection of terms that appear in the  $k^{\text{th}}$ -order Taylor Series approximation of  $c(s)$  about an arbitrary value  $s^0$ . Recall that this approximation is given by

$$\begin{aligned} c(s) \approx c(s^0) &+ \sum_{i=1}^{\ell} \frac{\partial c(s)}{\partial s_i}(s^0)(s_i - s_i^0) \\ &+ \frac{1}{2!} \sum_{i_1=1}^{\ell} \sum_{i_2=1}^{\ell} \frac{\partial^2 c(s)}{\partial s_{i_1} \partial s_{i_2}}(s^0)(s_{i_1} - s_{i_1}^0)(s_{i_2} - s_{i_2}^0) \\ &\dots \\ &+ \frac{1}{k!} \sum_{i_1=1}^{\ell} \dots \sum_{i_k=1}^{\ell} \frac{\partial^k c(s)}{\partial s_{i_1} \dots \partial s_{i_k}}(s^0)(s_{i_1} - s_{i_1}^0) \dots (s_{i_k} - s_{i_k}^0). \end{aligned}$$

Thus for  $k = 1$ , the complete set of polynomials for the  $\ell$ -dimensional case is given by

$$P_1^\ell = \{1, s_1, s_2, \dots, s_\ell\}.$$

For  $k = 2$  the set expands to

$$P_2^\ell = P_1^\ell \cup \{s_1^2, s_2^2, \dots, s_\ell^2, s_1 s_2, s_1 s_3, \dots, s_1 s_\ell, s_2 s_3, \dots, s_{\ell-1} s_\ell\},$$

and so on. So in the two-dimensional case, whereas the tensor product of third-order polynomials involves 16 terms, the complete set of third-degree polynomials involves only 10 terms.

An approximation of  $\hat{c}(s, \chi)$  based on the complete set of  $k$ -degree polynomials can be obtained, for example, using a weighted-least-squares procedure, as in (5.27).

### 5.1.6 Application to the Optimal Growth Model

We now demonstrate the application of these methods, working in the context of the optimal growth model. This is done for three alternative

assumptions regarding the behavior of  $z$ . We begin by working with a nonstochastic specification under which  $z$  is fixed at unity. Next, we specify  $z$  as following a two-state Markov process (defined below). Finally, we specify  $\ln z$  as following an AR(1) process, as in (5.7). Under the first two specifications the remaining parameters are fixed at  $(\alpha, \beta, \delta, \phi) = (0.33, 0.96, 0.1, 2)$ ; under the latter specification the parameters are given by  $(\alpha, \beta, \delta, \phi) = (0.33, 0.96, 1, 1)$ . We chose the latter specification because it admits an exact solution to the policy function. Specifically, the policy function in this case is given by

$$c = (1 - \alpha\beta)zk^\alpha, \quad (5.37)$$

a result we invite the reader to verify. Thus under this parameterization, we have an exact benchmark available for judging the accuracy of our approximation algorithm.<sup>4</sup>

Beginning with the nonstochastic specification, approximation was achieved using as an objective function the weighted-least-squares function (5.27), with  $w(k)$  specified as indicated in (5.28), with  $\kappa = 0.1$ . We used an 11<sup>th</sup>-order Chebyshev polynomial as a basis function for constructing  $\tilde{c}(k, \chi)$ . Recall that since Chebyshev polynomials are defined over  $[-1, 1]$ , we worked with the normalization

$$\tilde{k} = \frac{k - k^*}{\omega_k}. \quad (5.38)$$

Finally, we used the derivative-based optimization algorithm of Broyden, Fletcher, Goldfarb, and Shanno (BFGS) to minimize (5.27) with respect to  $\chi$  (implemented using the GAUSS package `optimum`). We used trial-and-error to specify starting values for  $\chi$  in this case, and we encourage the reader to investigate the delicate sensitivity of the estimated policy function we obtained to alternative specifications, along with the accuracy of the corresponding estimates (using `determa.prg`). The policy function we derived in this manner is illustrated in figure 5.4. The range over which the policy function is plotted is  $[k^* - \omega_k, k^* + \omega_k]$ , which is partitioned into 101 equally spaced intervals using  $\omega_k = 0.25 * k^*$  (i.e., the range is 25% above and below  $k^*$ ).

The top panel of the figure depicts the policy function itself; the bottom panel depicts its slope. The slope declines from 0.156 to 0.128 over the range of  $k$ , a decrease of nearly 22%. In a  $\pm 5\%$  range around the steady state,

<sup>4</sup>The GAUSS programs `determa.prg`, `markova.prg`, and `ara.prg` were used to perform the calculations described here. In addition, the collection of procedures in `stogro.src` and `stogrowm_nonlin_exact.src` served as inputs to these programs. The former solves the model under an arbitrary parameterization of the model; the latter is specialized to the case in which  $\delta = 1, \phi = 1$ .

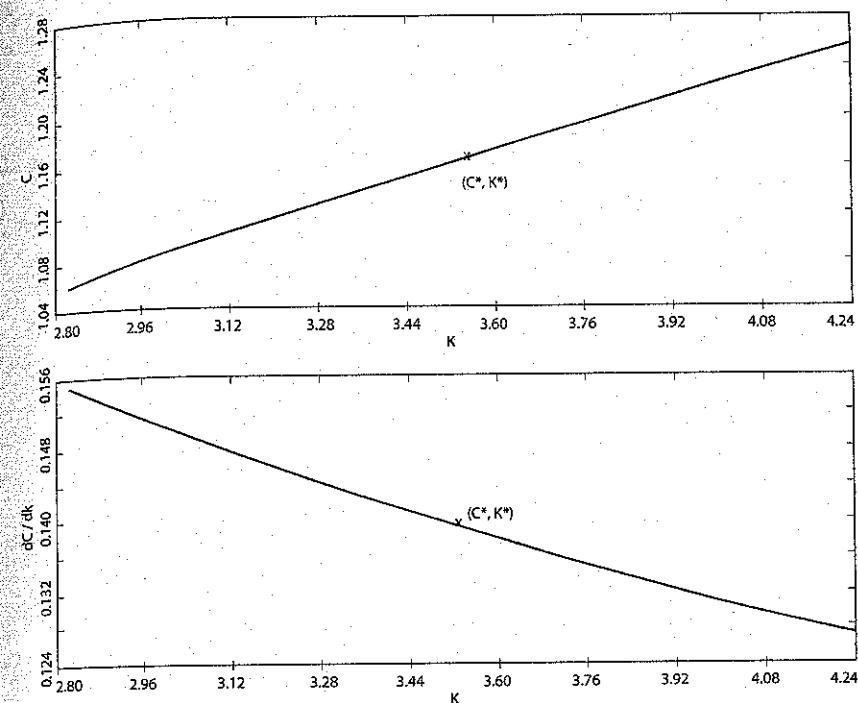


Figure 5.4 Policy Function for the Nonstochastic Optimal Growth Model

the percentage change in the slope is roughly 6%. These slope calculations are of interest for comparison with a linear approximation of the policy function, under which the slope is constant over the range of  $k$ .

Turning to the stochastic specification of the model, two critical differences arise: the functional equation (5.25) now requires the calculation of a conditional expectation, and the state vector  $s$  is now two-dimensional.

To describe how the first issue was addressed, we pause to characterize the two-state Markov process specified for  $z$ . (Further details regarding Markov processes are provided in chapter 8, section 5.3.) This was constructed by discretizing the state space over  $z$  into one of two outcomes:  $z = zL = 0.95$  in state  $L$  (low productivity) and  $z = zH = 1.05$  in state  $H$  (high productivity). Given this discretization, the stochastic behavior of  $z$ , which is manifested by movements between low and high states, was modeled by specifying that movements across states are dictated by the following matrix of transition probabilities:

$$P = \begin{bmatrix} p_{LL} & 1 - p_{HH} \\ 1 - p_{LL} & p_{HH} \end{bmatrix}. \quad (5.39)$$

Here,  $p_{LL}$  denotes the probability of  $z$  remaining in the low state given that it began in the low state, and so on. For a general  $n$ -state process, the transition matrix  $P$  is  $n \times n$ , with entries contained in the  $i^{\text{th}}$  column indicating transition probabilities given that the process is initially in state  $i$ ; that is, the (row- $j$ , column- $i$ ) element of  $P$  indicates the probability of ending up in state  $j$  given that the current state is  $i$ .

Following Hamilton (1994), we note that there is a close relationship between a two-state Markov process and an AR(1) specification. To see this, let  $\xi_t$  be a  $2 \times 1$  vector with a 1 in the first row and a zero in the second row given that  $z_t = zL$ , and a 0 in the first row and a 1 in the second row given that  $z_t = zH$ . Then the conditional expectation of  $\xi_{t+1}$  given  $\xi_t$  is given by

$$E_t \xi_{t+1} = P \xi_t, \quad (5.40)$$

and thus

$$\xi_{t+1} = P \xi_t + \varsigma_{t+1}, \quad (5.41)$$

where  $\varsigma_{t+1} = \xi_{t+1} - E_t \xi_t$  is a mean-zero unforecastable stochastic process.

Note that (5.41) is in the form of a first-order VAR specified for  $\xi_{t+1}$ . Further, noting that the second element of  $\xi_t$ , denoted as  $\xi_{2t}$ , can be written as  $\xi_{2t} = 1 - \xi_{1t}$ , (5.41) can be rewritten as

$$\begin{bmatrix} \xi_{1t+1} \\ \xi_{2t+1} \end{bmatrix} = \begin{bmatrix} p_{LL} & 1 - p_{HH} \\ 1 - p_{LL} & p_{HH} \end{bmatrix} \begin{bmatrix} \xi_{1t} \\ \xi_{2t} \end{bmatrix} + \begin{bmatrix} \varsigma_{1t+1} \\ \varsigma_{2t+1} \end{bmatrix}. \quad (5.42)$$

Rearranging the first row of (5.42), we obtain

$$\xi_{1t+1} = (1 - p_{22}) + (-1 + p_{11} + p_{22}) \xi_{1t} + \varsigma_{1t+1}, \quad (5.43)$$

which is in the form of an AR(1) representation for  $\xi_{1t+1}$ . With  $p_{11} = p_{22}$ , the selection of  $\rho = 0.8$  translates into probability of transition between states of 0.1. The results presented below were obtained using this specification. For a generalization to the use of Markov processes for approximating AR( $p$ ) models for univariate time series, and VAR models for multivariate time series; see Tauchen (1986a).

Returning to the first issue then, under the Markov specification the conditional expectation  $E_t z'$  is the weighted average of  $zL$  and  $zH$ , with relative weights given by [0.1, 0.9] given the initial value of  $a$  in the low state, and [0.9, 0.1] given the initial value of  $z$  in the high state.

Under the AR(1) specification for  $\ln z'$ , the situation is slightly more complicated, because the conditional expectation in (5.11) cannot be calculated analytically. Calculation in this case is accomplished via numerical approximation, using what are known as quadrature methods. To describe how this is done, we begin by defining the normalization  $\tilde{z}' = \frac{\ln z'}{\omega_z} = \alpha \tilde{z} + \tilde{s}$  where  $\tilde{z} = \frac{z - \bar{z}}{\sigma_z}$  which is restricted to the range [-1, 1] with the policy

### 5.1 Projection Methods

function for  $c$  defined over  $(\tilde{z}, \tilde{k})$ , and in this case with  $\phi = \delta = 1$ , the expectation we must approximate is given by

$$\begin{aligned} I &= \beta E_t \left[ \left( \frac{1}{\hat{c}(\tilde{z}', \tilde{k}')} \right) \alpha k'^{\alpha-1} e^{\rho \ln z + \varepsilon'} \right] \\ &= \frac{1}{\sigma_\varepsilon \sqrt{2\pi}} \int_{-\infty}^{\infty} f(\varepsilon' | z, k) e^{-\frac{\varepsilon'^2}{2\sigma_\varepsilon^2}} d\varepsilon' \\ &= \frac{\omega_z}{\sigma_\varepsilon \sqrt{2\pi}} \int_{-\infty}^{\infty} f(\tilde{\varepsilon}' | z, k) e^{-\frac{(\omega_z \tilde{\varepsilon}')^2}{2\sigma_\varepsilon^2}} d\tilde{\varepsilon}'. \end{aligned} \quad (5.44)$$

The second equality is obtained by defining

$$f(\varepsilon' | z, k) = \beta \left( \frac{1}{\hat{c}(\tilde{z}', \tilde{k}')} \right) \alpha k'^{\alpha-1} e^{\rho \ln z + \varepsilon'}$$

and replacing  $E_t$  with the integral over the  $N(0, \sigma_\varepsilon^2)$  density for  $\varepsilon'$  it entails. The third equality is obtained by implementing a change in variables from  $\varepsilon$  to  $\tilde{\varepsilon}$ . To see why, recall that for some general transformation  $\varepsilon' = g(\tilde{\varepsilon}')$ , the change-of-variables formula is given by

$$\int_a^b f(\varepsilon') d\varepsilon = \int_{g^{-1}(a)}^{g^{-1}(b)} f(g(\tilde{\varepsilon}')) g'(\tilde{\varepsilon}') d\tilde{\varepsilon}'. \quad (5.45)$$

With this notation set, note that the integral in (5.44) is a specific example of a generic integral of the form

$$\int_a^b f(x) dx.$$

Quadrature methods comprise a wide class of numerical tools available for approximating many alternative specific examples of such an integral (for an overview, e.g., see Judd, 1998). Gaussian quadrature methods comprise a specific subclass of methods that use the orthogonal approach to functional approximation to evaluate such integrals. Here we focus on the case in which  $x$  is univariate; for an extension to the multivariate case, see Heiss and Winschel (2008).

Gaussian quadrature approximations take the form

$$\int_a^b f(x) dx \approx \sum_{i=1}^n w_i f(x_i), \quad (5.46)$$

where the nodes  $x_i$  and corresponding weights  $w_i$  are chosen to provide accurate and efficient approximations. Specifically, the nodes and weights are chosen so that if  $f(\cdot)$  were in fact a polynomial of degree  $2n - 1$ , then the approximation (5.46) will be exact given the use of  $n$  nodes and weights

Gauss-Hermite nodes and weights are tailored specifically for cases in which integrals are of the form

$$\int_{-\infty}^{\infty} f(x) e^{-x^2} dx,$$

which arises naturally in working with Normal random variables, as is the case in this application. In particular, defining

$$\tilde{\varepsilon}' = \frac{\sigma_\varepsilon \sqrt{2}}{\omega_z} x,$$

use of the change-of-variables formula (5.45) enables the approximation of the conditional expectation (5.44) via

$$\begin{aligned} I &= \frac{\omega_z}{\sigma_\varepsilon \sqrt{2\pi}} \int_{-\infty}^{\infty} f(\tilde{\varepsilon}' | z, k) e^{-\frac{(\omega_z \tilde{\varepsilon}')^2}{2\sigma_\varepsilon^2}} d\tilde{\varepsilon}' \\ &\approx \frac{1}{\sqrt{\pi}} \sum_{i=1}^n w_i (f(\sigma_\varepsilon \sqrt{2}) x_i) \end{aligned} \quad (5.47)$$

using Gauss-Hermite weights and nodes.<sup>5</sup>

Turning now to the latter issue, with the state vector being two-dimensional, we approximated the policy function using an orthogonal collocation scheme applied to a tensor product of Chebyshev polynomials. Because under the Markov specification  $z$  can take on only one of two values, we specified a second-order polynomial along the  $z$  dimension, and a fifth-order polynomial along the  $k$  dimension. Under the AR(1) specification we specified a fourth-order polynomial along the  $\ln z$  dimension, but higher-order specifications are feasible in this case.

Regarding the range chosen for  $\ln z$  in the approximation, care must be taken to ensure that in the Gauss-Hermite approximation (5.47), the set of nodes  $x_i$  is chosen subject to the restriction that the corresponding values of  $\tilde{\varepsilon}'_i$  remain bounded between  $[-1, 1]$ , because we are using Chebyshev polynomials to construct  $\hat{c}(\tilde{\varepsilon}', \tilde{k}')$ . To achieve this here we follow Judd (1992) by constructing the upper and lower bounds for  $z$  as

$$(\underline{z}, \bar{z}) \pm \frac{1}{1 - \rho} 3\sigma_\varepsilon;$$

that is, the bounds are set as the values of  $\ln z$  that would be attained if every single realized  $\varepsilon$  innovation were either three times larger or smaller

<sup>5</sup> A table of Gauss-Hermite weights and nodes is available at the textbook website.

### 5.1 Projection Methods

than the expected value of 0. With  $\rho = 0.8$ , this range is  $\pm 1.5\sigma_\varepsilon$ , enabling the specification of a large number of Gauss-Hermite nodes that lie within the required bounds. We used seven nodes to obtain the results reported below.

Given these choices for ranges and nodes, the corresponding parameterization  $\chi$  was obtained as the solution to the system of equations

$$F(\hat{c}(\tilde{s}_j, \chi)) = 0, \quad i = 1, \dots, r_j, \quad j = z, k. \quad (5.48)$$

Under the AR(1) specification,  $\{\tilde{s}_j\}_{j=1}^{r_j}$  denotes the roots of the  $r_j^{\text{th}}$ -order component  $T_{r_j}(s_j)$ ,  $j = z, k$ . Under the Markov specification,  $\tilde{s}_j$  was constructed using the roots of the  $r_j^{\text{th}}$ -order component  $T_{r_j}(s_j)$  along the  $k$  dimension, and using  $zL$  and  $zH$  along the  $z$  dimension. (In this case, we used GAUSS's nlsys routine to solve the corresponding systems of equations.)

As with the deterministic specification, the ability to obtain accurate approximations again hinges critically upon the identification of reliable starting values for  $\chi$ . But encouragingly, the use of starting values fashioned from the log-linear approximation of the model around its steady state turned out to provide reliable values that quickly and consistently led to accurate approximations for a wide range of alternative parameterizations of the model.

To describe the construction of starting values, let  $\sigma_z$  and  $\sigma_k$  denote the elasticity of  $c$  with respect to  $z$  and  $k$ , respectively, calculated using the log-linear approximation of the model. (Here we are supposing that the policy function is defined over  $z$  and  $k$ ; the case for which the policy function is instead defined over  $\ln z$  and  $\ln k$  proceeds analogously.) In terms of levels of variables, in the neighborhood of the steady state an approximation of  $c$  can be constructed using  $\sigma_z$  and  $\sigma_k$  as

$$\begin{aligned} c &\approx c^* + \frac{\partial c}{\partial z}(z^*, k^*)(z - z^*) + \frac{\partial c}{\partial k}(z^*, k^*)(k - k^*) \\ &\quad + \frac{1}{2} \frac{\partial^2 c}{\partial z \partial k}(z^*, k^*)(z - z^*)(k - k^*) \\ &= c^* + \frac{c^*}{z^*} \sigma_z(z - z^*) + \frac{c^*}{k^*} \sigma_k(k - k^*) \\ &\quad + \frac{1}{2} \left( \frac{c^*}{z^*} \sigma_z \right) \left( \frac{c^*}{k^*} \sigma_k \right) (z - z^*)(k - k^*). \end{aligned} \quad (5.49)$$

$$(5.50)$$

Of course, additional terms could also be incorporated in this approximation. Given the use of a tensor product representation, the corresponding approximation of  $\hat{c}(z, k, \chi)$  we seek is of the form

$$\begin{aligned}\hat{c}(a, k, \chi) \approx & \chi_{11} + \chi_{12} \left( \frac{z - z^*}{\omega_z} \right) + \chi_{21} \left( \frac{k - k^*}{\omega_k} \right) \\ & + \chi_{22} \left( \frac{z - z^*}{\omega_z} \right) \left( \frac{k - k^*}{\omega_k} \right) + \dots\end{aligned}\quad (5.51)$$

Matching terms across representations yields the suggested starting values

$$\begin{aligned}\chi_{11} = c^*, \quad \chi_{12} = \sigma_z \omega_z \frac{c^*}{z^*}, \quad \chi_{21} = \sigma_k \omega_k \frac{c^*}{k^*}, \\ \chi_{22} = \frac{1}{2} \left( \sigma_z \omega_z \frac{c^*}{z^*} \right) \left( \sigma_k \omega_k \frac{c^*}{k^*} \right).\end{aligned}\quad (5.52)$$

The remaining values of  $\chi$  were initiated at 0. As noted, use of these starting values turned out to yield an effective means of approximating the policy function for both the Markov and AR specifications for  $\ln z$ . Indeed, nlsys typically generated a model solution using only two or three quasi-Newton iterations for a wide range of alternative parameterizations of the model. We invite the reader to verify this performance (working with markova.prg and ara.prg).

The policy functions obtained using the Markov specification for  $z$  are depicted in figure 5.5. The upper (lower) curve depicts the policy function given an initial value of  $z$  in the high (low) state. Note that the functions closely parallel each other, and shift roughly symmetrically above and below the policy function obtained under the nonstochastic specification. As with figure 5.4, the range for  $k$  over which the policy function is plotted is  $[k^* - \omega_k, k^* + \omega_k]$ , which is partitioned into 101 equally spaced intervals using  $\omega_k = 0.25k^*$ .

The policy function obtained using the AR(1) specification follows the pattern depicted in figure 5.5 closely, and thus is not illustrated here. Instead, we conclude this application section with a note regarding the accuracy of the approximation we obtained in this case. Recall that under the AR(1) specification, we are able to compare our approximation with the exact solution, which is given by (5.37). Over a set of values ranging over  $k \in [k^* - \omega_k, k^* + \omega_k]$ ,  $\omega_k = 0.25$  and  $z \in [-\omega_z, \omega_z]$ ,  $\omega_z = \frac{1}{1-\rho} 3\sigma_\epsilon$ , the maximum approximation error we obtained amounts to a difference in consumption values of only 0.0003% of the steady state value of consumption.

### Exercise 5.1

Using a projection method, construct policy functions for the choice of both consumption and leisure for the RBC model introduced in chapter 3, section 1. Use the GAUSS programs available at the textbook website for guidance.

[Hint: Note from the intratemporal optimality condition that determines the labor/leisure trade-off that the policy function for determining  $c$  implicitly serves to determine the policy function over labor supply. Exploit this fact in deriving policy functions.]

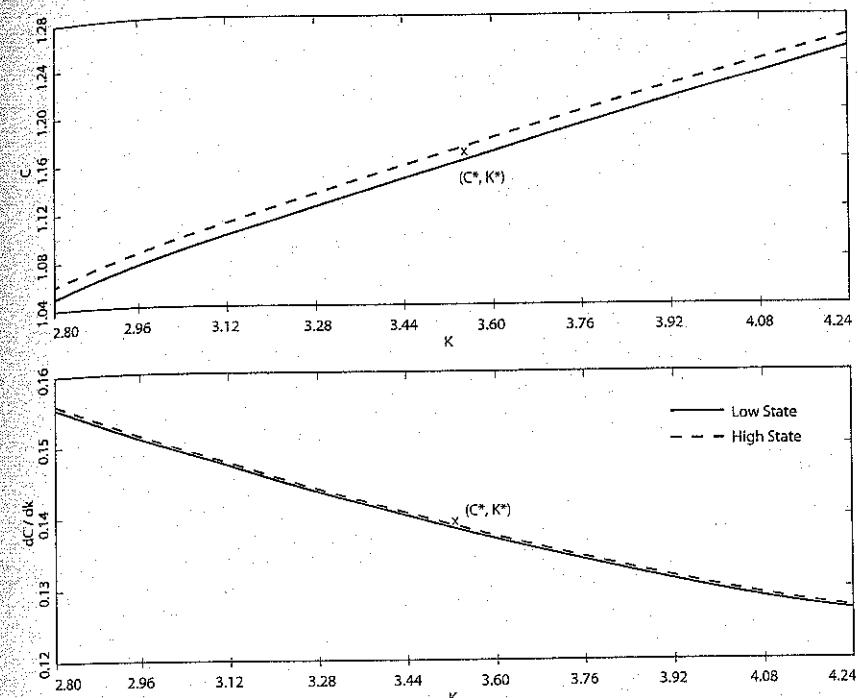


Figure 5.5 Policy Function for the Stochastic Optimal Growth Model

## 5.2 Iteration Techniques: Value-Function and Policy-Function Iterations

### 5.2.1 Dynamic Programming

Value-function and policy-function iterations are alternative methods for constructing  $\hat{c}(s)$  that are closely linked to the dynamic programming approach to solving sequential optimization problems. We begin here by providing a brief overview of dynamic programming. Far more detailed textbook treatments are provided by, for example, Sargent (1987b) and

Stokey and Lucas (1989). In addition, Judd (1998), Adda and Cooper (2003), Ljungqvist and Sargent (2004), and Heer and Maussner (2005) describe links between dynamic programming and the iterative methods for approximating  $\tilde{c}(s)$  described here.<sup>6</sup>

To begin, we establish explicitly the optimization problem that gives rise to the policy function  $c(s)$ . We seek to choose an infinite sequence  $\{c_t\}$  that maximizes

$$E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, s_t),$$

subject to the law of motion for the state

$$s_{t+1} = g(s_t, c_t, v_{t+1}), \quad (5.53)$$

taking  $s_0$  as given. The parameter  $0 < \beta < 1$  is the usual subjective discount factor, and  $E_0$  is the expectations operator conditional on information available at time 0 (i.e.,  $s_0$ ). Moreover, in order for the problem to be well defined,  $u(c_t, s_t)$  must be concave, and the set

$$\{(s_{t+1}, s_t) : s_{t+1} \leq g(s_t, c_t, v_t)\}$$

must be convex and compact.

The policy function is a contingency plan for choosing the optimal value of  $c_t$  given the realization of the state  $s_t$ :  $c_t = c(s_t)$ . Using this rule to substitute for  $c_t$  in (5.53) yields the alternative characterization of the law of motion for the state we have used throughout the text:  $s_t = f(s_{t-1}, v_t)$ .

Let  $V(s_0)$  denote the value of behaving optimally over the lifetime of the problem given the initial endowment of the state  $s_0$ . This as yet unknown function is linked with the policy function  $c(s_0)$  via Bellman's equation:

$$V(s_0) = \max_{c_0} \{u(c_0, s_0) + \beta E_0 V(s_1)\}, \quad (5.54)$$

subject to (5.53). In words, Bellman's equation states that the value of behaving optimally over the lifetime of the problem is equal to the value of behaving optimally in the initial period, plus the expected discounted value of behaving optimally thereafter.

Substituting for  $s_1$  in (5.54) using (5.53), we obtain

$$V(s_0) = \max_{c_0} \{u(c_0, s_0) + \beta E_0 V(g(s_0, c_0, v_1))\}. \quad (5.55)$$

<sup>6</sup> A good source of code for implementing these methods is <http://ideas.repec.org/s/dge/qmrbcd.html>.

## 5.2 Iteration Techniques

Also, substituting for  $c_0$  using the as yet unknown policy function  $c_t = c(s_t)$ , the max operator is eliminated from (5.55), yielding

$$V(s_0) = u(c(s_0), s_0) + \beta E_0 V(g(s_0, c(s_0), v_1)). \quad (5.56)$$

Having made these substitutions, we have transformed Bellman's equation into a functional equation involving  $V(s)$  and  $c(s)$ . Given the concavity of  $u(c_t, s_t)$  and the convexity and compactness of

$$\{(s_{t+1}, s_t) : s_{t+1} \leq g(s_t, c_t, v_t)\},$$

the  $V(s)$  satisfying (5.56) will be strictly concave, and will be approached in the limit via iterations on

$$V^{j+1}(s) = \max_c \{u(c, s) + \beta EV^j(g(s, c, v'))\}. \quad (5.57)$$

Moreover, the  $c(s)$  that maximizes the right-hand side of (5.55) will be unique and time-invariant. For details see Stokey and Lucas (1989).

The fact that iterations on (5.57) are convergent implies that (5.55) is a contraction mapping. This property is critical in establishing value-function iterations and policy-function iterations as viable techniques for constructing  $\tilde{c}(s)$ .

### 5.2.2 Value-Function Iterations

As its name implies, the value-function-iteration approach uses iterations on (5.57) to generate an approximation of  $\tilde{c}(s)$ . To operationalize (5.57), we begin by constructing a discretization of the state space. That is, we divide the state space in such a way that  $s$  can take on only one of  $n$  possible values:

$$s \in \{s_1, \dots, s_n\}.$$

Also, we restrict  $c$  to take on only one of  $m$  possible values:

$$c \in \{c_1, \dots, c_m\}.$$

The finer the grid specified over  $\{s, c\}$ , the more accurate will be the resulting approximation. Note that as the dimensionality of the state space increases, for a given number of grid points specified along each dimension,  $n$  will increase exponentially. This phenomenon is another manifestation of the curse of dimensionality.

Given a selected discretization, the next step is to calculate a corresponding set of transition probabilities. This entails the use of (5.53), coupled

with the distributional assumption made for  $v_t$ . For each possible value  $s_i$  in the state-space grid, (5.53) and the distributional assumption made for  $v_t$  imply that a given choice of  $c$  induces a probability associated with the transition to the state-space value  $s_j$ ,  $j = 1, \dots, n$  in the following period. Let this probability be given by  $\pi_{ij}(c)$ . In practice,  $\pi_{ij}(c)$  can be constructed using Markov transition matrices or quadrature-approximation methods, both of which were introduced in section 1.

The collection of probabilities  $\pi_{ij}(c)$  calculated for a given  $c$  comprise a Markov transition matrix  $\Pi(c)$ . Given the present state  $s_i$ , the  $i^{\text{th}}$  row of  $\Pi(c)$  indicates the probability distribution induced over the state space in the next period given the specific choice of  $c$ .

Having constructed transition probabilities, the iteration process can be implemented. Let  $V$  denote the  $n \times 1$  vector of values associated with the initial endowment of the  $n$  possible values of  $s$  in the discretized state space. That is, the  $i^{\text{th}}$  element of  $V$ , denoted as  $V_i$ , is given by

$$\begin{aligned} V_i &= \max_{c \in \{c_1, \dots, c_m\}} \left\{ u(c, s_i) + \beta \sum_{j=1}^n \pi_{ij}(c) V_j \right\} \\ &= \max_{c \in \{c_1, \dots, c_m\}} \{u(c, s_i) + \beta \Pi_i(c) V\}, \end{aligned} \quad (5.58)$$

where  $\Pi_i(c)$  is the  $i^{\text{th}}$  row of  $\Pi(c)$ . Our goal is to obtain the solution to (5.58), which takes the form of two vectors: the vector of optimized values  $V^*$ , and the associated  $n \times 1$  vector of controls  $C^*$ . Specifically, the  $i^{\text{th}}$  element of  $C^*$  is given by

$$C_i^* = \arg \max_{c \in \{c_1, \dots, c_m\}} \{u(c, s_i) + \beta \Pi_i(c) V\}. \quad (5.59)$$

Note that  $C_i^*$  emerges as the best of the  $m$  possible choices for  $c$  available in the set  $\{c_1, \dots, c_m\}$ .

To explain the solution strategy it is useful at this point to simplify notation. Let  $S$  be an  $n \times 1$  vector containing the  $n$  possible values of the state space. Replacing  $s_i$  with  $S$  in (5.58), and replacing  $\Pi_i(c)$  with  $\Pi(c)$ , (5.58) can be rewritten as

$$\begin{aligned} V &= \max_{c \in \{c_1, \dots, c_m\}} \{u(c, S) + \beta \Pi(c) V'\} \\ &\equiv TV', \end{aligned} \quad (5.60)$$

where  $V'$  is next period's value function. This defines  $T$  as a mapping from next period's value function into today's. Likewise, (5.59) can be rewritten as

## 5.2 Iteration Techniques

$$\begin{aligned} C^* &= \arg \max_{c \in \{c_1, \dots, c_m\}} \{u(c, S) + \beta \Pi(c) V'\} \\ &\equiv \Gamma V'. \end{aligned} \quad (5.61)$$

The iteration process is initiated with a guess for  $V$ ,  $V^0$ . Using this guess, we generate the sequence

$$V^{j+1} = TV^j, \quad j = 1, \dots \quad (5.62)$$

Because (5.55) is a contraction mapping, this sequence is guaranteed to converge; for details, see Santos and Vigo-Aguiar (1998). Using a convergence criterion of the form  $\|V^{j+1} - V^j\| < \varepsilon$ , we obtain a stopping rule. Once a stopping point has been realized, an approximation to  $V^*$  has been obtained; the associated policy function is then obtained as  $C^* = \Gamma V^*$ . The  $n \times 1$  elements of  $C^*$  map directly into the elements of  $S$ ; this mapping yields the approximation  $\hat{c}(s)$  we set out to obtain.

Returning to the optimal growth model, given that

$$c = zk^\alpha + (1 - \delta)k - k', \quad (5.63)$$

the model can be characterized using the Bellman equation

$$B(k, z) = \max_{k'} \left\{ \frac{(zk^\alpha + (1 - \delta)k - k')^{1-\phi}}{1 - \phi} + \beta E_t B(k', z') \right\} \quad (5.64)$$

$$\text{s.t. } \ln z' = \rho \ln z + \varepsilon'. \quad (5.65)$$

In turn, the first-order condition for choosing  $k'$  is given by

$$(zk^\alpha + (1 - \delta)k - k')^{-\phi} = \beta E_t B_{k'}(k', z'). \quad (5.66)$$

Then an algorithm for obtaining the policy function  $k'(k, z)$  is as follows:

- Specify a suitable grid over  $(k, z)$ .
- Set  $j = 0$ , and construct  $B^j(k, z)$  by guessing a value for each  $(k, z)$  combination in the grid.
- Then for  $j = 0, 1, 2, \dots$ , and for each  $(k, z)$  combination in the grid:
  - Obtain  $k^{*j}$  as the solution to (5.66), with  $E_t B_{k'}^j(k', z')$  replacing  $E_t B_{k'}(k', z')$  (e.g., using nlsys).
  - Construct

$$B^{j+1}(k, z) = \frac{(zk^\alpha + (1 - \delta)k - k^{*j})^{1-\phi}}{1 - \phi} + \beta E_t B^j(k^{*j}, z').$$

- If  $\|B^{j+1} - B^j\| < \varepsilon$ , terminate loop over  $j$ ; otherwise set  $j = j + 1$  and continue loop over  $j$ . Denote the final set of values as  $B^*(k, z)$ .
- Obtain  $k'(k, z)$  using (5.66), with  $E_t B_{k'}^*(k', z')$  replacing  $E_t B_{k'}(k', z')$ .

If desired,  $c(k, z)$  can be obtained using  $c = zk^\alpha + (1 - \delta)k - k'(k, z)$ , and  $i(k, z)$  can be obtained using  $i = k'(k, z) - (1 - \delta)k$ .

There are two technicalities associated with this algorithm. The first involves the need to calculate the conditional expectations  $E_t B_k^j(k', z')$ ,  $E_t B_k^j(k^{**}, z')$ , and  $E_t B_k^{**}(k', z')$ . Since the value-function approach requires discretization of the state space, it is typical to specify a Markov process for  $z$ , rendering the calculation of expectations as straightforward. The second involves the need to construct  $B_k^j(k, z)$  from  $B^j(k, z)$ . This can be accomplished using linear slopes calculated as  $k$  ranges across grid points.

### Exercise 5.2

Reconduct Exercise 5.1 using a value-function iteration method in place of a projection method. Experiment with alternative grid densities specified over the state space, and compare the resulting approximations with your previous findings.

The solution of nonlinear equations such as (5.66) is often time-consuming. As the need to solve such equations arises routinely in implementing value-function iteration techniques, this problem limits the functionality of this solution method. However, Carroll (2006) has proposed a modification of this method that avoids the need to solve such equations for one-dimensional problems, and Barillas and Fernandez-Villaverde (2007) have generalized the modification to multidimensional problems. These modifications are known as endogenous grid methods (EGMs).

EGMs entail distinct treatment of the exogenous and endogenous components of the vector of state variables  $s$ . Specifically, under EGMs the grid over the state space is specified in terms of time- $t$  values of the exogenous state variables, but time- $(t+1)$  values of the endogenous variables. As a result, the need to solve equations such as (5.66) during the iterative convergence process is avoided, and replaced with the simpler problem of inferring time- $t$  values of the endogenous state variables that coincide with points within the modified grid. For concreteness, we characterize the implementation of EGMs in the context of the optimal growth model, following the presentation of Barillas and Fernandez-Villaverde.

Implementation begins with the introduction of what Carroll (2006) refers to as the “market resources” variable  $\Upsilon$ , defined as the sum of output and undepreciated capital:

$$\Upsilon = zk^\alpha + (1 - \delta)k \quad (5.67)$$

$$= c + k'. \quad (5.68)$$

### 5.2 Iteration Techniques

Recast in terms of  $\Upsilon$ , the Bellman equation associated with the model can be written as

$$V(\Upsilon, z) = \max_{k'} \left\{ \frac{(\Upsilon - k')^{1-\phi}}{1-\phi} + \beta E_t V(\Upsilon', z') \right\} \quad (5.69)$$

$$\text{s.t. } \Upsilon = zk^\alpha + (1 - \delta)k. \quad (5.70)$$

Further, noting that  $\Upsilon'$  is purely a function of  $k'$  and  $z'$ , the expectational term  $\beta E_t V(\Upsilon', z')$  can be written as

$$\beta E_t V(\Upsilon', z') = \tilde{V}(k', z), \quad (5.71)$$

and thus

$$V(\Upsilon, z) = \max_{k'} \left\{ \frac{(\Upsilon - k')^{1-\phi}}{1-\phi} + \tilde{V}(k', z) \right\}. \quad (5.72)$$

In turn, the first-order condition for choosing  $k'$  is given by

$$(c)^{-\phi} = \tilde{V}_k(k', z), \quad (5.73)$$

where recall  $c = \Upsilon - k'$ .

Then an algorithm for obtaining the policy function  $k'(k, z)$  is as follows:

- Specify a suitable grid over  $(k', z)$ .
- Set  $j = 0$ , and construct  $\tilde{V}^j(k', z)$  by guessing a value for each  $(k', z)$  combination in the grid.
- Then for  $j = 0, 1, 2, \dots$ , and for each  $(k', z)$  combination in the grid:
  - Obtain  $(c^*, \Upsilon^*)$  using

$$c^* = (\tilde{V}_{k'}^j(k', z))^{-\frac{1}{\phi}},$$

$$\Upsilon^* = c^* + k'.$$

— Construct

$$V^j(\Upsilon^*, z) = \left\{ \frac{(c^*)^{1-\phi}}{1-\phi} + \tilde{V}^j(k', z) \right\}.$$

— Construct

$$\tilde{V}^{j+1}(k', z) = \beta E_t V^j(\Upsilon^{**}, z').$$

- If  $\|\tilde{V}^{j+1} - \tilde{V}^j\| < \varepsilon$ , terminate loop over  $j$ ; otherwise set  $j = j + 1$  and continue loop over  $j$ . Denote the final set of values as  $\tilde{V}^*(k', z)$ .

- Obtain  $k'(k, z)$  by solving the nonlinear equation

$$(zk^\alpha + (1 - \delta)k - k')^{-\phi} = \tilde{V}_k^*(k', z). \quad (5.74)$$

Once again, if desired,  $c(k, z)$  can be obtained using  $c = zk^\alpha + (1 - \delta)k - k'(k, z)$ , and  $i(k, z)$  can be obtained using  $i = k'(k, z) - (1 - \delta)k$ .

Comparing algorithms, we see the critical difference between solution methods: while the nonlinear equation (5.66) must be solved for each iterative step under the standard algorithm, the nonlinear counterpart (5.74) must be solved only once. In the present context, this difference amounts to a reduction in computational time required for convergence by a factor of roughly 12 using a richly specified grid (41,000 points).

### Example 1

Reconduct Exercise 5.1 using an EGM. [Hint: For guidance, consult Barillas and Fernandez-Villaverde, 2007.]

#### 5.2.3 Policy-Function Iterations

Even when the need to solve equations such as (5.66) can be avoided, iterations on (5.62) can often be slow to converge in practice, a point emphasized, for example, by Ljungqvist and Sargent (2004). Convergence can often be enhanced significantly using policy-function iterations, sometimes referenced as the Howard improvement algorithm.

In order to describe this algorithm, we require some additional notation. Let the  $n \times 1$  vector  $C^1$  be the candidate guess for  $C^*$  associated with the guess chosen for  $V$ ; that is,  $C^1 = \Gamma V^0$ . Next, let  $U^1$  denote an  $n \times 1$  vector containing values of the current-period objective function evaluated at the  $n$  values that comprise  $S$  and  $C^1$ :

$$U^1 = u(C^1, S). \quad (5.75)$$

Finally, let  $\Pi^1$  denote the Markov transition matrix induced by coupling the corresponding elements of  $S$  and  $C^1$ ; that is, given the state  $s_i$  in the current period, the  $i^{\text{th}}$  element of  $C^1$  is chosen as the control, inducing the  $i^{\text{th}}$  row of probabilities contained in  $\Pi^1$ .

Now, note from the Bellman equation expressed in (5.58) that if  $C^1$  was in fact the optimal specification  $C^*$ , then replacing the individual elements  $c$  and  $s_i$  with the vectors  $C^*$  and  $S$ , the max operator would be eliminated, yielding

$$V^* = U^* + \beta \Pi^* V^*. \quad (5.76)$$

Solving for  $V^*$ , we obtain

$$V^* = (I - \beta \Pi^*)^{-1} U^*. \quad (5.77)$$

### 5.3 Perturbation Techniques

If  $C^1$  is not optimal, (5.77) instead serves as an updating equation in the Howard improvement algorithm.

Specifically, the algorithm works as follows. Beginning with a guess  $V^0$ , construct the corresponding guess  $C^1 = \Gamma V^0$ . Next, construct  $U^1$  using (5.75), and the associated transition matrix  $\Pi^1$ . Inserting  $(U^1, \Pi^1)$  into (5.77), obtain an updated vector  $V^2$ . By the  $j^{\text{th}}$  iteration, the updating equation becomes

$$V^{j+1} = (I - \beta \Pi^j)^{-1} U^j. \quad (5.78)$$

Once again, because (5.55) is a contraction mapping, the resulting sequence  $\{V^j\}$  is guaranteed to converge. Using a convergence criterion of the form  $\|V^{j+1} - V^j\| < \varepsilon$ , obtain a stopping rule. Once a stopping point has been realized, an approximation to  $V^*$  has been obtained; the associated policy function is again given by  $C^* = \Gamma V^*$ .

### Exercise 5.3

Reconduct Exercise 5.1 using a policy-function iteration method in place of a projection method. For alternative grid densities specified over the state space, compare the convergence speeds you obtain using this algorithm with those obtained using the value function iterations used in Exercise 5.2.

### 5.3 Perturbation Techniques

#### 5.3.1 Notation

Perturbation techniques were introduced to macroeconomists by Judd and Guu (1997), and their use in solving DSGE models was introduced by Schmitt-Grohé and Uribe (2004). Here we shall follow the notation of Schmitt-Grohé and Uribe closely, as their methodology (and corresponding code) remains in wide use.<sup>7</sup>

For this purpose, it is convenient to reexpress the model representation

$$\Gamma(E_t z_{t+1}, z_t, v_{t+1}) = 0$$

as

$$E_t f(c_{t+1}, c_t, s_{t+1}, s_t) = 0, \quad (5.79)$$

<sup>7</sup> Matlab code for executing their procedure is available at [http://econ.duke.edu/~uribe/2nd\\_order.htm](http://econ.duke.edu/~uribe/2nd_order.htm)

Analogous GAUSS code is available at the textbook website.

where recall  $c_t$  is  $n_c \times 1$ ,  $s_t$  is  $n_s \times 1$ , and  $n_c + n_s = n$ . Further,  $s_t$  is decomposed as

$$s_t = \begin{bmatrix} s_t^1 \\ s_t^2 \end{bmatrix},$$

where  $s_t^1$  is an  $n_{s1} \times 1$  vector consisting of endogenous state variables (such as physical capital),  $s_t^2$  is an  $n_{s2} \times 1$  vector of exogenous state variables (such as total factor productivity),  $n_{s2} = m$ , and  $n_{s1} = n_s - n_{s2}$ . Finally,  $s_t^2$  evolves according to

$$s_{t+1}^2 = \Lambda s_t^2 + \sigma \tilde{\eta} \varepsilon_{t+1}, \quad (5.80)$$

where  $\tilde{\eta}$  is an  $m \times m$  covariance matrix,  $\varepsilon_{t+1}$  is iid with zero mean and covariance matrix  $I$ ,  $v_{t+1} = \tilde{\eta} \varepsilon_{t+1}$ , and  $\sigma$  is a scalar known as a perturbation parameter. Note that when  $\sigma$  is set to zero, stochastic innovations are shut down, and the evolution of the model is purely deterministic.

The solution we seek is of the form

$$c_t = c(s_t, \sigma), \quad (5.81)$$

$$s_{t+1} = s(s_t, \sigma) + \sigma \eta \varepsilon_{t+1}, \quad (5.82)$$

where

$$\eta = \begin{bmatrix} 0 \\ n_{s1} \times n_{s2} \\ \tilde{\eta} \\ n_{s2} \times n_{s2} \end{bmatrix}. \quad (5.83)$$

It is constructed via the use of Taylor Series expansions of (5.81) and (5.82) around the nonstochastic steady state of the model, defined in the context of (5.79) as

$$f(\bar{c}, \bar{s}, \bar{s}) = 0. \quad (5.84)$$

Note in addition that

$$\bar{c} = c(\bar{s}, 0), \quad (5.85)$$

$$\bar{s} = s(\bar{s}, 0), \quad (5.86)$$

since for  $\sigma = 0$ ,

$$E_t f() = f().$$

### 5.3.2 Overview

In sum, we seek to approximate (5.81) and (5.82) via Taylor Series expansions around

$$(c_t, s_t, \sigma) = (\bar{c}, \bar{s}, 0).$$

A complication in constructing Taylor Series approximations in this case is that  $c(s_t, \sigma)$  and  $s(s_t, \sigma)$  are unknown. This complication is overcome via application of the Implicit Function Theorem:

#### Definition

For  $k$ -times differentiable  $H(x, y) : R^n \times R^m \rightarrow R^m$ , with

$$H(x_0, y_0) = 0,$$

and  $H_y(x_0, y_0)$  nonsingular, there is a unique function  $h : R^n \rightarrow R^m$  such that

$$y_0 = h(x_0),$$

and for  $x$  near  $x_0$ ,

$$H(x, h(x)) = 0.$$

Furthermore,  $h(x)$  is  $k$ -times differentiable, and its derivatives can be computed by implicit differentiation of the identity  $H(x, h(x)) = 0$ .

As we shall see, perturbation proceeds recursively via the tandem application of Taylor's Theorem and the Implicit Function Theorem, with the  $j^{\text{th}}$ -order step taking as input results from the  $(j-1)^{\text{st}}$ -order step, and with the  $0^{\text{th}}$ -order terms  $(\bar{c}, \bar{s}, 0)$  initializing the process. Before detailing these steps in application to DSGE models, we begin with a simple heuristic characterization of a one-dimensional problem, following Judd (1998).

Consider a problem of the form

$$f(x, \sigma) = 0, \quad (5.87)$$

with  $(x, \sigma)$  scalars. We seek an approximation to the solution

$$x = x(\sigma), \quad (5.88)$$

given that  $x(0)$  is known. The approximation is in the form of a Taylor Series expansion:

$$x \approx \bar{x} + x'(\sigma = 0)\sigma + \frac{1}{2}x''(\sigma = 0)\sigma^2 + \dots \quad (5.89)$$

Substituting for  $x$  using the solution we seek, we have

$$f(x(\sigma), \sigma) = 0.$$

Then differentiating with respect to  $\sigma$ , we have by the Implicit Function Theorem

$$f_x(x(\sigma), \sigma)x'(\sigma) + f_{\sigma}(x(\sigma), \sigma) = 0. \quad (5.90)$$

Then since  $x(0)$  is known, and the functional form of  $f()$  is given, this yields

$$x'(0) = -\frac{f_{\sigma}(x(0), 0)}{f_x(x(0), 0)}. \quad (5.91)$$

Thus taking the 0<sup>th</sup>-order term

$$\bar{x} \equiv x(0)$$

as an input, we obtain the first-order term  $x'(0)$  directly from (5.91). Finally, having calculated  $x'(0)$ , the first-order approximation to  $x(\sigma)$  we seek is given by

$$x(\sigma) \approx x(0) - \frac{f_{\sigma}(x(0), 0)}{f_x(x(0), 0)}\sigma. \quad (5.92)$$

The key observation from this sketch is as follows: given  $x(0)$ ,  $x'(0)$  obtains linearly from the first difference of  $f(x(\sigma), \sigma)$  with respect to  $\sigma$ .

To expand the approximation of  $x(\sigma)$  in (5.88) to second-order, we require an expression for  $x''(0)$ . Relying again on the Implicit Function Theorem, we can differentiate (5.90) with respect to  $\sigma$ , yielding

$$\begin{aligned} & f_x(x(\sigma), \sigma)x''(\sigma) + f_{xx}(x(\sigma), \sigma)(x'(\sigma))^2 \\ & + 2f_{x\sigma}(x(\sigma), \sigma)x'(\sigma) + f_{\sigma\sigma}(x(\sigma), \sigma) \\ & = 0. \end{aligned} \quad (5.93)$$

Given the expression for  $x'(0)$  calculated in (5.91) for the first-order approximation, and once again given that  $x(0)$  is known, solving for  $x''(0)$  yields

$$x''(0) = -\frac{f_{xx}(x(0), 0)(x'(0))^2 + 2f_{x\sigma}(x(0), 0)x'(0) + f_{\sigma\sigma}(x(0), 0)}{f_x(x(0), 0)}. \quad (5.94)$$

The second-order approximation we seek is then given by

$$x(\sigma) \approx x(0) - \frac{f_{\sigma}(x(0), 0)}{f_x(x(0), 0)}\sigma + \frac{1}{2}x''(0)\sigma^2. \quad (5.95)$$

### Exercise 5.4

Construct the third-order approximation of  $x(\sigma)$  by differentiating (5.93) with respect to  $\sigma$ , and solving for  $x'''(0)$  in the expression you obtain.

### 5.3.3 Application to DSGE Models

Application of the foregoing recursive algorithm to DSGE models is straightforward: the only significant difference involves the dimensionality of the problem. Specifically, instead of solving for a single term using a single equation in a given stage of the recursion as in the one-dimensional case, we must now solve for multiple terms using systems of equations. Code is available for this purpose, and it is straightforward to implement.

To reboot, for a DSGE model expressed as

$$E_t f(c_{t+1}, c_t, s_{t+1}, s_t) = 0, \quad (5.96)$$

$$s_{t+1}^2 = \Lambda s_t^2 + \sigma \tilde{\eta} \varepsilon_{t+1}, \quad (5.97)$$

our goal is to construct a  $k^{\text{th}}$ -order Taylor Series approximation to the unknown solution

$$c_t = c(s_t, \sigma), \quad (5.98)$$

$$s_{t+1} = s(s_t, \sigma) + \sigma \eta \varepsilon_{t+1}. \quad (5.99)$$

Let

$$[c_s]_a^i, \quad [c_{ss}]_{ab}^i$$

denote the  $(i, a)$  and  $(i, ab)$  elements of the  $n_c \times n_s$  and  $n_c \times n_s^2$  matrices

$$\frac{\partial c(s_t, \sigma)}{\partial s_t}, \quad \frac{\partial^2 c(s_t, \sigma)}{\partial s \partial s'},$$

evaluated at  $(\bar{s}, 0)$ . Also, let

$$[c_s]_a^i[s - \bar{s}]_a = \sum_{a=1}^{n_s} [c_s]_a^i(s_a - \bar{s}_a),$$

$$[c_{ss}]_{ab}^i[s - \bar{s}]_a[s - \bar{s}]_b = \sum_{a=1}^{n_s} \sum_{b=1}^{n_s} [c_{ss}]_{ab}^i(s_a - \bar{s}_a)(s_b - \bar{s}_b),$$

and so on. Focusing on the derivation of second-order approximations, the representation of  $c(s_t, \sigma)$  we seek is of the form

$$\begin{aligned} [c(s_t, \sigma)]^i &= [\bar{c}]^i + [c_s]_a^i [s - \bar{s}]_a + [c_\sigma]^i \sigma \\ &\quad + \frac{1}{2} [c_{ss}]_{ab}^i [s - \bar{s}]_a [s - \bar{s}]_b \\ &\quad + \frac{1}{2} [c_{s\sigma}]_a^i [s - \bar{s}]_a \sigma \\ &\quad + \frac{1}{2} [c_{\sigma s}]_a^i [s - \bar{s}]_a \sigma \\ &\quad + \frac{1}{2} [c_{\sigma\sigma}]^i \sigma^2, \end{aligned} \quad (5.100)$$

$i = 1, \dots, n_c$ ;  $a, b = 1, \dots, n_s$ . Likewise, the representation of  $s(s_t, \sigma)$  we seek is of the form

$$\begin{aligned} [s(s_t, \sigma)]^j &= [\bar{s}]^j + [s_s]_a^j [s - \bar{s}]_a + [s_\sigma]^j \sigma \\ &\quad + \frac{1}{2} [s_{ss}]_{ab}^j [s - \bar{s}]_a [s - \bar{s}]_b \\ &\quad + \frac{1}{2} [s_{s\sigma}]_a^j [s - \bar{s}]_a \sigma \\ &\quad + \frac{1}{2} [s_{\sigma s}]_a^j [s - \bar{s}]_a \sigma \\ &\quad + \frac{1}{2} [s_{\sigma\sigma}]^j \sigma^2, \end{aligned} \quad (5.101)$$

$j = 1, \dots, n_s$ ;  $a, b = 1, \dots, n_s$ . Note that if

$$[c_s]_a^i, \quad [c_{ss}]_{ab}^i,$$

and so on are in the form of elasticities, then  $[s - \bar{s}]_a$ , and so on represent logged deviations from steady states. That is, our approximations can accommodate both linear and log-linear model representations.

To construct these approximations, we proceed by substituting for  $(c_t, c_{t+1}, s_{t+1})$  in (5.96) using (5.98) and (5.99); eliminating time subscripts, and denoting time- $(t+1)$  variables with primes, substitution yields

$$F(s, \sigma) \equiv E_\theta f(c(s(s, \sigma) + \sigma \eta \varepsilon', \sigma), c(s, \sigma), s(s, \sigma) + \sigma \eta \varepsilon', s) = 0. \quad (5.102)$$

To construct our linear approximation, we use the Implicit Function Theorem to obtain

$$F_s(s, \sigma) = 0$$

$$F_\sigma(s, \sigma) = 0,$$

### 5.3 Perturbation Techniques

where the first expression represents a set of  $n \cdot n_s$  equalities, and the second a set of  $n$  equalities. Denoting

$$[f_s]_a^i [c_s]_b^a [s_s]_j^b = \sum_{a=1}^{n_c} \sum_{b=1}^{n_s} \frac{\partial f^i}{\partial c^a} \frac{\partial c^a}{\partial s^b} \frac{\partial s^b}{\partial s_j},$$

and so on,  $F_s(\bar{s}, 0) = 0$  is given by

$$\begin{aligned} [F_s(\bar{s}, 0)]_j^i &= [f_c]_a^i [c_s]_b^a [s_s]_j^b + [f_c]_a^i [c_\sigma]_j^a + [f_s]_b^i [s_s]_j^b + [f_s]_j^i \\ &= 0, \end{aligned} \quad (5.103)$$

$i = 1, \dots, n$ ;  $j, b = 1, \dots, n_s$ ;  $a = 1, \dots, n_c$ . Since the derivatives of  $f()$  evaluated at  $(\bar{c}, \bar{s})$  are known,

$$[F_s(\bar{s}, 0)]_j^i = 0$$

is a system of  $n \cdot n_s$  quadratic equations in the  $n \cdot n_s$  unknown elements of  $c_s()$  and  $s_s()$ .

This takes us halfway towards our linear approximations of  $c()$  and  $s()$ . Below we shall discuss an alternative approach to obtaining  $c_s()$  and  $s_s()$ . To complete the construction of our linear approximations, we use

$$F_\sigma(\bar{s}, 0) = 0,$$

which is given by

$$\begin{aligned} [F_\sigma(\bar{s}, 0)]^i &= [f_c]_a^i [c_s]_b^a [s_\sigma]^b + [f_c]_a^i [c_\sigma]^a + [f_c]_a^i [c_\sigma]^a + [f_s]_b^i [s_\sigma]^b \\ &= 0, \end{aligned} \quad (5.104)$$

$i = 1, \dots, n$ ;  $a = 1, \dots, n_c$ ;  $b = 1, \dots, n_s$ . (Note: expressions involving  $\varepsilon'$  are eliminated by application of the expectations operator.) As these equations are linear and homogeneous in  $([s_\sigma]^b, [c_\sigma]^a)$ , these terms must be zero. Thus our first-order approximations are given by

$$[c(s_t, \sigma)]^i = [\bar{c}]^i + [c_s]_a^i [s - \bar{s}]_a \quad (5.105)$$

$$[s(s_t, \sigma)]^j = [\bar{s}]^j + [s_s]_a^j [s - \bar{s}]_a, \quad (5.106)$$

$i = 1, \dots, n$ ;  $j = 1, \dots, n_s$ , with  $[c_s]_a^i$  and  $[s_s]_a^j$  obtained from (5.104).

As an aside, note that an alternative approach to obtaining

$$(c_s(), s_s())$$

involves the transformation of the linear model representation

$$x_{t+1} = Fx_t + Gv_{t+1}$$

(obtained, e.g., using Sims' method described in chapter 4) into

$$\begin{aligned} c_t &= Cs_t, \\ s_{t+1} &= \Gamma s_t. \end{aligned}$$

A simple method for doing so is as follows:

- Simulate  $\{v_t\}_{t=1}^T$  from its known distribution (note:  $T$  need not be large).
- Using  $x_0 = \bar{x}$  and  $\{v_t\}_{t=1}^T$ , simulate  $\{x_t\}_{t=1}^T$  using

$$x_{t+1} = Fx_t + Gv_{t+1}.$$

- Divide  $\{x_t\}_{t=1}^T$  into  $\{c_t\}_{t=1}^T$ ,  $\{s_t\}_{t=1}^T$ , construct  $y$  as the  $T \times n_c$  matrix with  $t^{\text{th}}$  row  $c'_t$ , and  $X$  as the  $T \times n_s$  matrix with  $t^{\text{th}}$  row  $s'_t$ , and obtain

$$C' = (X'X)^{-1}X'y.$$

Then to construct  $s_{t+1} = \Gamma s_t$ :

- Divide  $\{x_t\}_{t=1}^T$  into  $\{s_t^1\}_{t=1}^T$ ,  $\{s_t^2\}_{t=1}^T$ , construct  $y$  as the  $T \times n_{s1}$  matrix with  $t^{\text{th}}$  row  $s_t^{1'}$ , and  $X$  as the  $T \times n_s$  matrix with  $t^{\text{th}}$  row  $s_t'$ , and obtain

$$\tilde{F}' = (X'X)^{-1}X'y.$$

- Then recalling that

$$s_{t+1}^2 = \Lambda s_t^2 + \sigma \tilde{\eta} \varepsilon_{t+1},$$

construct

$$\Gamma = \begin{bmatrix} \tilde{F}' \\ 0 \sim \Lambda \end{bmatrix}, \quad \underbrace{\tilde{F}'}_{n_{s1} \times n_s} \quad \underbrace{0 \sim \Lambda}_{n_{s2} \times n_s}.$$

To obtain the second-order approximations (5.100) and (5.101), we first differentiate (5.103) with respect to  $s$  to identify

$$c_{ss}(\bar{s}, 0), \quad s_{ss}(\bar{s}, 0).$$

Next, we differentiate (5.104) with respect to  $\sigma$  to identify

$$c_{\sigma\sigma}(\bar{s}, 0), \quad s_{\sigma\sigma}(\bar{s}, 0).$$

Finally, we differentiate (5.104) with respect to  $s$  to identify

$$c_{s\sigma}(\bar{s}, 0), \quad s_{s\sigma}(\bar{s}, 0).$$

Specifically, differentiating (5.103) with respect to  $s$ , we obtain

$$\begin{aligned} [F_{ss}(\bar{s}, 0)]_{jk}^i &= ([f_{c'c'}]_{\alpha\gamma}^i [c_s]_{\delta}^{\gamma} [s_s]_{k}^{\delta} + [f_{c'c}]_{\alpha\gamma}^i [c_s]_{\delta}^{\gamma}) \\ &\quad + [f_{c's'}]_{\alpha\delta}^i [s_s]_{k}^{\delta} + [f_{c's}]_{\alpha k}^i) [c_s]_{\beta}^{\alpha} [s_s]_{j}^{\beta} \\ &\quad + [f_{c'}]_{\alpha}^i [c_{ss}]_{\beta\delta}^{\alpha} [s_s]_{k}^{\delta} [s_s]_{j}^{\beta} \\ &\quad + [f_{c'}]_{\alpha}^i [c_s]_{\beta}^{\alpha} [s_{ss}]_{jk}^{\beta} \\ &\quad + ([f_{cc'}]_{\alpha\gamma}^i [c_s]_{\delta}^{\gamma} [s_s]_{k}^{\delta} + [f_{cc'}]_{\alpha\gamma}^i [c_s]_{\delta}^{\gamma} + [f_{cc'}]_{\alpha\delta}^i [s_s]_{k}^{\delta} \\ &\quad + [f_{cs}]_{\alpha k}^i) [c_s]_{\beta}^{\alpha} + [f_{c'}]_{\alpha}^i [c_{ss}]_{jk}^{\alpha} \\ &\quad + ([f_{s'c'}]_{\beta\gamma}^i [c_s]_{\delta}^{\gamma} [s_s]_{k}^{\delta} + [f_{s'c'}]_{\beta\gamma}^i [c_s]_{\delta}^{\gamma} + [f_{s's'}]_{\beta\delta}^i [s_s]_{k}^{\delta} \\ &\quad + [f_{s's}]_{\beta k}^i) [s_s]_{j}^{\beta} + [f_{s'}]_{\beta}^i [s_{ss}]_{jk}^{\beta} \\ &\quad + [f_{sc'}]_{j\gamma}^i [c_s]_{\delta}^{\gamma} [s_s]_{k}^{\delta} + [f_{sc'}]_{j\gamma}^i [c_s]_{\delta}^{\gamma} + [f_{ss'}]_{j\delta}^i [s_s]_{k}^{\delta} + [f_{ss'}]_{jk}^i \\ &= 0; \quad i = 1, \dots, n, \quad j, k, \beta, \delta = 1, \dots, n_s; \\ &\quad \alpha, \gamma = 1, \dots, n_c. \end{aligned} \quad (5.107)$$

Since we know the derivatives of  $f$  as well as the first derivatives of  $c$  and  $s$  evaluated at  $(c', c, s', s) = (\bar{c}, \bar{c}, \bar{s}, \bar{s})$ , (5.107) is a system of  $n \times n_s \times n_s$  linear equations in the  $n \times n_s \times n_s$  unknowns given by the elements of  $c_{ss}$  and  $s_{ss}$ .

Next, differentiating (5.104) with respect to  $\sigma$ , we obtain

$$\begin{aligned} [F_{\sigma\sigma}(\bar{s}, 0)]^i &= \\ &[f_c]^i_a [c_s]^a_b [s_{\sigma\sigma}]^b \\ &+ [f_{c'c}]^i_a \gamma [c_s]^\gamma_\delta [\eta]^\delta_\xi [c_s]^a_b [\eta]^\phi_\phi [I]^\phi_\xi \\ &+ [f_{c's'}]_a^i [\eta]^\delta_\xi [c_s]^a_b [\eta]^\phi_\phi [I]^\phi_\xi \\ &+ [f_{c'}]^i_a [c_{ss}]^a_b [\eta]^\delta_\xi [\eta]^\phi_\phi [I]^\phi_\xi \\ &+ [f_{c'}]^i_a [c_{\sigma\sigma}]^a \\ &+ [f_c]^i_a [c_{\sigma\sigma}]^a \\ &+ [f_s]^i_b [s_{\sigma\sigma}]^b \\ &+ [f_{s'c'}]^i_b [\eta]^\delta_\xi [\eta]^\phi_\phi [I]^\phi_\xi \\ &+ [f_{s's'}]^i_b [\eta]^\delta_\xi [\eta]^\phi_\phi [I]^\phi_\xi \\ &= 0; \quad i = 1, \dots, n; \quad a, \gamma = 1, \dots, n_c; \quad b, \delta = 1, \dots, n_s; \\ &\phi, \xi = 1, \dots, n_e. \end{aligned} \tag{5.108}$$

This is a system of  $n$  linear equations in the  $n$  unknowns given by the elements of  $c_{\sigma\sigma}$  and  $s_{\sigma\sigma}$ .

Finally, differentiating (5.104) with respect to  $s$ , taking into account that all terms containing either  $c_\sigma$  or  $s_\sigma$  are zero at  $(\bar{s}, 0)$ , we have

$$\begin{aligned} [F_{\sigma s}(\bar{s}, 0)]_j^i &= [f_c]^i_a [c_s]^a_b [s_{\sigma s}]_j^b + [f_c]^i_a [c_{\sigma s}]_j^a [s_s]^\gamma_j \\ &+ [f_c]^i_a [c_{\sigma s}]_j^a + [f_s]^i_b [s_{\sigma s}]_j^b \\ &= 0; \quad i = 1, \dots, n; \quad a = 1, \dots, n_c; \quad b, \gamma, j = 1, \dots, n_s. \end{aligned} \tag{5.109}$$

This is a system of  $n \times n_s$  equations in the  $n \times n_s$  unknowns given by the elements of  $c_{\sigma s}$  and  $s_{\sigma s}$ . However, the system is homogeneous in the unknowns, thus

$$c_{\sigma s} = 0 \quad s_{\sigma s} = 0.$$

In sum, having solved for

$$c_{ss}(\bar{s}, 0), \quad s_{ss}(\bar{s}, 0), \quad c_{\sigma\sigma}(\bar{s}, 0), \quad s_{\sigma\sigma}(\bar{s}, 0), \quad c_{s\sigma}(\bar{s}, 0), \quad s_{s\sigma}(\bar{s}, 0)$$

in (5.107) through (5.109), we have all of the terms needed for the second-order approximations (5.100) and (5.101).

### 5.3.4 Application to an Asset-Pricing Model

We conclude with an example sufficiently simple to render the calculations described above as easily tractable analytically. The example is a simplification of the single-asset pricing environment presented in chapter 3, section 3.1, in which dividends serve as the lone stochastic process. Recalling that in equilibrium

$$c_t = d_t \quad \forall t,$$

the model is given by

$$\begin{aligned} u'(d_t)p_t &= \beta E_t(u'(d_{t+1})(p_{t+1} + d_{t+1})) \\ d_{t+1} &= (1 - \rho)\bar{d} + \rho d_t + \sigma \varepsilon_{t+1} \quad \varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2). \end{aligned}$$

Note in this case that the state is comprised exclusively as the exogenous dividend process

$$s_t = d_t - \bar{d},$$

implying

$$\eta = \tilde{\eta} = \sigma_\varepsilon.$$

Moreover, this implies that the state-transition equation need not be constructed, but is given directly as

$$\begin{aligned} s_{t+1} &= s(s_t, \sigma) + \sigma \eta \varepsilon_{t+1} \\ &= (1 - \rho)\bar{d} + \rho d_t + \sigma \sigma_\varepsilon \varepsilon_{t+1}. \end{aligned}$$

Thus in this case we merely require the approximation of the policy function

$$c_t = c(s_t, \sigma),$$

where the controls are comprised exclusively as

$$c_t = p_t,$$

with steady state

$$\bar{p} = \frac{1}{r}\bar{d}.$$

The approximation we seek is of the form

$$\begin{aligned} c(s_t, \sigma) &= \bar{p} + [c_s](d - \bar{d}) + [c_\sigma]\sigma \\ &+ \frac{1}{2}[c_{ss}](d - \bar{d})^2 \\ &+ \frac{1}{2}[c_{\sigma\sigma}]\sigma^2, \end{aligned} \tag{5.110}$$

since it is known that

$$[c_{ss}] = [c_{\sigma s}] = 0$$

in light of (5.109).

In terms of the representation

$$F(s, \sigma) \equiv E_t f(c(s, \sigma) + \sigma \eta \varepsilon', \sigma), c(s, \sigma), s(s, \sigma) + \sigma \eta \varepsilon', s) = 0,$$

under the redefinition

$$c(s, \sigma) \equiv p(d, \sigma), \quad s(s, \sigma) = (1 - \rho)\bar{d} + \rho d_t,$$

the model is given by  $F(\bar{d}, \sigma) =$

$$E_t \left[ \begin{array}{l} u'(\bar{d})p(\bar{d}, \sigma) - \beta u'((1 - \rho)\bar{d} + \rho d + \sigma \sigma_\varepsilon \varepsilon') \\ (p((1 - \rho)\bar{d} + \rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) + (1 - \rho)\bar{d} + \rho d + \sigma \sigma_\varepsilon \varepsilon') \end{array} \right] = 0.$$

Hereafter, to ease notation, we shall drop the appearance of the constant term  $(1 - \rho)\bar{d}$  from all expressions in  $F(\bar{d}, \sigma)$  involving  $d_{t+1}$  expressed as a function of  $d_t$ . Thus the model is expressed as

$$F(\bar{d}, \sigma) = E_t \left[ \begin{array}{l} u'(\bar{d})p(\bar{d}, \sigma) \\ -\beta u'(\rho d + \sigma \sigma_\varepsilon \varepsilon') \cdot (p(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) + \rho d + \sigma \sigma_\varepsilon \varepsilon') \end{array} \right] = 0.$$

Proceeding with the calculation of first-order terms, differentiating  $F(\bar{d}, \sigma)$  with respect to  $\bar{d}$ , we obtain

$$F_{\bar{d}}(\bar{d}, \sigma) = E_t \left[ \begin{array}{l} u''(\bar{d})p(\bar{d}, \sigma) + u'(\bar{d})p_{\bar{d}}(\bar{d}, \sigma) \\ -\beta u''(\rho d + \sigma \sigma_\varepsilon \varepsilon') \cdot \rho \\ (p(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) + \rho d + \sigma \sigma_\varepsilon \varepsilon') \\ -\beta u'(\rho d + \sigma \sigma_\varepsilon \varepsilon') \cdot (p_{\bar{d}}(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot \rho + \rho) \end{array} \right] = 0.$$

Applying the expectations operator, and evaluating at  $(\bar{d}, \bar{p}, \sigma = 0)$ , we obtain

$$\begin{aligned} & u''(\bar{d})p(\bar{d}, 0) + u'(\bar{d})p_{\bar{d}}(\bar{d}, 0) - \beta u''(\bar{d})\rho \cdot (\bar{p} + \bar{d}) \\ & - \beta u'(\bar{d})(p_{\bar{d}}(\bar{d}, 0)\rho + \rho) \\ & = 0. \end{aligned}$$

### 5.3 Perturbation Techniques

Note in the linear utility case,

$$u'(\bar{d}) = 1, \quad u''(\bar{d}) = 0,$$

and thus

$$p_{\bar{d}}(\bar{d}, 0) - \beta(p_{\bar{d}}(\bar{d}, 0)\rho + \rho) = 0,$$

or

$$p_{\bar{d}}(\bar{d}, 0) = \frac{\rho\beta}{1 - \rho\beta}.$$

In the case of CRRA preferences, with

$$u'(\bar{d}) = \bar{d}^{-\gamma},$$

solving for  $p_{\bar{d}}(\bar{d}, 0)$  yields

$$\begin{aligned} p_{\bar{d}}(\bar{d}, 0) &= \frac{-u''(\bar{d})p(\bar{d}, 0) + \beta u''(\bar{d})\rho \cdot (\bar{p} + \bar{d}) + \rho\beta u'(\bar{d})}{u'(\bar{d})(1 - \rho\beta)} \quad (5.111) \\ &= \left( \frac{-u''(\bar{d})\bar{d}}{u'(\bar{d})} \right) \left( \frac{1/r - \beta\rho \cdot (1/r + 1) - \rho\beta u'(\bar{d})/u''(\bar{d})\bar{d}}{(1 - \rho\beta)} \right) \\ &= \gamma \left( \frac{1/r - \beta\rho \cdot (1/r + 1) + \rho\beta\gamma^{-1}}{(1 - \rho\beta)} \right) \\ &= \gamma \left( \frac{1/r(1 - \rho\beta) + \rho\beta(\gamma^{-1} - 1)}{(1 - \rho\beta)} \right). \end{aligned}$$

Next, differentiating  $F(\bar{d}, \sigma)$  with respect to  $\sigma$ , we obtain

$$F_{\sigma}(\bar{d}, \sigma) = E_t \left[ \begin{array}{l} u'(\bar{d})p_{\sigma}(\bar{d}, \sigma) - \beta u''(\rho d + \sigma \sigma_\varepsilon \varepsilon') \cdot \sigma \sigma_\varepsilon \varepsilon' \\ (p(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) + \rho d + \sigma \sigma_\varepsilon \varepsilon') \\ -\beta u'(\rho d + \sigma \sigma_\varepsilon \varepsilon') \cdot (p_{\sigma}(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot \sigma \sigma_\varepsilon \varepsilon' + p_{\sigma}(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) + \sigma \sigma_\varepsilon \varepsilon') \end{array} \right] = 0.$$

Applying the expectations operator, and evaluating at  $(\bar{d}, \bar{p}, \sigma = 0)$ , we obtain

$$u'(\bar{d})p_{\sigma}(\bar{d}, 0) - \beta u'(\bar{d})p_{\sigma}(\bar{d}, 0) = 0,$$

implying

$$p_{\sigma}(\bar{d}, 0) = 0.$$

Thus our linear approximation of the policy function is given by

$$[c(s_t, \sigma)] = \bar{p} + \gamma \left( \frac{1/r(1 - \rho\beta) + \rho\beta(\gamma^{-1} - 1)}{(1 - \rho\beta)} \right) (d - \bar{d}). \quad (5.112)$$

Pursuing the second-order approximation, differentiating  $F_d(d, \sigma)$  with respect to  $d$ , we obtain

$$F_{dd}(d, \sigma) = E_t \left[ \begin{array}{l} u'''(d)p(d, \sigma) + 2u''(d)p_d(d, \sigma) \\ \quad + u'(d)p_{dd}(d, \sigma) \\ \quad - \beta u'''(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot \rho^2 \\ \quad (p(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) + \rho d + \sigma \sigma_\varepsilon \varepsilon') \\ \quad - 2\beta u''(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot \rho \\ \quad (p_d(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) + \rho) \\ \quad - \beta u'(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot (p_{dd}(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot \rho^2) \end{array} \right] = 0.$$

Applying the expectations operator and evaluating at  $(\bar{d}, \bar{p}, \sigma = 0)$ , we obtain

$$\begin{aligned} & u'''(\bar{d})p(\bar{d}, 0) + 2u''(\bar{d})p_d(\bar{d}, 0) \\ & + u'(\bar{d})p_{dd}(\bar{d}, 0) - \beta u'''(\bar{d}) \cdot \rho^2 \cdot (p(\bar{d}, 0) + \bar{d}) \\ & - 2\beta u''(\bar{d}) \cdot \rho \cdot (p_d(\bar{d}, 0) + 1) \\ & - \beta u'(\bar{d})p_{dd}(\bar{d}, 0) \cdot \rho^2 \\ & = 0. \end{aligned}$$

Solving for  $p_{dd}(\bar{d}, 0)$ , we obtain

$$p_{dd}(\bar{d}, 0) = \frac{-a - 2b + c + 2d}{u'(\bar{d})(1 - \rho\beta)}, \quad (5.113)$$

where

$$\begin{aligned} a &= u'''(\bar{d})p(\bar{d}, 0) \\ b &= u''(\bar{d})p_d(\bar{d}, 0) \\ c &= \rho^2 \beta u'''(\bar{d}) \cdot (p(\bar{d}, 0) + \bar{d}) \\ d &= \rho^2 \beta u''(\bar{d})(p_d(\bar{d}, 0) + 1). \end{aligned}$$

### 5.3 Perturbation Techniques

Next, differentiating  $F_\sigma(d, \sigma)$  with respect to  $\sigma$ , we obtain

$$F_{\sigma\sigma}(d, \sigma) = E_t \left[ \begin{array}{l} u'(d)p_{\sigma\sigma}(d, \sigma) \\ -\beta u'''(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot \sigma_\varepsilon^2 \varepsilon'^2 \\ (p(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) + \rho d + \sigma \sigma_\varepsilon \varepsilon') \\ -\beta u''(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot \sigma_\varepsilon \varepsilon' \\ (p_d(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot \sigma_\varepsilon \varepsilon' + p_\sigma(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) + \sigma \sigma_\varepsilon \varepsilon') \\ -\beta u''(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot \sigma_\varepsilon \varepsilon' \\ (p_{dd}(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) \cdot \sigma_\varepsilon^2 \varepsilon'^2 + p_\sigma(\rho d + \sigma \sigma_\varepsilon \varepsilon', \sigma) + p_{\sigma\sigma}(d, \sigma)) \end{array} \right] = 0,$$

where we have exploited the fact that  $p_{d\sigma}() = p_{\sigma d}() = 0$ .

Applying the expectations operator, evaluating at  $(\bar{d}, \bar{p}, \sigma = 0)$ , and recalling  $p_\sigma(\bar{d}, 0) = 0$ , we obtain

$$\begin{aligned} & u'(\bar{d})p_{\sigma\sigma}(\bar{d}, 0) \\ & -\beta u'''(\bar{d}) \cdot \sigma_\varepsilon^2 \cdot (\bar{p} + \bar{d}) \\ & -\beta u'(\bar{d}) \\ & (p_{dd}(\bar{d}, 0) \cdot \sigma_\varepsilon^2 + p_{\sigma\sigma}(\bar{d}, 0)) \\ & = 0. \end{aligned}$$

Given the expression for  $p_{dd}(\bar{d}, 0)$ , we solve for  $p_{\sigma\sigma}(\bar{d}, 0)$  as

$$p_{\sigma\sigma}(\bar{d}, 0) = \sigma_\varepsilon^2 \frac{\beta}{1 - \beta} \left( \frac{u'''(\bar{d})}{u'(\bar{d})} + p_{dd}(\bar{d}, 0) \right). \quad (5.114)$$

Thus our quadratic approximation of the policy function is given by

$$\begin{aligned} [c(s_t, \sigma)] &= \bar{p} + \gamma \left( \frac{1/r(1 - \rho\beta) + \rho\beta(\gamma^{-1} - 1)}{(1 - \rho\beta)} \right) (d - \bar{d}) \\ &+ \frac{1}{2} p_{dd}(\bar{d}, 0)(d - \bar{d})^2 \\ &+ \frac{1}{2} p_{\sigma\sigma}(\bar{d}, 0)\sigma^2, \end{aligned} \quad (5.115)$$

where  $p_{dd}(\bar{d}, 0)$  is defined in (5.113), and  $p_{\sigma\sigma}(\bar{d}, 0)$  is defined in (5.114).

#### Exercise 5.5

GAUSS code for implementing a second-order approximation of the RBC model is available at the textbook website. Using this code, compare the

accuracy of model approximations generated using the projection method implemented in Exercise 5.1, and those generated using the iterative methods, implemented in Exercises 5.2 and 5.3. To do so, evaluate the relative accuracy of  $F(\bar{c}(s)) \approx 0$ , where  $F(c(s)) = 0$  is the system of functional equations used to establish policy functions under the projection method.

#### Exercise 5.6

Adapt the GAUSS code just noted for the RBC model to the asset-pricing model described above. Compare the policy function you obtain for prices with that of (5.115): they should be identical. Next, extend the model to the full specification presented in chapter 3, section 3.1.

### Part III

---

## Data Preparation and Representation

## Chapter 6 - Structural Macroeconomics

### *Removing Trends and Isolating Cycles*

There is no branch of mathematics, however abstract, which may not someday be applied to the phenomena of the real world.

—Nicolai Lobachevsky

Just as DSGE models must be primed for empirical analysis, so too must the corresponding data. Broadly speaking, data preparation involves three steps. A guiding principle behind all three involves the symmetric treatment of the actual data and their theoretical counterparts. First, correspondence must be established between what is being characterized by the model and what is being measured in the data. For example, if the focus is on a business cycle model that does not include a government sector, it would not be appropriate to align the model's characterization of output with the measure of aggregate GDP reported in the National Income and Product Accounts. The papers in Cooley (1995) provide a good set of examples for dealing with this issue, and do so for a broad range of models.

The second and third steps involve the removal of trends and the isolation of cycles. Regarding the former, model solutions are typically in terms of stationary versions of variables: the stochastic behavior of the variables is in the form of temporary departures from steady state values. Corresponding data are represented analogously. So again using a business cycle model as an example, if the model is designed to characterize the cyclical behavior of a set of time series, and the time series exhibit both trends and cycles, the trends are eliminated prior to analysis. In such cases, it is often useful to build both trend and cyclical behavior into the model, and eliminate trends from the model and actual data in parallel fashion. Indeed, a typical objective in the business cycle literature is to determine whether models capable of capturing salient features of economic growth can also account for observed patterns of business cycle activity. Under this objective, the specification of the model is subject to the constraint that it must successfully characterize trend behavior. That constraint having been satisfied, trends are eliminated appropriately and the analysis proceeds with an investigation of cyclical behavior. Steady states in this case are interpretable as the relative heights of trend lines.

Regarding the isolation of cycles, this is closely related to the removal of trends. Indeed, for a time series exhibiting cyclical deviations about a trend, the identification of the trend automatically serves to identify the cyclical deviations as well. However, even after the separation of trend from cycle is accomplished, additional steps may be necessary to isolate cycles by the frequency of their recurrence. Return again to the example of a business cycle model. By design, the model is intended to characterize patterns of fluctuations in the data that recur at business cycle frequencies between approximately 6 and 40 quarters. It is not intended to characterize seasonal fluctuations. Yet unless additional steps are taken, the removal of the trend will leave such fluctuations intact, and their presence can have a detrimental impact on inferences involving business cycle behavior. (For an example of a model designed to jointly characterize both cyclical and seasonal variations in aggregate economic activity, see Wen, 2002.)

The isolation of cycles is also related to the task of aligning models with appropriate data, because the frequency with which data are measured in part determines their cyclical characteristics. For example, empirical analyses of economic growth typically involve measurements of variables averaged over long time spans (e.g., over half-decade intervals). This is because the models in question are not designed to characterize business cycle activity, and time aggregation at the five-year level is typically sufficient to eliminate the influence of cyclical variations while retaining relevant information regarding long-term growth. For related reasons, analyses of aggregate asset-pricing behavior are typically conducted using annual data, which mitigates the need to control, for example, for seasonal fluctuations. Analyses of business cycle behavior are typically conducted using quarterly data. Measurement at this frequency is not ideal, because it introduces the influence of seasonal fluctuations into the analysis; but on the other hand, aggregation to an annual frequency would entail an important loss of information regarding fluctuations observed at business cycle frequencies. Thus an alternative to time aggregation is needed to isolate cycles in this case.

This chapter presents alternative approaches available for eliminating trends and isolating cycles. Supplements to the brief coverage of these topics provided here are available from any number of texts devoted to time-series analysis (e.g., Harvey, 1993; Hamilton, 1994). Specific textbook coverage of cycle isolation in the context of macroeconomic applications is provided by Sargent (1987a), Kaiser and Maravall (2001), and Canova (2007).

To illustrate the concepts introduced in this chapter, we work with a prototypical data set used to analyze business cycle behavior. It is designed for alignment with the real business cycle model introduced in chapter 3, section 1. The data are contained in the text file `rbcdata.txt`, available for

### 6.1 Removing Trends

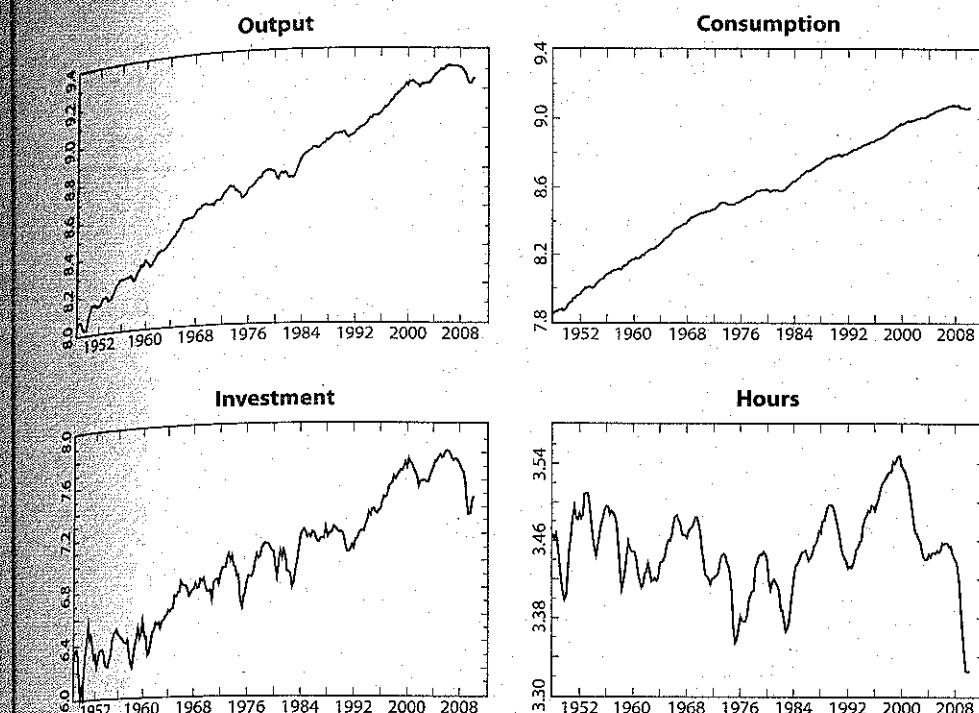


Figure 6.1 Business Cycle Data

downloading at the textbook website. A description of the data is contained in an accompanying file. Briefly, the data set consists of four time series: consumption of nondurables and services; gross private domestic investment; output, measured as the sum of consumption and investment; and hours of labor supplied in the nonfarm business sector. Each variable is real, measured in per capita terms, and is seasonally adjusted. The data are quarterly, and span 1948:I through 2010:I. In addition, we also work with the nonseasonally adjusted counterpart to consumption. Logged time-series trajectories of the seasonally adjusted data are illustrated in figure 6.1.

### 6.1 Removing Trends

There are three leading approaches to removing trends from macroeconomic time series. The goal under all three is to transform the data into mean-zero covariance-stationary stochastic processes (CSSPs). By definition, such processes have time-invariant second moments; therefore,

sample averages may be used to estimate population averages of these moments, and functions thereof. Trend removal is not sufficient to induce covariance stationarity, but is of first-order importance.

Before describing the three approaches, we note that it is common to work with logged versions of data represented in levels (e.g., as in figure 6.1). This is because changes in the log of a variable  $y_t$  over time represent the growth rate of the variable:

$$\frac{\partial}{\partial t} \log y_t = \frac{\frac{\partial}{\partial t} y_t}{y_t} \equiv \frac{\dot{y}_t}{y_t} \equiv g_{y_t}, \quad (6.1)$$

where  $\dot{y}_t = \frac{\partial}{\partial t} y_t$ . In addition, when using log-linear approximations to represent the corresponding structural model, working with logged versions of levels of the data provides symmetric treatment of both sets of variables.

The first two approaches to trend removal, detrending and differencing, are conducted under the implicit assumption that the data follow roughly constant growth rates. Detrending proceeds under the assumption that the level of  $y_t$  obeys

$$y_t = y_0(1 + g_y)^t e^{u_t}, \quad u_t \sim CSSP. \quad (6.2)$$

Then taking logs,

$$\log y_t = \log y_0 + g_y t + u_t, \quad (6.3)$$

where  $\log(1 + g_y)$  is approximated as  $g_y$ . Trend removal is accomplished by fitting a linear trend to  $\log y_t$  using an ordinary least squares (OLS) regression, and subtracting the estimated trend:

$$\tilde{y}_t = \log y_t - \hat{\alpha}_0 - \hat{\alpha}_1 t = \hat{u}_t, \quad (6.4)$$

where the  $\hat{\alpha}$ 's are coefficient estimates. In this case,  $\log y_t$  is said to be trend stationary.

In working with a set of  $m$  variables characterized by the corresponding model as sharing a common trend component (i.e., exhibiting balanced growth), symmetry dictates the removal of a common trend from all variables. Defining  $\alpha_1^j$  as the trend coefficient associated with variable  $j$ , this is accomplished via the imposition of the linear restrictions

$$\alpha_1^1 - \alpha_1^j = 0, \quad j = 2, \dots, m,$$

easily imposed in an OLS estimation framework.<sup>1</sup>

<sup>1</sup> The GAUSS procedure ct.prc, available at the textbook website, serves this purpose.

## 6.1 Removing Trends

Differencing proceeds under the assumption that  $y_t$  obeys

$$y_t = y_0 e^{\varepsilon_t}, \quad (6.5)$$

$$\varepsilon_t = \gamma + \varepsilon_{t-1} + u_t, \quad u_t \sim CSSP. \quad (6.6)$$

Note from (6.6) that iterative substitution for  $\varepsilon_{t-1}, \varepsilon_{t-2}, \dots$ , yields an expression for  $\varepsilon_t$  of the form

$$\varepsilon_t = \gamma t + \sum_{j=0}^{t-1} u_{t-j} + \varepsilon_0, \quad (6.7)$$

and thus the growth rate of  $y_t$  is given by  $\gamma$ . From (6.5),

$$\log y_t = \log y_0 + \varepsilon_t. \quad (6.8)$$

Thus the first difference of  $\log y_t$ , given by

$$\log y_t - \log y_{t-1} \equiv (1 - L) \log y_t,$$

where the lag operator  $L$  is defined such that  $L^p y_t = y_{t-p}$ , is stationary:

$$\begin{aligned} \log y_t - \log y_{t-1} &= \varepsilon_t - \varepsilon_{t-1} \\ &= \gamma + u_t. \end{aligned} \quad (6.9)$$

In this case,  $\log y_t$  is said to be difference stationary. Estimating  $\gamma$  using the sample average of  $\log y_t - \log y_{t-1}$  yields the desired transformation of  $y_t$ :

$$\tilde{y}_t = \log y_t - \log y_{t-1} - \widehat{\gamma} = \widehat{u}_t.$$

Once again, a common growth rate may be imposed across a set of variables via restricted OLS by estimating  $\widehat{\gamma}^j$  subject to the restriction<sup>2</sup>

$$\widehat{\gamma}^1 - \widehat{\gamma}^j = 0, \quad j = 2, \dots, m.$$

The choice between detrending versus differencing hinges on assumptions regarding whether (6.3) or (6.9) provides a more appropriate representation for  $\log y_t$ . Nelson and Plosser (1982) initiated an intense debate regarding this issue, and despite the large literature that has followed, the issue has proven difficult to resolve. (For overviews of this literature, see, e.g., DeJong and Whiteman, 1993; and Stock, 2004.) A remedy for this difficulty is to work with both specifications in turn, and evaluate the sensitivity of results to the chosen specification.

<sup>2</sup> The GAUSS procedure ct.prc is also available for this purpose.

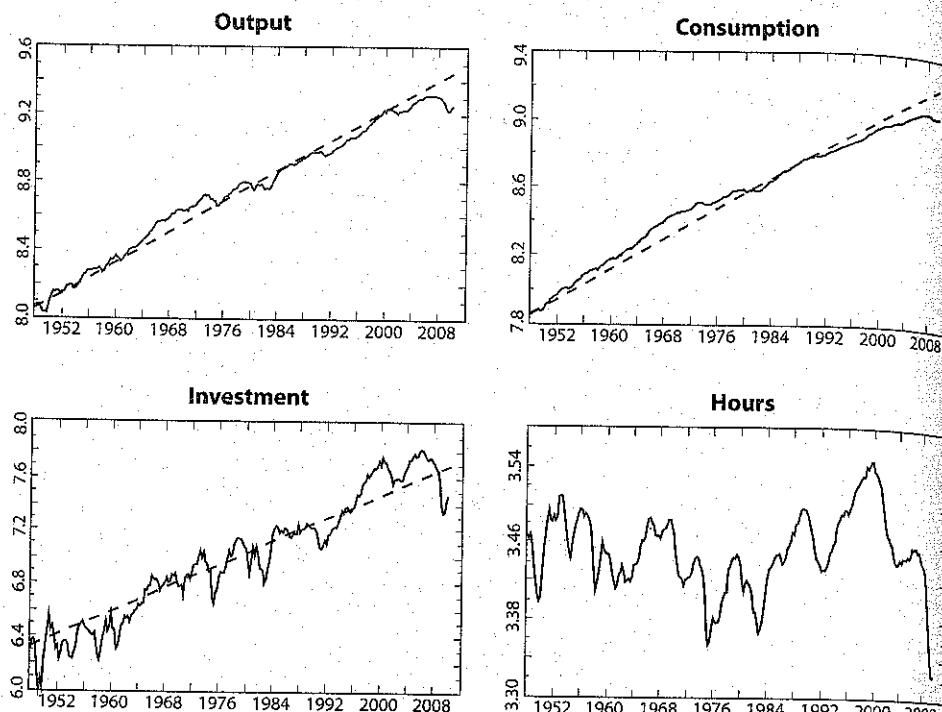


Figure 6.2 Logged Trajectories and Fitted Trends

As figure 6.2 illustrates, the choice of either specification is problematic in the present empirical context, because the data do not appear to follow a constant average growth rate throughout the sample period. The figure depicts the logged variables, along with fitted trends estimated for consumption, investment, and output (subject to the common-trend restriction). A trend was not fitted to hours, which as expected, does not exhibit trend behavior.

Note from the figure that consumption, investment, and output exhibit a distinct reduction in growth in approximately 1974, coinciding with the reduction in productivity observed during this period (for a discussion of this phenomenon, see Nordhaus, 2004). Note in particular the persistent tendency for consumption to lie above its estimated trend line over the first half of the sample period, and below its trend line during the second half of the period. This illustrates that if the series are truly trend stationary, but around a broken trend line, the detrended series will exhibit a spurious degree of persistence, tainting inferences regarding their cyclical behavior (see Perron, 1989 for a discussion of this issue). Likewise, the removal

### 6.1 Removing Trends

of a constant from first differences of the data will result in series that persistently lie above and below zero, also threatening to taint inferences regarding cyclicity.

The third approach to detrending involves the use of filters designed to separate trend from cycle, but given the admission of a slowly evolving trend. In this section we introduce the Hodrick-Prescott (H-P) filter, which has proven popular in business cycle applications. In section 3.2, we introduce a leading alternative to the H-P filter: the band pass filter.

Decomposing  $\log y_t$  as

$$\log y_t = g_t + c_t, \quad (6.10)$$

where  $g_t$  denotes the growth component of  $\log y_t$  and  $c_t$  denotes the cyclical component, the H-P filter estimates  $g_t$  and  $c_t$  in order to minimize

$$\sum_{t=1}^T c_t^2 + \lambda \sum_{t=3}^T [(1 - L)^2 g_t]^2, \quad (6.11)$$

taking  $\lambda$  as given.<sup>3</sup> Trend removal is accomplished simply as

$$\tilde{y}_t = \log y_t - \hat{g}_t = \hat{c}_t. \quad (6.12)$$

The parameter  $\lambda$  in (6.11) determines the importance of having a smoothly evolving growth component: the smoother is  $g_t$ , the smaller will be its second difference. With  $\lambda = 0$ , smoothness receives no value, and all variation in  $\log y_t$  will be assigned to the trend component. As  $\lambda \rightarrow \infty$ , the trend is assigned to be maximally smooth, that is, linear.

In general,  $\lambda$  is specified to strike a compromise between these two extremes. In working with business cycle data, the standard choice is  $\lambda = 1,600$ . To explain the logic behind this choice and what it accomplishes, it is necessary to venture into the frequency domain. Before doing so, we illustrate the trajectories of  $\hat{g}_t$  resulting from this specification for the example data, including hours. (In business cycle applications, it is conventional to apply the H-P filter to all series, absent a common-trend restriction.) These are presented in figure 6.3. The evolution of the estimated  $\hat{g}_t$ 's serves to underscore the mid-1970s reduction in the growth rates of consumption, investment, and output discussed above.

The versions of detrended output  $\tilde{y}_t$  generated by these three trend-removal procedures are illustrated in figure 6.4. Most striking is the difference in volatility observed across the three measures: the standard deviation of the linearly detrended series is 0.049, compared with 0.018

<sup>3</sup> The GAUSS procedure hpfilter.prc is available for this purpose.

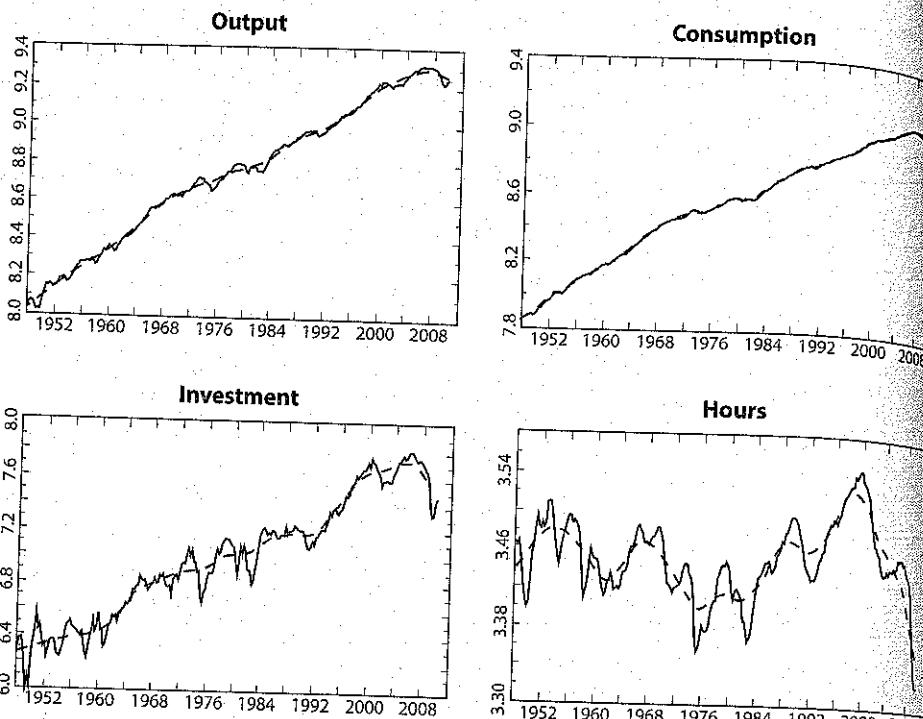


Figure 6.3 Logged Trajectories and H-P Trends

for the differenced series and 0.010 for the H-P filtered series. The behavior of the linearly detrended series is dominated by the large and extended departure above zero observed during the mid-1960s through the 1970s, and the subsequent reversal at the end of the sample. This behavior provides an additional indication of the trend break observed for this series in the mid-1970s. The correlation between the linearly detrended and differenced series is only 0.21; the correlation between the linearly detrended and H-P filtered series is 0.48; and the correlation between the H-P filtered and differenced series is 0.25.

## 6.2 Isolating Cycles

### 6.2.1 Mathematical Background

In venturing into the frequency domain, background information on complex variables is useful. Brown and Churchill (2003) and Palka (1991)

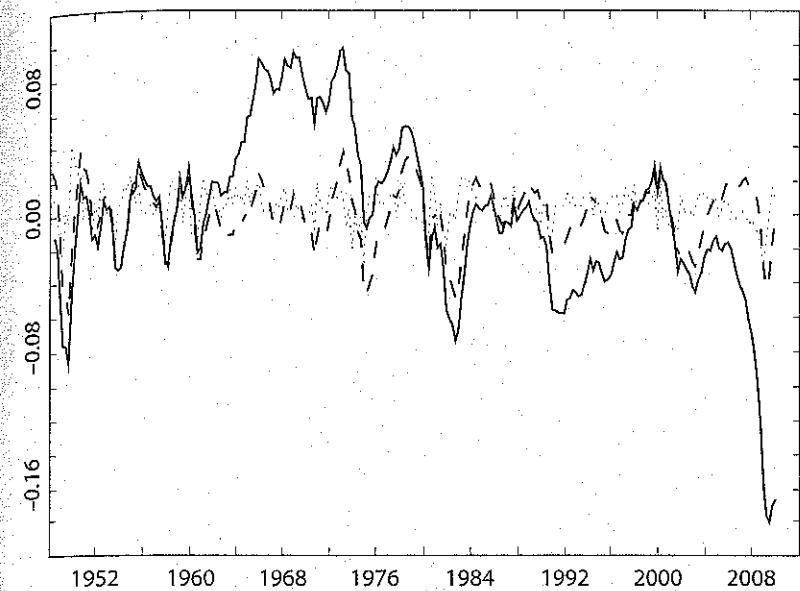


Figure 6.4 Detrended Output. Linearly detrended: solid; differenced: dots; H-P filtered: dashes.

are good undergraduate and graduate textbook sources of background information. And as noted above, expanded textbook treatments on the isolation of cycles are available from Sargent (1987a), Harvey (1993), Hamilton (1994), and Kaiser and Maravall (2001).

To help keep this section relatively self-contained, here we briefly sketch some key points regarding complex variables. Let  $i$  be imaginary, so that

$$\sqrt{i} = -1.$$

A variable  $z$  is complex if it can be represented as

$$z = x + iy,$$

where  $x$  and  $y$  are real;  $x$  is referenced as the real component of  $z$ , and  $y$  as the imaginary component. This representation of  $z$  is in terms of rectangular coordinates. In a graphical depiction of  $z$ , with the real component of  $z$  depicted on the horizontal axis and the imaginary component depicted on the vertical axis, the distance of  $z$  from the origin is given by

$$\begin{aligned} \sqrt{x^2 + y^2} &= \sqrt{(x + iy)(x - iy)} \\ &\equiv |z|. \end{aligned} \tag{6.13}$$

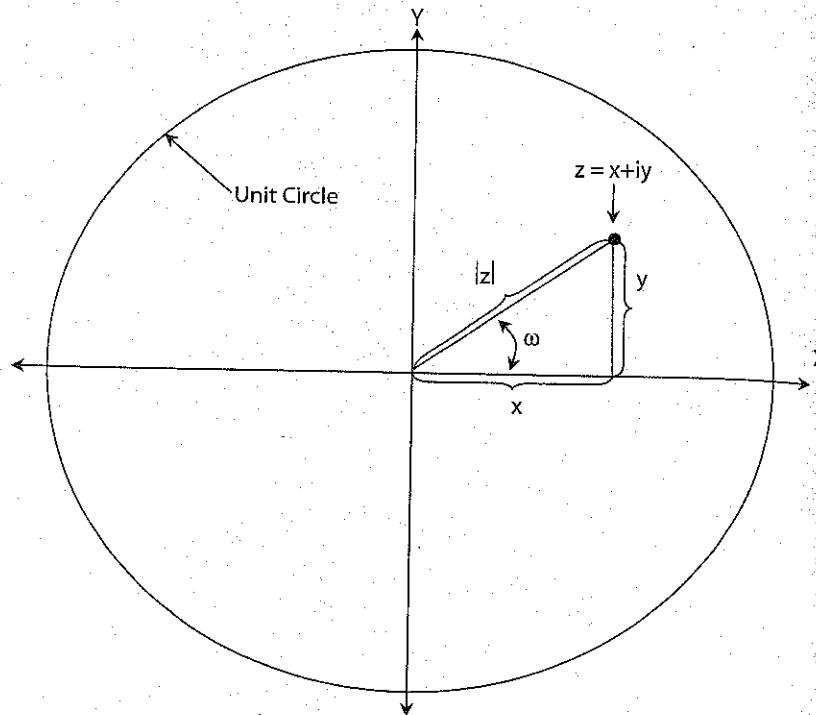


Figure 6.5 The Unit Circle

In (6.13),  $(x - iy)$  is known as the complex conjugate of  $z$ , and  $|z|$  is known as the modulus of  $z$ . If  $|z| = 1$ ,  $z$  is said to lie on the unit circle. See figure 6.5 for an illustration.

An additional representation of  $z$  is in polar coordinates. Let  $\omega$  denote the radian angle of  $z$  in  $(x, y)$  space; that is,  $\omega$  is the distance traveled counterclockwise along a circle starting on the  $x$  axis before reaching  $z$  (again, see figure 6.5 for an illustration). In terms of polar coordinates,  $z$  may be represented as

$$\begin{aligned} z &= |z|(\cos \omega + i \sin \omega) \\ &= |z|e^{i\omega}, \end{aligned} \quad (6.14)$$

where the second equality may be derived by taking Taylor Series expansions of  $\cos \omega$ ,  $\sin \omega$ , and  $e^{i\omega}$  about 0 and matching terms. Using (6.14), we obtain DeMoivre's Theorem:

$$\begin{aligned} z^j &= |z|^j e^{i\omega j} \\ &= |z|^j (\cos \omega j + i \sin \omega j). \end{aligned} \quad (6.15)$$

In addition to DeMoivre's Theorem, another important result from complex analysis we shall reference below comes from the Riesz-Fischer Theorem, which we now sketch. For a sequence of complex numbers  $\{a_j\}_{j=-\infty}^{\infty}$  that satisfy

$$\sum_{j=-\infty}^{\infty} |a_j|^2 < \infty, \quad (6.16)$$

known as a square-summability condition, there exists a complex function  $f(\omega)$  such that

$$f(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j}, \quad (6.17)$$

where  $\omega \in [-\pi, \pi]$ . Note from (6.14) that

$$e^{-i\omega} = (\cos \omega - i \sin \omega).$$

The construction of  $f(\omega)$  from  $\{a_j\}_{j=-\infty}^{\infty}$  is known as the Fourier transform of  $\{a_j\}_{j=-\infty}^{\infty}$ . Given  $f(\omega)$ , the elements of  $\{a_j\}_{j=-\infty}^{\infty}$  can be recovered using the inversion formula

$$a_j = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(\omega) e^{i\omega j} d\omega. \quad (6.18)$$

Finally, note that for any two functions  $f(\omega)$  and  $g(\omega)$ , where

$$f(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j}, \quad g(\omega) = \sum_{j=-\infty}^{\infty} b_j e^{-i\omega j},$$

we have

$$f(\omega) + g(\omega) = \sum_{j=-\infty}^{\infty} (a_j + b_j) e^{-i\omega j};$$

$$\alpha f(\omega) = \sum_{j=-\infty}^{\infty} \alpha a_j e^{-i\omega j}.$$

This establishes that the Fourier transform of the sum of sequences is the sum of their Fourier transforms, and that the Fourier transform of  $\{\alpha a_j\}_{j=-\infty}^{\infty}$  is  $\alpha$  times the Fourier transform of  $\{a_j\}_{j=-\infty}^{\infty}$ .

### 6.2.2 Cramér Representations

Consider the behavior of a time series  $y_t^\omega$  given by

$$y_t^\omega = \alpha(\omega) \cos(\omega t) + \beta(\omega) \sin(\omega t), \quad (6.19)$$

where  $\alpha(\omega)$  and  $\beta(\omega)$  are uncorrelated zero-mean random variables with equal variances. As above,  $\omega$  is measured in radians; here it determines the frequency with which  $\cos(\omega t)$  completes a cycle relative to  $\cos(t)$  as  $t$  evolves from 0 to  $2\pi$ ,  $2\pi$  to  $4\pi$ , and so on (the frequency for  $\cos(t)$  being 1). The upper panels of figure 6.6 depict  $\cos(\omega t)$  and  $\sin(\omega t)$  as  $t$  evolves from 0 to  $2\pi$  for  $\omega = 1$  and  $\omega = 2$ . Accordingly, given realizations for  $\alpha(\omega)$  and  $\beta(\omega)$ ,  $y_t^\omega$  follows a deterministic cycle that is completed  $\omega$  times as  $t$  ranges from 0 to  $2\pi$ , and so on. This is depicted in the lower panels of figure 6.6, using  $\alpha(\omega) = \beta(\omega) = 1$ ,  $\omega = 1$  and  $\omega = 2$ .

Consider now the construction of a time series  $y_t$  obtained by combining a continuum of  $y_t^\omega$ 's, differentiated by infinitesimal variations in  $\omega$  over the interval  $[0, \pi]$ :

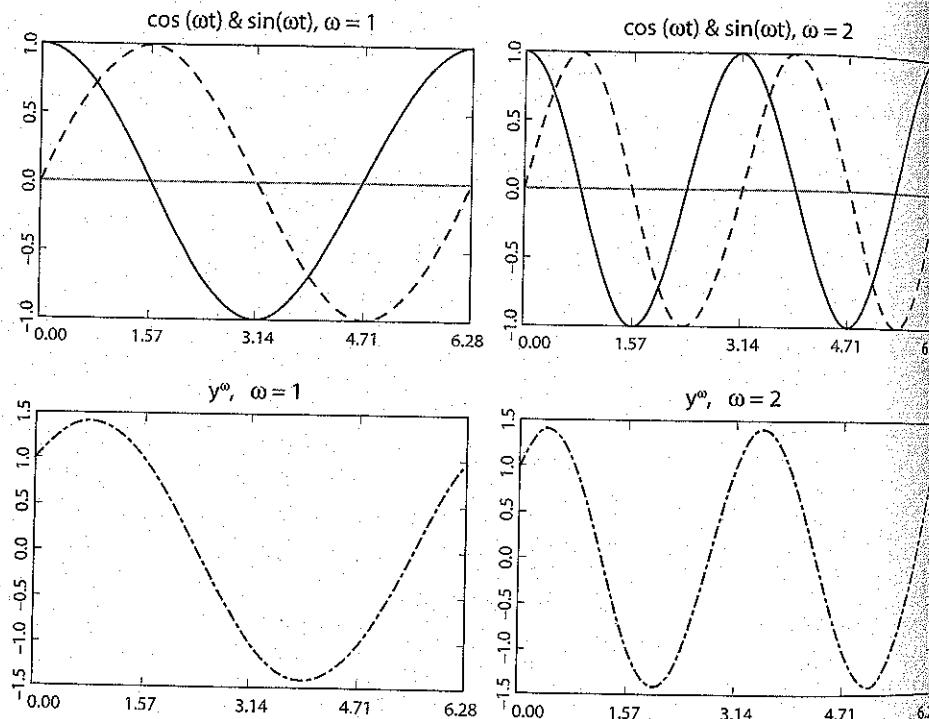


Figure 6.6 Evolution of  $\cos(\omega t)$  (solid),  $\sin(\omega t)$  (dashes), and  $y_t^\omega$  (bottom panels)

### 6.2 Isolating Cycles

$$y_t = \int_0^\pi \alpha(\omega) \cos(\omega t) d\omega + \int_0^\pi \beta(\omega) \sin(\omega t) d\omega. \quad (6.20)$$

(The symmetry of  $\cos(\omega t)$  and  $\sin(\omega t)$  between  $t \in [0, \pi]$  and  $t \in [\pi, 2\pi]$  renders the latter range as redundant.) Given appropriate specifications for  $\alpha(\omega)$  and  $\beta(\omega)$ , any time series  $y_t$  may be represented in this manner. This is referred to as the spectral representation, or Cramér representation, of  $y_t$ . It represents  $y_t$  as resulting from the combined influence of a continuum of cyclical components differing by the frequency with which they complete their cycles.

### 6.2.3 Spectra

Closely related to the spectral representation of  $y_t$  is its spectrum. This is a tool that measures the contribution to the overall variance of  $y_t$  made by the cyclical components  $y_t^\omega$  over the continuum  $[0, \pi]$ . Specifically, the spectrum is a decomposition of the variance of  $y_t$  by frequency.

To explain why, let  $\gamma(\tau)$  denote the autocovariance between  $y_t$  and  $y_{t+\tau}$  (or equivalently, between  $y_t$  and  $y_{t-\tau}$ ):

$$\gamma(\tau) = E(y_t - \mu_t)(y_{t+\tau} - \mu_{t+\tau}), \quad E(y_t) = \mu_t.$$

Note that  $\gamma(0)$  denotes the variance of  $y_t$ . So long as the sequence  $\{\gamma(\tau)\}_{\tau=-\infty}^{\infty}$  is square-summable, by the Riesz-Fischer Theorem, we may calculate its Fourier transform:

$$f_y(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau) e^{-i\omega\tau}. \quad (6.21)$$

Moreover, by the inversion formula,

$$\gamma(\tau) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f_y(\omega) e^{i\omega\tau} d\omega. \quad (6.22)$$

The power spectrum of  $y_t$  (spectrum hereafter) is defined as

$$s_y(\omega) = \frac{1}{2\pi} f_y(\omega). \quad (6.23)$$

From (6.22) and (6.23), note the sense in which the spectrum can be viewed as a decomposition of the variance of  $y_t$  by frequency: setting  $\tau = 0$ , the integral of  $s_y(\omega)$  over the range  $[-\pi, \pi]$  yields  $\gamma(0)$ , and comparisons of the height of  $s_y(\omega)$  for alternative values of  $\omega$  indicate the relative importance of fluctuations at the chosen frequencies in influencing variations in  $y_t$ .

For an alternative representation of the spectrum, note that DeMoivre's Theorem allows us to rewrite  $e^{-i\omega\tau}$  as  $(\cos \omega\tau - i \sin \omega\tau)$ . Thus combining (6.21) and (6.23), we obtain

$$s_y(\omega) = \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma(\tau) (\cos \omega\tau - i \sin \omega\tau). \quad (6.24)$$

Moreover, since  $\gamma(\tau) = \gamma(-\tau)$ ,  $\cos(0) = 1$ ,  $\sin(0) = 0$ ,  $\sin(-\omega) = -\sin \omega$ , and  $\cos(-\omega) = \cos \omega$ , (6.24) simplifies to

$$s_y(\omega) = \left( \frac{1}{2\pi} \right) \left[ \gamma(0) + 2 \sum_{\tau=1}^{\infty} \gamma(\tau) \cos(\omega\tau) \right]. \quad (6.25)$$

Because  $\cos(\omega\tau)$  is symmetric over  $[-\pi, 0]$  and  $[0, \pi]$ , so too is  $s_y(\omega)$ ; thus it is customary to represent  $s_y(\omega)$  over  $[0, \pi]$ .

To obtain an interpretation for frequency in terms of units of time rather than radians, it is useful to relate  $\omega$  to its associated period  $p$ , defined as the number of units of time necessary for  $y_t^\omega$  in (6.19) to complete a cycle:  $p = 2\pi/\omega$ . In turn,  $1/p = \omega/2\pi$  indicates the number of cycles completed by  $y_t^\omega$  per period. For example, with a period representing a quarter, a 10-year or 40-quarter cycle has an associated value of  $\omega$  of  $2\pi/40 = 0.157$ . For a 6-quarter cycle,  $\omega = 2\pi/6 = 1.047$ . Thus values for  $\omega$  in the range  $[0.157, 1.047]$  are of central interest in analyzing business cycle behavior.

#### 6.2.4 Using Filters to Isolate Cycles

Returning to the problem of trend removal, it is useful to think of a slowly evolving trend as a cycle with very low frequency; in the case of a constant trend, the associated frequency is zero. Filters are tools designed to eliminate the influence of cyclical variation at various frequencies. Detrending filters such as the first-difference and H-P filters target low frequencies; seasonal filters target seasonal frequencies; and so on.

The general form of a linear filter applied to  $y_t$ , producing  $y_t^f$ , is given by

$$y_t^f = \sum_{j=-r}^s c_j y_{t-j} \equiv C(L)y_t.$$

In other words, the filtered series  $y_t^f$  is a linear combination of the original series  $y_t$ . In the frequency domain, the counterpart to  $C(L)$  is obtained by replacing  $L^j$  with  $e^{-i\omega j}$ . The result is the frequency response function:  $C(e^{-i\omega})$ .

#### 6.2 Isolating Cycles

To gain an appreciation for how the filter works to isolate cycles, it is useful to derive the spectrum of  $y_t^f$ ; here we do so following Sargent (1987a). Suppose  $\{y_t\}$  is a mean-zero process with autocovariance sequence  $\{\gamma(\tau)\}_{\tau=-\infty}^{\infty}$ . The autocovariance between  $y_t^f$  and  $y_{t-\tau}^f$  is given by

$$\begin{aligned} E(y_t^f y_{t-\tau}^f) &= E \left( \sum_{j=-r}^s c_j y_{t-j} \right) \left( \sum_{k=-r}^s c_k y_{t-k-\tau} \right) \\ &= E \sum_{j=-r}^s \sum_{k=-r}^s c_j c_k y_{t-j} y_{t-k-\tau} \\ &= \sum_{j=-r}^s \sum_{k=-r}^s c_j c_k \gamma(t+j+k) \\ &\equiv \gamma_{yf}(\tau). \end{aligned} \quad (6.26)$$

Taking the Fourier transform of  $\gamma_{yf}(\tau)$ , the spectrum of  $y_t^f$  is given by

$$\begin{aligned} s_{yf}(\omega) &= \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \gamma_{yf}(\tau) e^{-i\omega\tau} \\ &= \frac{1}{2\pi} \sum_{\tau=-\infty}^{\infty} \sum_{j=-r}^s \sum_{k=-r}^s c_j c_k \gamma(t+j+k) e^{-i\omega\tau}. \end{aligned} \quad (6.27)$$

Let  $b = \tau + k - j$ , and rewrite  $e^{-i\omega\tau}$  in (6.27) as

$$\begin{aligned} e^{-i\omega\tau} &= e^{-i\omega(b+j-k)} \\ &= e^{-i\omega b} e^{-i\omega j} e^{i\omega k}. \end{aligned} \quad (6.28)$$

Finally, substituting for  $e^{-i\omega\tau}$  in (6.27) using (6.28), we obtain

$$\begin{aligned} s_{yf}(\omega) &= \frac{1}{2\pi} \sum_{j=-r}^s c_j e^{-i\omega j} \sum_{k=-r}^s c_k e^{i\omega k} \sum_{b=-\infty}^{\infty} \gamma(b) e^{-i\omega b} \\ &= \sum_{j=-r}^s c_j e^{-i\omega j} \sum_{k=-r}^s c_k e^{i\omega k} s_y(\omega) \\ &= C(e^{-i\omega}) C(e^{i\omega}) s_y(\omega), \end{aligned} \quad (6.29)$$

where the second equality stems from the definition of  $s_y(\omega)$ .

Before interpreting this expression, we introduce the gain function:

$$G(\omega) = |C(e^{-i\omega})|, \quad (6.30)$$

where  $|C(e^{-i\omega})|$  denotes the modulus of  $C(e^{-i\omega})$ :

$$|C(e^{-i\omega})| = \sqrt{C(e^{-i\omega})C(e^{i\omega})}. \quad (6.31)$$

For example, for the first-difference filter  $(1 - L)$ , the gain function is given by

$$\begin{aligned} G(\omega) &= \sqrt{(1 - e^{-i\omega})(1 - e^{i\omega})} \\ &= \sqrt{2}\sqrt{1 - \cos(\omega)}, \end{aligned} \quad (6.32)$$

where the second equality follows from the identity

$$e^{-i\omega} + e^{i\omega} = 2 \cos(\omega). \quad (6.33)$$

Given the definition of the gain function, the relationship between  $s_{y^f}(\omega)$  and  $s_y(\omega)$  in (6.29) can be written as

$$\begin{aligned} s_{y^f}(\omega) &= |C(e^{-i\omega})|^2 s_y(\omega) \\ &\equiv G(\omega)^2 s_y(\omega), \end{aligned} \quad (6.34)$$

where  $G(\omega)^2$  is referred to as the squared gain of the filter. This relationship illustrates how filters serve to isolate cycles: they attenuate or amplify the spectrum of the original series on a frequency-by-frequency basis. For example, note from (6.32) that the first-difference filter  $(1 - L)$  shuts down cycles of frequency zero.

### 6.2.5 The Hodrick-Prescott Filter

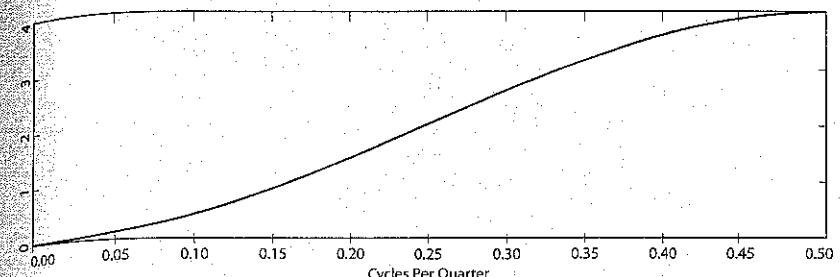
Regarding the H-P filter, the specification of  $\lambda$  determines the division of the influence of  $y_t^\omega$  on  $y_t$  between  $g_t$  and  $c_t$  in (6.10). Following Kaiser and Maravall (2001), its gain function is given by

$$G(\omega) = \left[ 1 + \left( \frac{\sin(\omega/2)}{\sin(\omega_0/2)} \right)^4 \right]^{-1}, \quad (6.35)$$

where

$$\omega_0 = 2 \arcsin \left( \frac{1}{2\lambda^{1/4}} \right). \quad (6.36)$$

### First-Difference Filter



### H-P Filter

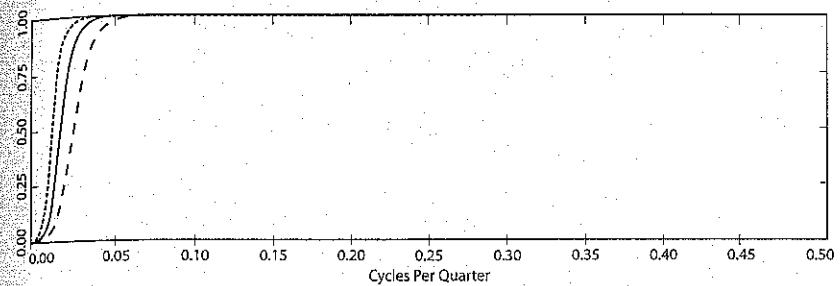


Figure 6.7 Squared Gains of the First-Difference and H-P Filters. Bottom panel:  $\lambda = 6,400$ : dots;  $\lambda = 1,600$ : solid;  $\lambda = 400$ : dashes.

The parameter  $\omega_0$ , selected through the specification of  $\lambda$ , determines the frequency at which  $G(\omega) = 0.5$ , or at which 50% of the filter gain has been completed. The specification  $\lambda = 1,600$  for quarterly data implies 50% completion at a 40-quarter cycle. The choice of  $\lambda = 400$  moves the 50% completion point to a 20-quarter cycle, and  $\lambda = 6,400$  to a 56-quarter cycle.

Squared gains  $G(\omega)^2$  associated with the first-difference and H-P filter (for the choices of  $\lambda$  highlighted above) are illustrated in figure 6.7. In all cases, the filters shut down zero-frequency fluctuations, and rise monotonically with frequency (reported hereafter in terms of cycles per quarter:  $\omega/2\pi$ ).

Although the first-difference and H-P filters are capable of eliminating trends, they are not designed to eliminate seasonal fluctuations. For quarterly data, seasonal frequencies correspond with  $\frac{1}{4}$  and  $\frac{1}{2}$  cycles per quarter, and the squared gains associated with each of these filters are positive at these values. As noted, business cycle models are typically not designed to explain seasonal variation, thus it is desirable to work with variables that have had seasonal variations eliminated.

### 6.2.6 Seasonal Adjustment

As with the example analyzed above, it is most often the case that aggregate variables are reported in seasonally adjusted (SA) form. Seasonal adjustment is typically achieved using the so-called X-11 filter (as characterized, e.g., by Bell and Monsell, 1992). So typically, seasonal adjustment is not an issue of concern in the preliminary stages of an empirical analysis. However, it is useful to consider this issue in order to appreciate the importance of the seasonal adjustment step; the issue also serves to motivate the introduction of the band pass filter, which provides an important alternative to the H-P filter.

As an illustration of the importance of seasonal adjustment, figure 6.8 presents the consumption series discussed above (here running through 2004:IV), along with its nonseasonally adjusted (NSA) counterpart (including H-P trends for both). Trend behavior dominates both series, but the recurrent seasonal spikes associated with the NSA series are distinctly apparent. The spikes are even more apparent in figure 6.9, which presents the H-P filtered series.

Figure 6.10 presents spectra estimated for both versions of the H-P filtered data (chapter 7, section 2 presents methods for estimating spectra).

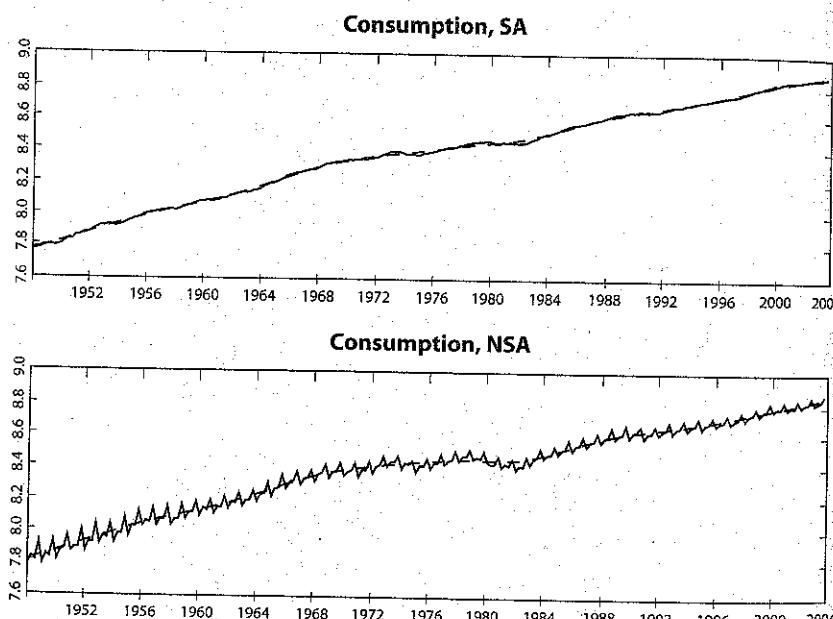
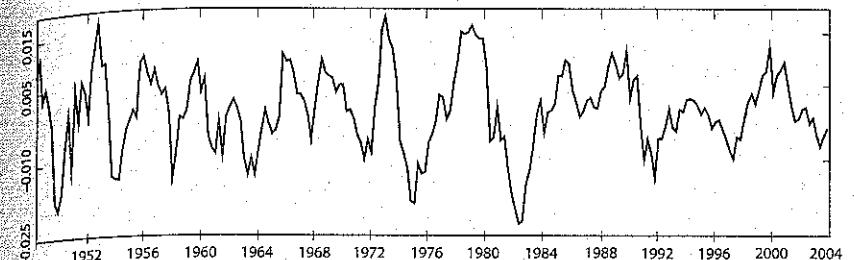


Figure 6.8 Seasonally Adjusted and Nonseasonally Adjusted Consumption

### 6.2 Isolating Cycles

#### H-P Filtered Consumption, SA



#### H-P Filtered Consumption, NSA

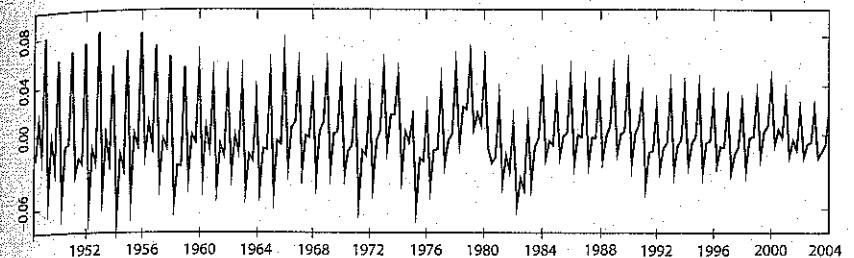


Figure 6.9 H-P Filtered Consumption

The bottom panel of the figure truncates the seasonal spikes associated with the spectrum of the NSA series to better illustrate the extent to which the seasonal fluctuations dominate the contribution of business cycle fluctuations to the overall variance of the series (recall that business cycle fluctuations lie roughly between 1/40 and 1/6 cycles per quarter).

### 6.2.7 Band Pass Filters

We turn now to the band pass (B-P) filter. This is a filter designed to shut down all fluctuations outside of a chosen frequency band. Given an interest in cycles with periods between  $p_l$  and  $p_u$  (again, roughly between 6 and 40 quarters in business cycle applications), the ideal B-P filter has a squared gain that satisfies

$$G(\omega)^2 = \begin{cases} 1, & \omega \in [2\pi/p_u, 2\pi/p_l] \\ 0, & \text{otherwise.} \end{cases} \quad (6.37)$$

As (6.38) below indicates, it is not feasible to implement the ideal B-P filter, because doing so requires as input an infinite number of observations of the unfiltered series. However, several approaches to estimating

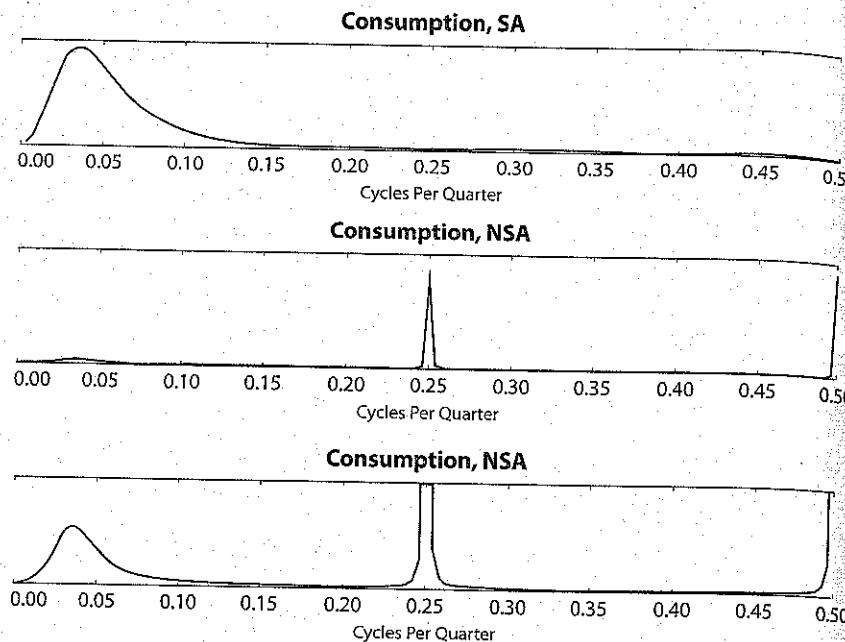


Figure 6.10 Spectra of H-P Filtered Consumption. Note: The third panel zooms in on the middle panel.

approximate B-P filters have been proposed. Here, we present the approach developed by Baxter and King (1999); for alternatives, for example, see Woitek (1998) and Christiano and Fitzgerald (1999).<sup>4</sup>

Let the ideal symmetric B-P filter for a chosen frequency range be given by

$$\alpha(L) = \sum_{j=-\infty}^{\infty} \alpha_j L^j, \quad (6.38)$$

where symmetry implies  $\alpha_{-j} = \alpha_j \forall j$ . This is an important property for filters because it avoids inducing what is known as a phase effect. Under a phase effect, the timing of events between the unfiltered and filtered series, such as the timing of business cycle turning points, will be altered. The Fourier transformation of a symmetric filter has a very simple form. In the present case,

<sup>4</sup> The collection of GAUSS procedures contained in `bpx.sxc` is available for constructing Baxter and King's B-P filter.

## 6.2 Isolating Cycles

$$\begin{aligned} \alpha(e^{-i\omega}) &\equiv \alpha(\omega) = \sum_{j=-\infty}^{\infty} \alpha_j e^{-i\omega j} \\ &= \alpha_0 + \sum_{j=1}^{\infty} \alpha_j (e^{-i\omega j} + e^{i\omega j}) \\ &= \alpha_0 + 2 \sum_{j=1}^{\infty} \alpha_j \cos(\omega), \end{aligned} \quad (6.39)$$

where the second equality follows from symmetry and the last equality results from (6.33).

Baxter and King's approximation to  $\alpha(\omega)$  is given by the symmetric, finite-ordered filter

$$A(\omega) = \alpha_0 + 2 \sum_{j=1}^K \alpha_j \cos(\omega), \quad (6.40)$$

where

$$A(0) = \sum_{j=-K}^K \alpha_j = 0,$$

ensuring that  $A(\omega)$  is capable of removing a trend from the unfiltered series (see their appendix A for details).  $A(\omega)$  is chosen to solve

$$\min_{\alpha_j} \int_{-\pi}^{\pi} |\alpha(\omega) - A(\omega)|^2 d\omega \quad \text{subject to } A(0) = 0; \quad (6.41)$$

that is,  $A(\omega)$  minimizes departures from  $\alpha(\omega)$  (measured in squared-error sense) accumulated over frequencies. The solution to this objective is given by

$$\begin{aligned} \alpha_j &= \alpha_j + \theta, \quad j = -K, \dots, K; \\ \alpha_j &= \begin{cases} \frac{\omega_u - \omega_l}{\pi}, & j = 0 \\ \frac{\sin(\omega_2 j) - \sin(\omega_1 j)}{\pi j}, & j = \pm 1, \dots, K; \end{cases} \\ \theta &= \frac{-\sum_{j=-K}^K \alpha_j}{2K+1}, \end{aligned} \quad (6.42)$$

where  $\omega_l = 2\pi/p_u$  and  $\omega_u = 2\pi/p_l$ .

Baxter and King propose the selection of  $K = 12$  in working with quarterly data, entailing the loss of 12 filtered observations at the beginning

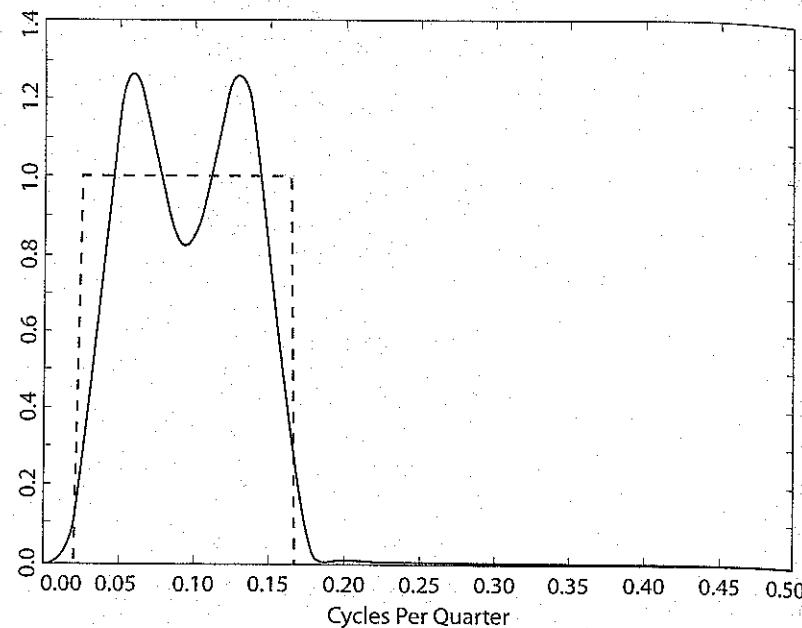


Figure 6.11 Squared Gains of Band Pass Filter and Optimal Filter. B-P filter: solid; optimal filter: dashes.

and end of the sample period. Figure 6.11 illustrates the squared gains associated with the ideal and approximated B-P filters constructed over the 1/40 and 1/6 cycles per quarter range.

Application of the B-P filter constructed using  $K = 12$  to the SA and NSA consumption data produces the smoothed series illustrated in figure 6.12. The series exhibit close correspondence, with no discernible trends or seasonal variations. The spectra estimated for these series, illustrated in figure 6.13, confirm the absence of both influences on the variations in the series.

### 6.3 Spuriousness

We conclude this chapter with some words of caution. The common presence of trends in macroeconomic time series requires the completion of a preliminary trend-removal step in most empirical applications. But because it is difficult to convincingly establish the precise nature of the driving process that gives rise to trend behavior, and the separation of trend from cycle can also pose challenges, it is difficult to ensure that appropriate steps have

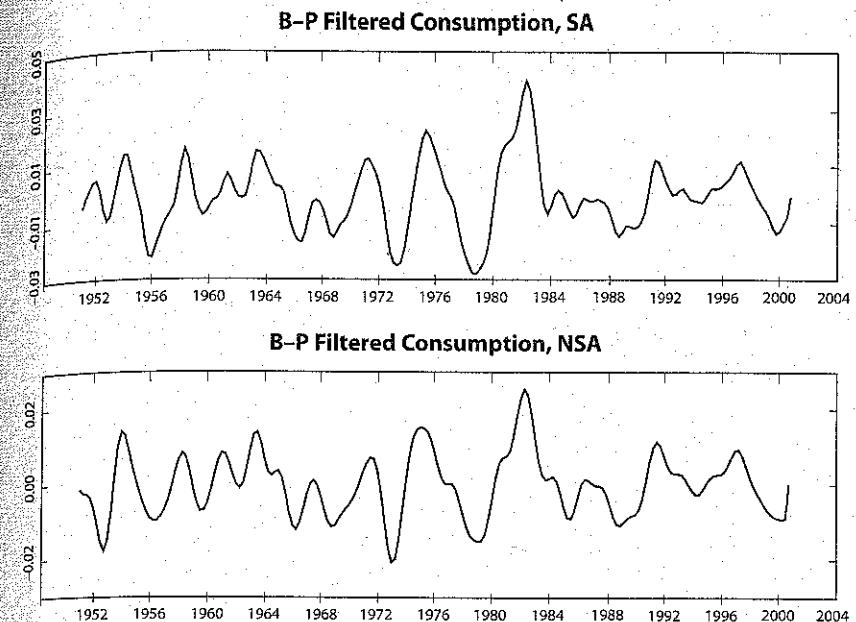


Figure 6.12 B-P Filtered Consumption

been taken in this preliminary stage. Unfortunately, this is an important issue due to a common problem involving spurious stochastic behavior.

In general, spuriousness is used to characterize situations in which the stochastic behavior of a filtered variable differs systematically from its unfiltered counterpart along the dimension of original interest in the empirical analysis. Of course, the stochastic behavior of the two series will differ in general, but for example, if the removal of a trend induces systematic differences in the business cycle properties of filtered variables, spuriousness is said to have been induced.

Spuriousness can arise both in removing trends and isolating cycles. Regarding the latter, consider the extreme but illustrative case in which an H-P or B-P filter is applied to a collection of zero-mean serially uncorrelated CSSPs. Such CSSPs are referred to as white noise: their spectra are uniform. In this case, spectra of the filtered series will identically assume the shape of the squared gains of the filters, and thus the filtered series will exhibit spurious cyclical behavior. Harvey and Jaeger (1993) provide an analysis of spurious behavior arising from H-P filtered data.

Regarding trend removal, we have seen that the removal of fixed trends from the levels of series that have evidently undergone trend breaks can induce considerable persistence in the detrended series. So even if the

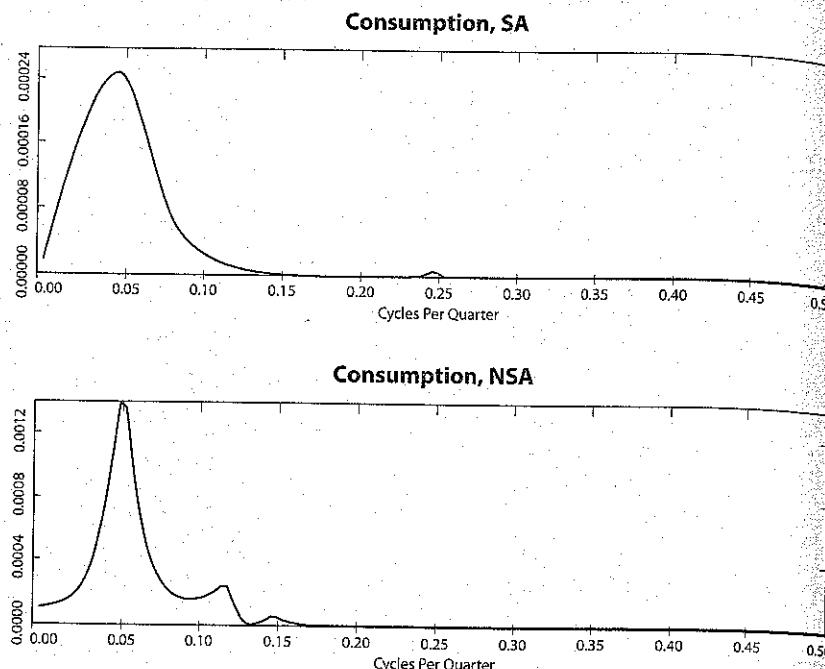


Figure 6.13 Spectra of B-P Filtered Consumption

underlying data were trend-stationary, the application of a fixed trend specification in this case would induce undue persistence in the detrended series. Moreover, as noted, it is difficult to distinguish between trend- and difference-stationary specifications even given the ideal case of constant average growth over the sample period. And as shown by Chan, Hayya, and Ord (1977) and Nelson and Kang (1981), both the removal of a deterministic trend from a difference-stationary specification and the application of the difference operator  $(1 - L)$  to a trend-stationary process induces spurious autocorrelation in the resulting series. Similarly, Cogley and Nason (1995) and Murray (2003) illustrate spuriousness arising from the application of the H-P and B-P filters to nonstationary data.

Having painted this bleak picture, we conclude by noting that steps are available for helping to mitigate these problems. For example, regarding the trend- versus difference-stationarity issue, although it is typically difficult to reject either specification in applications of classical hypothesis tests to macroeconomic time series, it is possible to obtain conditional inferences regarding their relative plausibility using Bayesian methods. Such

### 6.3 Spuriousness

inferences can be informative in many instances. (See DeJong and White- man, 1991a, 1991b, for examples in macroeconomic applications; and Phillips, 1991, for skepticism. Tools for implementing Bayesian methods in general are presented in chapter 14.) And the use of alternative filtering methods in a given application is a useful way to investigate the robustness of inferences to steps taken in this preliminary stage.

# Chapter 7 - Structural Macroeconomics

## Summarizing Time Series Behavior When All Variables Are Observable

The sign of a truly educated man is to be deeply moved by statistics.

—George Bernard Shaw

Empirical implementations of DSGE models often entail the comparison of theoretical predictions regarding collections of summary statistics with empirical counterparts. Such comparisons are broadly characterized as limited-information analyses, and encompass as special cases calibration exercises (the subject of chapter 11) and moment-matching exercises (the subject of chapter 12).

Here we present two workhorse reduced-form models that provide flexible characterizations of the time-series behavior of the  $m \times 1$  collection of observable variables  $X_t$ . We then present a collection of summary statistics that frequently serve as targets for estimating the parameters of structural models, and as benchmarks for judging their empirical performance. The statistics are all calculable as functions of the parameters of the reduced-form models we present. Finally, we briefly describe how theoretical predictions regarding the same set of statistics may be obtained, setting the stage for the detailed descriptions provided in chapters 11 and 12. Continuing with the roadmap, we then go on in chapter 8 to present state-space models, which characterize the time-series behavior of the  $n \times 1$  collection of variables  $x_t$ , which includes unobservable variables that map into the observables via

$$X_t = H'x_t + u_t,$$

where  $u_t$  represents measurement errors. State-space representations provide the foundation upon which full-information (or likelihood-bases) analyses of DSGE models are based. Full-information analyses of these models are then described in detail in chapters 13 and 14.

We take as granted here the completion of the data-preparation stage described in chapter 6, involving the removal of trends and the isolation

### 7.1 Two Useful Reduced-Form Models

139

of cycles. This implies that the variables contained in  $X_t$  are mean-zero covariance stationary stochastic processes (CSSPs) exhibiting fluctuations isolated to desired frequencies. Anticipating the mapping of these data to their theoretical counterparts, these data represent deviations of the observables from steady state values.

Hereafter we employ the following notation. A univariate time series considered in isolation is denoted as  $y_t$ . The variable  $X_{it}$  denotes the  $i^{\text{th}}$  element of  $X_t$ . The parameters of a given reduced-form model are collected in the vector  $\vartheta$ , and a generic function of these parameters is denoted as  $g(\vartheta)$ . The corresponding generic function of the parameters of a given structural model is denoted as  $g(\mu)$ .

As with chapter 6, supplements to the brief coverage of these topics provided here are available from any number of texts devoted to time series analysis (e.g., Harvey, 1993; Hamilton, 1994). Also, the example business cycle data set introduced in chapter 6 is used to illustrate the material presented in this chapter.

### 7.1 Two Useful Reduced-Form Models

#### 7.1.1 The ARMA Model

The foundation of the models presented in this section is a mean-zero covariance-stationary stochastic process  $\{\varepsilon_t\}$ , which is serially uncorrelated at all leads and lags:  $E(\varepsilon_t \varepsilon_s) = 0$ ,  $t \neq s$ . The variance of  $\varepsilon_t$  is denoted as  $\sigma^2$ . Recall from chapter 6 that such a process is referred to as white noise, due to the shape of its spectrum. The variable  $y_t$  constructed as

$$\begin{aligned} y_t &= \varepsilon_t + \theta \varepsilon_{t-1} \\ &\equiv (1 - \theta L)\varepsilon_t \end{aligned} \tag{7.1}$$

is said to follow a moving average process of order 1, or MA(1) process: it is a moving average of the two most recent observations of  $\varepsilon_t$ . The variance of  $y_t$ , denoted as  $\sigma_y^2$  or  $\gamma(0)$ , is given by

$$\begin{aligned} \sigma_y^2 &= E(y_t^2) = E(\varepsilon_t^2 + 2\theta\varepsilon_t\varepsilon_{t-1} + \theta^2\varepsilon_{t-1}^2) \\ &= (1 + \theta^2)\sigma^2, \end{aligned} \tag{7.2}$$

and its autocovariance pattern is simply

$$\begin{aligned} \gamma(1) &= E(y_t y_{t-1}) = E(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2}) \\ &= \theta\sigma^2; \\ \gamma(s) &= E(y_t y_{t-s}) = 0, \quad s > 1. \end{aligned} \tag{7.3}$$

Denoting the  $s^{\text{th}}$ -order autocorrelation of  $y_t$  as  $\varphi(s) = \gamma(s)/\gamma(0)$ , the corresponding autocorrelation pattern is

$$\begin{aligned}\varphi(1) &= \frac{\theta}{(1+\theta^2)}; \\ \varphi(s) &= 0, \quad s > 1.\end{aligned}\quad (7.4)$$

Thus the impact of  $\varepsilon_t$  on  $\{y_t\}$  persists one period beyond its initial realization, imparting first-order serial correlation in  $\{y_t\}$ .

An MA( $q$ ) process specified for  $y_t$  generalizes to

$$\begin{aligned}y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \\ &= \sum_{j=0}^q \theta_j \varepsilon_{t-j} \\ &= \sum_{j=0}^q (\theta_j L^j) \varepsilon_t \\ &\equiv \theta(L) \varepsilon_t,\end{aligned}\quad (7.5)$$

where  $\theta_0 = 1$ . The variance of  $y_t$  in this case is given by

$$\sigma_y^2 = \sigma^2 \sum_{j=0}^q \theta_j^2, \quad (7.6)$$

and its covariance pattern is

$$\gamma(s) = \left\{ \begin{array}{ll} \sigma^2 [\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \cdots + \theta_q\theta_{q-s}], & s = 1, \dots, q \\ 0 & s > q. \end{array} \right\} \quad (7.7)$$

Note that the persistence of  $\varepsilon_t$  is limited to the horizon corresponding with  $q$ .

Finally, an infinite-order MA process specified for  $y_t$  is given by

$$\begin{aligned}y_t &= \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j} \\ &\equiv \psi(L) \varepsilon_t.\end{aligned}\quad (7.8)$$

The Wold Decomposition Theorem holds that any CSSP may be represented as in (7.8), where

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty, \quad (7.9)$$

### 7.1 Two Useful Reduced-Form Models

with the white noise process  $\{\varepsilon_t\}$  representing a sequence of one-step-ahead forecast errors:

$$\varepsilon_t = y_t - E_{t-1}(y_t | y_{t-1}, y_{t-2}, \dots). \quad (7.10)$$

The expectations operator  $E_{t-1}(\cdot | \Omega)$  is conditional on information contained in  $\Omega$  available at time  $t-1$ . The condition (7.9) is referred to as square summability. It is necessary to ensure a finite variance and covariances for  $y_t$ :

$$\sigma_y^2 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2; \quad (7.11)$$

$$\gamma(s) = \sigma^2 [\psi_s \psi_0 + \psi_{s+1} \psi_1 + \psi_{s+2} \psi_2 + \dots]. \quad (7.12)$$

Consider now a specification for  $y_t$  of the form

$$y_t = \rho y_{t-1} + \varepsilon_t, \quad |\rho| < 1. \quad (7.13)$$

In this case,  $y_t$  is said to follow an autoregressive process of order 1, or AR(1) process. To derive the variance and covariance pattern for  $y_t$  implied by (7.13), it is useful obtain an expression for  $y_t$  in terms of  $\{\varepsilon_t\}$ . One way to do so is through recursive substitution. This begins by substituting for  $y_{t-1}$  in (7.13), yielding

$$\begin{aligned}y_t &= \rho(\rho y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t \\ &= \varepsilon_t + \rho \varepsilon_{t-1} + \rho^2 y_{t-2}.\end{aligned}$$

Repeated substitution for  $y_{t-2}, y_{t-3}$ , and so on, yields

$$y_t = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}. \quad (7.14)$$

Notice the role played by the restriction  $|\rho| < 1$ : this ensures satisfaction of the square-summability condition; thus (7.14) constitutes a Wold Representation. For the special case in which  $\rho = 1$ ,  $(1-L)y_t = \varepsilon_t$ , and thus  $y_t$  is said to be difference-stationary. Square-summability is violated in this case, and the variance of  $y_t$  does not exist. In this special case  $y_t$  is also said to be integrated, since from (7.14)  $y_t$  is obtained by integrating over the set of realizations of  $\{\varepsilon_t\}$  observed up to time  $t$ .

Given the restriction  $|\rho| < 1$ , the expressions for  $\sigma_y^2$  and  $\gamma(s)$  in (7.11) and (7.12) specialize to

$$\sigma_y^2 = \frac{\sigma^2}{1 - \rho^2}; \quad (7.15)$$

$$\gamma(s) = \rho^s \frac{\sigma^2}{1 - \rho^2}, \quad s = 1, 2, \dots \quad (7.16)$$

Correspondingly,  $\varphi(s) = \rho^s$ . From both (7.14) and (7.16), it is evident that  $\rho$  determines the persistence of  $\varepsilon_t$ : the nearer is  $\rho$  to 1, the stronger is the influence of  $\varepsilon_{t-j}$  on  $y_t$ .

Before generalizing beyond the AR(1) case, it is useful to consider a more direct approach to the derivation of (7.14). Rewriting (7.13) as

$$y_t = \rho L y_t + \varepsilon_t,$$

(7.14) may be derived as follows:

$$\begin{aligned} (1 - \rho L) y_t &= \varepsilon_t, \\ y_t &= \frac{1}{(1 - \rho L)} \varepsilon_t \\ &= \sum_{j=0}^{\infty} (\rho^j L^j) \varepsilon_t \\ &= \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}. \end{aligned}$$

Think of  $(1 - \rho L)$  as a first-order polynomial in the lag operator  $L$ . The root of this polynomial is  $1/\rho$ , which lies outside the unit circle given that  $|\rho| < 1$ . This is a necessary condition for the stationarity of  $y_t$ . Returning again to the special case in which  $\rho = 1$ , the root of the polynomial is unity, and in this case  $y_t$  is said to follow a unit root process.

Generalizing to the AR( $p$ ) case,  $y_t$  is given by

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \dots + \rho_p y_{t-p} + \varepsilon_t, \quad (7.17)$$

or

$$(1 - \rho_1 L - \dots - \rho_p L^p) y_t \equiv \rho(L) y_t = \varepsilon_t. \quad (7.18)$$

Factoring  $(1 - \rho_1 L - \dots - \rho_p L^p)$  as  $(1 - \lambda_1 L)(1 - \lambda_2 L) \dots (1 - \lambda_p L)$ ,  $y_t$  may be expressed as

$$\begin{aligned} y_t &= \left( \frac{1}{\rho(L)} \right) \varepsilon_t \\ &= \left( \frac{1}{1 - \lambda_1 L} \right) \left( \frac{1}{1 - \lambda_2 L} \right) \dots \left( \frac{1}{1 - \lambda_p L} \right) \varepsilon_t \\ &= \psi(L) \varepsilon_t. \end{aligned} \quad (7.19)$$

### 7.1 Two Useful Reduced-Form Models

So long as the roots  $1/\lambda_j, j = 1, \dots, p$  of the polynomial  $\rho(L)$  lie outside the unit circle,  $\psi(L)$  is square-summable and (7.19) constitutes a Wold Representation.

The persistence of  $\varepsilon_t$  is determined by the proximity of the smallest root to unity: the closer to unity, the greater is the associated value of  $\lambda$ , and thus the greater is the persistence. If  $d$  roots lie exactly on the unit circle, then  $d$  applications of the difference operator to  $y_t$ , denoted as  $(1 - L)^d y_t$ , will cancel the  $d$  terms  $\left(\frac{1}{1-L}\right)$  on the right-hand side of (7.19), and thus  $(1 - L)^d y_t$  will be stationary. In this case,  $y_t$  is said to be integrated of order  $d$ , or an  $I(d)$  process.

A means for deriving the coefficients of  $\psi(L)$  (and thus expressions for  $\sigma_y^2$  and  $\gamma(s)$  using (7.11) and (7.12)) as functions of the coefficients of  $\rho(L)$  is provided by the method of undetermined coefficients. This begins by combining (7.8) and (7.18) to obtain

$$\rho(L)^{-1} \varepsilon_t = \psi(L) \varepsilon_t,$$

implying

$$\begin{aligned} 1 &= \rho(L) \psi(L) \\ &= (1 - \rho_1 L - \dots - \rho_p L^p)(\psi_0 + \psi_1 L + \psi_2 L^2 + \dots). \end{aligned} \quad (7.20)$$

Both sides of (7.20) may be thought of as infinite-order polynomials in  $L$ . Equating the coefficients associated with  $L^0, L^1, \dots$  on both sides of (7.20) yields the following system of equations:

$$\begin{aligned} 1 &= \psi_0 \\ 0 &= \psi_1 - \rho_1 \Rightarrow \psi_1 = \rho_1 \\ 0 &= \psi_2 - \rho_1 \psi_1 - \rho_2 \Rightarrow \psi_2 = \rho_1^2 + \rho_2 \\ &\dots \\ 0 &= \psi_p - \rho_1 \psi_{p-1} - \rho_2 \psi_{p-2} - \dots - \rho_p \psi_0 \\ 0 &= \psi_{p+j} - \rho_1 \psi_{p+j-1} - \rho_2 \psi_{p+j-2} - \dots - \rho_p \psi_j, \quad j = 1, 2, \dots \end{aligned} \quad (7.21)$$

#### Exercise 7.1

Derive expressions for  $\psi_j, j = 0, 1, 2, \dots$  as functions of the coefficients of  $\rho(L)$  for  $p = 1, 2$ , and 3.

The specifications for  $y_t$  outlined above are encompassed by an ARMA( $p, q$ ) model, expressed as

$$(1 - \rho_1 L - \dots - \rho_p L^p) y_t = (1 + \theta_1 L + \dots + \theta_q L^q) \varepsilon_t, \quad \text{or} \quad (7.22)$$

$$\alpha(I) v_t = \beta(I) \varepsilon_t$$

Stationarity is once again determined by the roots of  $\rho(L)$ . Assuming these lie outside the unit circle, the inversion of  $\rho(L)$  yields the Wold Representation

$$y_t = \left( \frac{\theta(L)}{\rho(L)} \right) \varepsilon_t. \quad (7.23)$$

Combining (7.8) and (7.23) yields the relationship

$$\theta(L) = \rho(L)\psi(L), \quad (7.24)$$

which generalizes (7.20). Accordingly, the method of undetermined coefficients may be applied to (7.24) to obtain expressions for  $\psi_j, j = 0, 1, 2, \dots$  as functions of the coefficients of  $\theta(L)$  and  $\rho(L)$ , again yielding expressions for  $\sigma_y^2$  and  $\gamma(s)$  via (7.11) and (7.12).

### Exercise 7.2

Derive expressions for  $\psi_j, j = 0, 1, 2, \dots$  as functions of the coefficients of  $\rho(L)$  and  $\theta(L)$  for  $p = 1, q = 1$ .

A numerical algorithm for implementing the method of undetermined coefficients involves the construction of a hypothetical realization of  $\{y_t\}_{t=0}^\infty$  resulting from a special realization of innovations  $\{\varepsilon_t\}_{t=0}^\infty$ . Specifically, let  $\{y_{-(p+1)}, y_{-(p)}, \dots, y_{-(1)}\} = 0, \varepsilon_0 = 1$ , and  $\{\varepsilon_t\}_{t=1}^\infty = 0$ . Then from (7.8), the resulting sequence  $\{y_t\}_{t=0}^\infty$  is identically  $\{\psi_j\}_{j=0}^\infty$ . From (7.22), the sequence  $\{\psi_j\}_{j=0}^\infty$  may thus be constructed iteratively as follows:

$$\begin{aligned} \psi_0 &= 1 \\ \psi_1 &= \rho_1 + \theta_1 \\ \psi_2 &= \rho_1 \psi_1 + \rho_2 + \theta_2 \\ &\vdots \\ \psi_p &= \rho_1 \psi_{p-1} + \rho_2 \psi_{p-2} + \dots + \rho_p + \theta_p \\ &\vdots \\ \psi_j &= \rho_1 \psi_{j-1} + \rho_2 \psi_{j-2} + \dots + \rho_p \psi_{j-p} + \theta_j, \end{aligned} \quad (7.25)$$

where  $\theta_j = 0$  for  $j > q$ . A plot of the resulting sequence  $\{\psi_j\}_{j=0}^\infty$  is referred to as an impulse response function, as it traces out the response of  $y_t$  to the realization of a representative shock.

We conclude with a brief discussion of estimation. Note from (7.17) that the coefficients of an AR( $p$ ) model may be estimated in straightforward fashion using an ordinary least squares (OLS) regression. In the

regression, the  $(T - p) \times 1$  vector  $y = [y_{p+1}, y_{p+2}, \dots, y_T]'$  serves as the dependent variable, and  $Ly, L^2y, \dots, L^p y$  serve as independent variables, where  $L^j y = [y_{p+1-j}, y_{p+2-j}, \dots, y_{T-j}]'$ , and  $T$  denotes the total number of observations of  $\{y_t\}$ . Moreover, given the assumption of Normality for  $\{\varepsilon_t\}$ , the resulting OLS estimates of  $\rho(L)$  coincide with conditional maximum likelihood (ML) estimates.

Estimation is not as straightforward when the specification for  $y_t$  contains an MA component. Complication arises from the presence of unobservable variables as explanatory variables for  $y_t$ : namely, lagged values of  $\varepsilon_t$ . But as demonstrated in chapter 8, this complication may be overcome via use of the Kalman filter.<sup>1</sup>

### 7.1.2 Allowing for Heteroskedastic Innovations

While the foregoing presentation has characterized the variance of  $\varepsilon_t$  as time-invariant, it is straightforward to incorporate generalizations that allow for heteroskedasticity. Indeed, suppose that the square of  $\varepsilon_t$  follows an autoregressive process, say, of order  $r$ :

$$\varepsilon_t^2 = \tau + \rho_{\varepsilon 1} \varepsilon_{t-1}^2 + \rho_{\varepsilon 2} \varepsilon_{t-2}^2 + \dots + \rho_{\varepsilon r} \varepsilon_{t-r}^2 + u_t, \quad (7.26)$$

where  $u_t$  is a white-noise process with variance  $\sigma_u^2$ . Then  $\varepsilon_t$  is defined as an autoregressive conditionally heteroskedastic (ARCH) process of order  $r$ , following Engle (1982).

An alternative representation of an ARCH process is useful for characterizing a leading extension. Representing  $\varepsilon_t$  as

$$\varepsilon_t = \sqrt{h_t} \cdot v_t, \quad (7.27)$$

with  $v_t$  a white-noise process with unit variance,  $\varepsilon_t$  is an ARCH( $r$ ) process given that  $h_t$  obeys

$$h_t = \tau + \rho_{\varepsilon 1} \varepsilon_{t-1}^2 + \rho_{\varepsilon 2} \varepsilon_{t-2}^2 + \dots + \rho_{\varepsilon r} \varepsilon_{t-r}^2. \quad (7.28)$$

Extending (7.28) to infinite order yields the generalized autoregressive conditionally heteroskedastic (GARCH) representation, following Bollerslev (1986). Denote this extension as

$$h_t = \tau + \rho_\varepsilon(L) \varepsilon_t^2, \quad (7.29)$$

<sup>1</sup> GAUSS's Time Series Module is also available for estimating ARMA models. Alternatively, estimation code is available at <http://www.american.edu/academic.depts/cas/econ/gaussres/timeseri/timeseri.htm>.

with

$$\rho_\varepsilon(L) = \rho_{\varepsilon 1}L + \rho_{\varepsilon 2}L^2 + \dots,$$

and note its resemblance to a Wold Representation. Following the process for mapping an ARMA representation into a Wold Representation described above, we obtain a convenient means of parameterizing (7.29). Specifically, define  $\rho_\varepsilon(L)$  as the ratio of finite-ordered polynomials

$$\rho_\varepsilon(L) = \frac{\alpha(L)}{1 - \delta(L)},$$

$$\alpha(L) = \alpha_1 L + \alpha_2 L^2 + \dots + \alpha_a L^a$$

$$\delta(L) = \delta_1 L + \delta_2 L^2 + \dots + \delta_d L^d.$$

Then multiplying both sides of (7.29) by  $1 - \delta(L)$  and rearranging, we obtain

$$\begin{aligned} b_t &= \bar{\tau} + \delta_1 b_{t-1} + \delta_2 b_{t-2} + \dots + \delta_d b_{t-d} \\ &\quad + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_a \varepsilon_{t-a}^2, \\ \bar{\tau} &= \tau(1 - \delta_1 - \delta_2 - \dots - \delta_d). \end{aligned} \quad (7.30)$$

Note the resemblance of (7.30) to an ARMA process; in fact, this expression maps directly into an ARMA specification for  $\varepsilon_t^2$ .

To obtain this mapping, we first subtract  $\varepsilon_t^2$  from both sides of (7.30), and then add and subtract  $\delta_i \varepsilon_{t-i}^2$  on the right-hand side for  $i = 1, \dots, d$ , this yields

$$\begin{aligned} b_t - \varepsilon_t^2 &= \bar{\tau} - \delta_1(\varepsilon_{t-1}^2 - b_{t-1}) + \delta_2(\varepsilon_{t-2}^2 - b_{t-2}) + \dots + \delta_d(\varepsilon_{t-d}^2 - b_{t-d}) \\ &\quad + \delta_1 \varepsilon_{t-1}^2 + \delta_2 \varepsilon_{t-2}^2 + \dots + \delta_d \varepsilon_{t-d}^2 + \alpha_1 \varepsilon_{t-1}^2 \\ &\quad + \alpha_2 \varepsilon_{t-2}^2 + \dots + \alpha_a \varepsilon_{t-1}^2 - \varepsilon_t^2. \end{aligned} \quad (7.31)$$

Then defining

$$w_t = \varepsilon_t^2 - b_t,$$

(7.31) can be rewritten as

$$\begin{aligned} \varepsilon_t^2 &= \bar{\tau} + (\alpha_1 + \delta_1)\varepsilon_{t-1}^2 + (\alpha_2 + \delta_2)\varepsilon_{t-2}^2 + \dots + (\alpha_p + \delta_p)\varepsilon_{t-p}^2 \\ &\quad + w_t - \delta_1 w_{t-1} - \delta_2 w_{t-2} - \dots - \delta_d w_{t-d}, \end{aligned} \quad (7.32)$$

$$\rho_\varepsilon = \max(a, d), \quad (7.33)$$

which is the ARMA representation for  $\varepsilon_t^2$  we seek. In sum, if  $\varepsilon_t^2$  follows the ARMA representation (7.32),  $\varepsilon_t$  is defined as a GARCH( $\rho_\varepsilon, d$ ) process.

There exist many variations on ARCH and GARCH specifications designed to capture alternative manifestations of conditional heteroskedasticity; for a survey, see Bollerslev, Chou, and Kroner (1992). One such generalization is known as the stochastic volatility model, which is introduced in section 5.2 of chapter 8. In addition, there are a wide range of options for estimating regression models featuring ARCH/GARCH-type disturbances, including those featured in GAUSS's time series library. For a detailed discussion of options, see Brooks, Burke, and Persand (2003).

### 7.1.3 The VAR Model

A vector autoregressive (VAR) model specified for the  $m \times 1$  vector  $X_t$  is the multivariate analogue of an AR model specified for the single variable  $y_t$ . The multivariate analogue to the AR( $p$ ) specification (7.18) for  $X_t$  is given by

$$\begin{bmatrix} \rho_{11}(L) & \rho_{12}(L) & \cdots & \rho_{1m}(L) \\ \rho_{21}(L) & \rho_{22}(L) & \cdots & \rho_{2m}(L) \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{m1}(L) & \rho_{m2}(L) & \cdots & \rho_{mm}(L) \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ \vdots \\ X_{mt} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{mt} \end{bmatrix}, \quad E(\varepsilon_t \varepsilon_t') = \Sigma, \quad (7.34)$$

where  $\rho_{ij}(L) = (1 + \rho_{ij1}L + \dots + \rho_{ijp}L^p)$  is a  $p^{\text{th}}$ -order polynomial in the lag operator, expressing the influence of  $X_j$  on  $X_i$ . Defining the lead matrix in (7.34) as  $\Phi(L)$ , the roots of the VAR representation correspond to the  $mp$  factors of the determinant of  $\Phi(z)$ ,  $z$  complex. These roots will lie outside the unit circle given stationarity of the individual elements of  $X_t$ ; thus  $\Phi(L)$  may be inverted to obtain a multivariate analogue of the Wold Representation (7.8):

$$X_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \quad (7.35)$$

where  $\Psi_j$  is  $m \times m$ .

Covariation patterns between the elements of  $X_t$  may be conveniently characterized by writing the VAR in companion form. The companion form of an AR( $p$ ) model for a single element  $X_{it}$  of  $X_t$  is given by

$$\begin{bmatrix} X_{it} \\ X_{it-1} \\ X_{it-2} \\ \vdots \\ X_{it-(p+1)} \end{bmatrix} = \begin{bmatrix} \rho_{ii1} & \rho_{ii2} & \rho_{ii3} & \cdots & \rho_{iip} \\ 1 & 0 & 0 & \cdots & 0 \\ 0 & 1 & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \cdots & 1 \end{bmatrix} \begin{bmatrix} X_{it-1} \\ X_{it-2} \\ X_{it-3} \\ \vdots \\ X_{it-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{it} \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}. \quad (7.36)$$

Defining the  $(mp) \times 1$  vector

$$\begin{aligned} z_t = & [X_{1t} \ X_{1t-1} \ \dots \ X_{1t-(p+1)} \ X_{2t} \ X_{2t-1} \ \dots \ X_{2t-(p+1)} \ \dots \\ & X_{mt} \ X_{mt-1} \ \dots \ X_{mt-(p+1)}]', \end{aligned}$$

and the  $(mp) \times 1$  vector  $e_t = [\varepsilon_{1t} \ 0 \ \dots \ 0 \ \varepsilon_{2t} \ 0 \ \dots \ 0 \ \varepsilon_{mt} \ 0 \ \dots \ 0]',$  the companion form for the VAR is given by

$$z_t = Az_{t-1} + e_t, \quad (7.37)$$

where the  $(mp) \times (mp)$  companion matrix  $A$  contains VAR equations in the 1<sup>st</sup>,  $(p+1)^{\text{th}}, \dots,$  and  $[(m-1)p+1]^{\text{th}}$  rows, and maintains identities between elements of  $z_t$  and  $z_{t-1}$  in the remaining rows. Note the correspondence between (7.37) and the AR(1) expression for  $y_t$  in (7.13).

Exploiting this correspondence, let  $\Gamma(0) = E(z_t z_t')$  denote the contemporaneous variance-covariance matrix of  $z_t$ , and  $\Gamma(s) = E(z_t z_{t-s}')$  the  $s^{\text{th}}$ -order covariance matrix. From (7.37),

$$\begin{aligned} \Gamma(0) &= E(Az_{t-1} + e_t)(Az_{t-1} + e_t)' \\ &= A\Gamma(0)A' + \Sigma, \end{aligned} \quad (7.38)$$

the solution to which is given by

$$\text{vec}[\Gamma(0)] = [I - A \otimes A]^{-1} \text{vec}[\Sigma], \quad (7.39)$$

where  $\otimes$  denotes the Kronecker product. Further,  $\Gamma(1) = E(z_t z_{t-1}') = E((Az_{t-1} + e_t)z_{t-1}') = A\Gamma(0),$  and in general,

$$\begin{aligned} \Gamma(s) &= A\Gamma(s-1) \\ &= A^s\Gamma(0). \end{aligned} \quad (7.40)$$

Note the symmetry between (7.39) and (7.15), and between (7.40) and (7.16).

As with the AR( $p$ ) specification, the parameters of the VAR model may be estimated using OLS by rewriting (7.34) in standard regression notation:

$$\Upsilon = \Xi B + u, \quad (7.41)$$

where the  $(T-p) \times m$  matrix  $\Upsilon = [X_1 X_2 \dots X_m],$  with  $X_i = [X_{ip+1} \ X_{ip+2} \ \dots \ X_{iT}]',$  the  $(T-p) \times (mp)$  matrix  $\Xi$  contains in its  $t^{\text{th}}$  row

$$[X_{1t-1} X_{1t-2} \ \dots \ X_{1t-p} X_{2t-1} X_{2t-2} \ \dots \ X_{2t-p} \ \dots \ X_{mt-1} X_{mt-2} \ \dots \ X_{mt-p}],$$

the  $(mp) \times m$  matrix  $B$  contains VAR parameters for  $X_i$  in its  $i^{\text{th}}$  column, and the  $(T-p) \times m$  matrix  $u$  contains the  $(T-p) \times 1$  vector of innovations  $[\varepsilon_{ip+1} \ \varepsilon_{ip+2} \ \dots \ \varepsilon_{iT}]'$  corresponding with the VAR equation for  $X_i.$  And as with the AR( $p$ ) model, OLS estimates coincide with ML estimates given the assumption of Normality for  $\varepsilon_t.$ <sup>2</sup> Finally, the ARCH/GARCH specifications described for the univariate disturbance terms associated with ARMA models extend to the disturbance terms associated with VARs; for details, for example, see, Hamilton (1994).

## 7.2 Summary Statistics

We begin by discussing summary statistics for a single variable  $y_t.$  A good initial characterization of  $y_t$  is provided by its autocorrelation function, which plots  $\varphi(s)$  as a function of  $s$  (since  $\varphi(s) = \varphi(-s)$ , negative values of  $s$  are ignored). This plot provides an indication of the persistence of innovations to  $y_t,$  since as indicated in (7.12), the greater the persistence (i.e., the greater the horizon over which  $\psi_j$  differs from zero), the greater will be the horizon over which autocovariance terms (and thus autocorrelation terms) will differ from zero. The plot also illustrates cyclical patterns followed by  $y_t.$

Estimates of the elements of  $\varphi(s)$  may be obtained using the following collection of sample averages:<sup>3</sup>

$$\begin{aligned} \bar{y} &= \left(\frac{1}{T}\right) \sum_{t=1}^T y_t \\ \hat{\gamma}(0) &= \left(\frac{1}{T}\right) \sum_{t=1}^T (y_t - \bar{y})^2 \\ \hat{\gamma}(s) &= \left(\frac{1}{T}\right) \sum_{t=s+1}^T (y_t - \bar{y})(y_{t-s} - \bar{y}) \\ \hat{\varphi}(s) &= \hat{\gamma}(s)/\hat{\gamma}(0). \end{aligned} \quad (7.42)$$

Alternatively, given estimates of an ARMA specification for  $y_t,$  the corresponding Wold Representation of  $y_t$  may be constructed using (7.25), which can then be mapped into  $\hat{\varphi}(s)$  using (7.12).

Plots of  $\hat{\varphi}(s)$  (estimated using sample averages) for the four filtered versions of output described in chapter 6 are illustrated in figure 7.1. The four

<sup>2</sup>The collection of procedures contained in var.src is available for estimating VAR specifications, constructing the companion matrix  $A,$  and calculating  $\Gamma(s), s = 0, 1, \dots$

<sup>3</sup>The GAUSS command autocor performs these calculations.

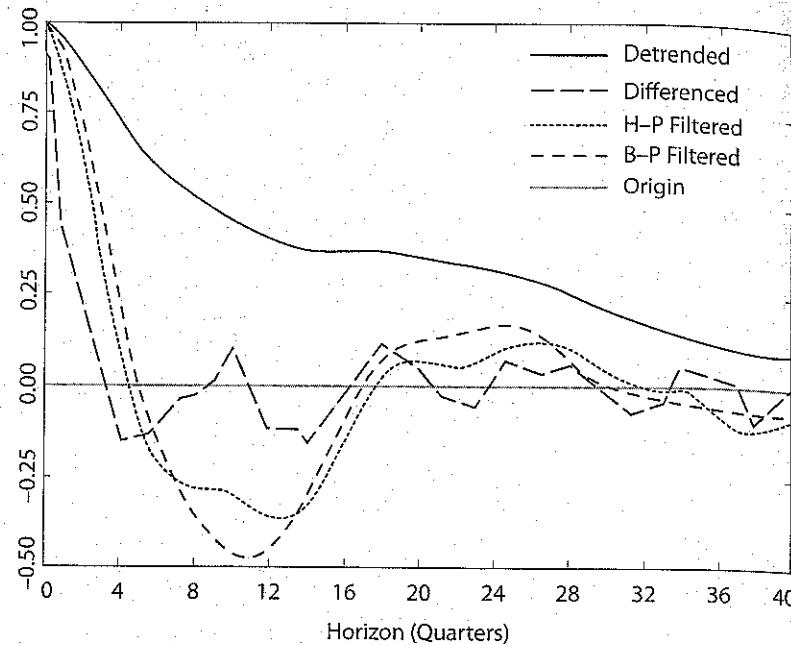


Figure 7.1 Sample Autocorrelations of Output

versions were obtained by detrending, differencing, Hodrick-Prescott (H-P) filtering, and band pass (B-P) filtering; each series represents logged deviations from trend.

Immediately apparent from figure 7.1 is the high degree of persistence exhibited by the detrended series. Recall from figure 6.2 that the level of the series follows a broken trend line, indicating a substantial reduction in growth approximately midway through the sample period. Given the removal of an unbroken trend line, the resulting detrended series persistently lies above zero during the first half of the sample period, and below zero during the second half. This persistence translates into the behavior of  $\hat{\phi}(s)$  depicted in figure 7.1, which decays very slowly, and remains at approximately 0.10 even at the 40-quarter horizon. Plots of  $\hat{\phi}(s)$  estimated for the alternative versions of  $y_t$  reveal a pattern of cyclical behavior: positive autocorrelation over the first four to six quarters gives way to negative autocorrelation until roughly the 16-quarter horizon, indicating predictable movements of  $y_t$  above and below trend.

Another useful means of summarizing the persistence and cyclicity of  $y_t$  is through the construction of its spectrum. The spectrum was described in detail in chapter 6; here we provide a brief summary. The spectrum

$s_y(\omega)$  is a decomposition of the variance of  $y_t$  by frequency  $\omega$ . Frequency, measured in radians, is usefully interpreted through its relationship with period  $p$ , which measures the number of time periods needed to complete a cycle:  $\omega = 2\pi/p$ . Thus business cycle frequencies, associated with periods between 6 and 40 quarters, fall within the range [0.157, 1.047] in working with quarterly data. The spectrum is closely related to the autocovariance function:

$$s_y(\omega) = \left( \frac{1}{2\pi} \right) \left[ \gamma(0) + 2 \sum_{\tau=1}^{\infty} \gamma(\tau) \cos(\omega\tau) \right]. \quad (7.43)$$

The integral of  $s_y(\omega)$  over the range  $[-\pi, \pi]$  yields  $\gamma(0)$ , and comparisons of the height of  $s_y(\omega)$  for alternative values of  $\omega$  indicate the relative importance of fluctuations at the chosen frequencies in influencing variations in  $y_t$ . Recall that since  $\cos(\omega\tau)$  is symmetric over  $[-\pi, 0]$  and  $[0, \pi]$ , so too is  $s(\omega)$ ; it is customary to represent  $s_y(\omega)$  over  $[0, \pi]$ .

As is clear from (7.43), the construction of an estimate of  $s_y(\omega)$  for  $y_t$  is straightforward given estimates  $\hat{\gamma}(s)$ . Alternatively, an estimate of  $s_y(\omega)$  may be obtained through its relationship with the parameters of an ARMA specification estimated for  $y_t$ ; the relationship is given by

$$s_y(\omega) = \left( \frac{\sigma^2}{2\pi} \right) \frac{|\theta(e^{-i\omega})|^2}{|\rho(e^{-i\omega})|^2}, \quad (7.44)$$

where  $e^{-i\omega} = \cos(\omega) - i \sin(\omega)$ ,  $\sqrt{i} = -1$ , and  $|\cdot|$  denotes the modulus operator; for example,  $|\theta(e^{-i\omega})| = \sqrt{\theta(e^{-i\omega})\theta(e^{i\omega})}$ .<sup>4</sup>

Spectra estimated for the four versions of output described above are illustrated in figure 7.2. Also, spectra estimated for the H-P filtered versions of output, consumption, investment, and hours are illustrated in figure 7.3. The estimates were obtained by estimating ARMA models for each series, and constructing  $s_y(\omega)$  as in (7.44). In the figures, the horizontal axis is in terms of  $\omega/2\pi$ : cycles per quarter.

Note how the behavior of the autocorrelation functions depicted in figure 7.1 translate into the behavior of the spectra depicted in figure 7.2. The persistence evident in the detrended series translates as a spike in its corresponding spectrum at zero frequency. (The height of this spectrum actually grows unboundedly as  $\omega$  approaches zero; the spectrum was truncated at 0.1 in the figure to better depict the additional spectra.) The frequency zero corresponds to a period  $p$  of infinity, or to a cycle that never repeats. As indicated by its spectrum, innovations with this characteristic dominate the behavior of the detrended series.

<sup>4</sup> See Harvey (1993) for a derivation of (7.44). The procedure `spec_arma.prc` may be used to construct  $s_y(\omega)$ , taking  $\rho(L)$  and  $\theta(L)$  as inputs.

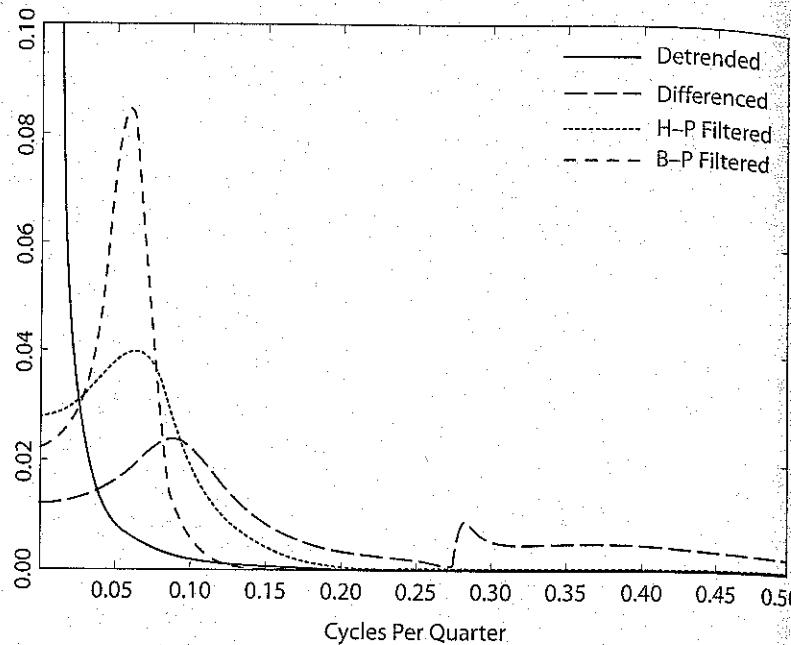


Figure 7.2 Spectra of Output

For the remaining series, spectral peaks lie within business cycle frequencies (between  $1/40$  and  $1/6$  cycles per quarter). In the case of the H-P and B-P filtered series, this is by design, as the characterization of the squared gains corresponding with the H-P and B-P filters provided in chapter 6 illustrates (see in particular figures 6.7 and 6.11). The spectral peak associated with the differenced series is much less pronounced in comparison, and the nontrivial component of the spectrum of this series over the range  $[0.15, 0.5]$  reflects the influence of relatively high-frequency fluctuations on the overall behavior of this series.

We conclude the characterization of  $y_t$  by recalling the discussion of impulse response functions in section 1. Recall that these trace the response of  $y_t$  to the realization of a single shock at time 0:  $\varepsilon_0 = 1$ ,  $\varepsilon_t = 0$ ,  $t > 0$ . Equivalently, these functions plot the coefficients  $\psi_j$ ,  $j = 1, 2, \dots$  of the Wold Representation of  $y_t$ , as expressed in (7.8). Impulse response functions are illustrated for the four H-P filtered series in figure 7.4. Each function was constructed from estimates of an ARMA specification, which were used to construct  $\psi_j$ ,  $j = 1, 2, \dots$  as in (7.25).<sup>5</sup>

<sup>5</sup> The procedure `armaimp.prc` may be used to calculate impulse response functions in this fashion.

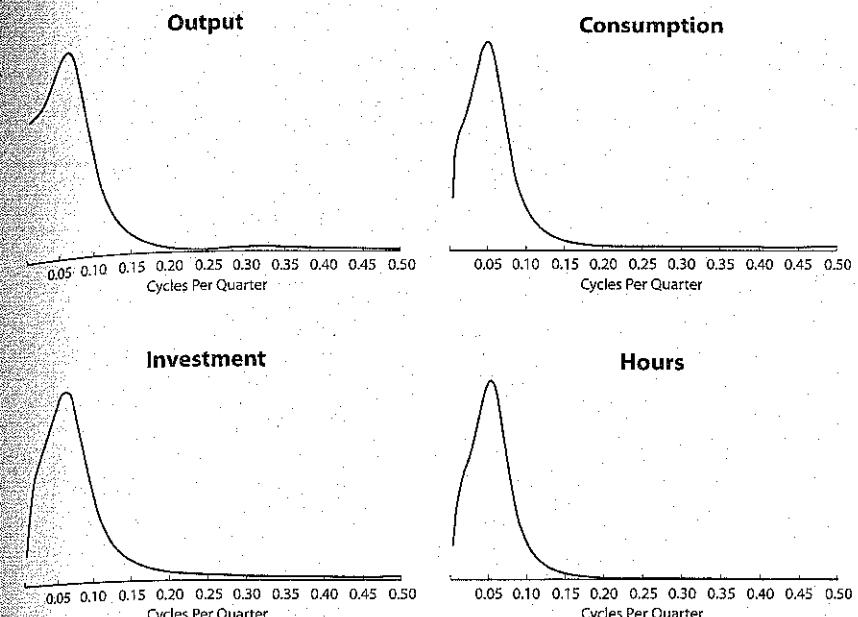


Figure 7.3 Spectra of H-P Filtered Data

Note how the cyclical nature of the responses mimics the autocorrelation pattern of H-P filtered output in figure 4.1. A positive shock drives each series above trend for approximately four to six quarters; the series then overshoot their trend lines before ultimately recovering at approximately the 20-quarter horizon. Note also the close relationship between the dynamic responses of the series.

We now discuss summary statistics for the collection of variables contained in the  $m \times 1$  vector  $X_t$ , or when convenient, for the expanded collection of variables contained in the  $(mp) \times 1$  vector  $z_t$ , constructed from  $X_t$  as indicated by the companion form of the VAR specified in (7.37). Patterns of auto- and cross-covariation are usefully characterized using  $\Gamma(s) = E(z_t z_t')$ ,  $s = 0, 1, \dots$ . These may be obtained from estimated VAR parameters following (7.38)–(7.40). A collection of statistics frequently of interest in empirical business cycle applications is reported for the differenced, H-P filtered, and B-P filtered data in table 7.1. The detrended data were excluded from the table because the VAR specification estimated for these data contains a unit root; thus the construction of  $\Gamma(0)$  via (7.39) is not possible in this case.

The first column of the table reports standard deviations of the individual series, and the second reports standard deviations relative to the standard

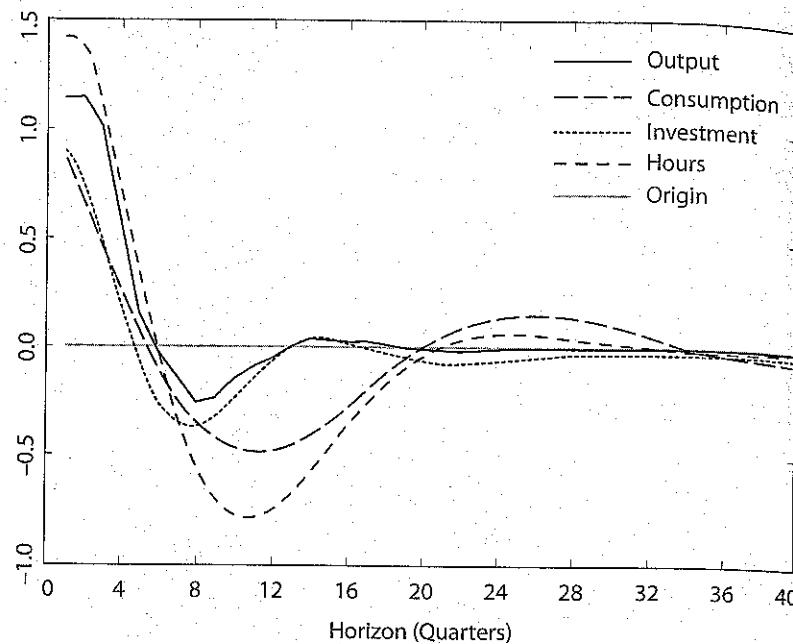


Figure 7.4 Univariate Impulse Response Functions of H-P Filtered Data

deviation of output. Note that under all three versions of the data, investment is far more volatile than output ( $\sigma_i/\sigma_y$  ranges from approximately 4 to 5), while consumption is far smoother ( $\sigma_c/\sigma_y$  is roughly 0.5); the volatilities of hours and output are roughly equal. Measures of first-order serial correlation are quite low among the differenced series (particularly for investment: 0.18), while ranging from 0.79 to 0.95 among the H-P and B-P filtered data. Finally, with  $\varphi_{j,y}(s)$  denoting the  $s^{\text{th}}$ -order correlation between variable  $j$  and output, note that output is most closely correlated with investment, and that correlation patterns in general are relatively weak among the differenced series in comparison with the H-P and B-P filtered series.

### Exercise 7.3

Construct the collection of summary statistics presented in table 7.1 using a version of detrended output obtained by fitting a broken trend to the data. Do this using the following steps.

1. Split the sample into two periods: observations through 1993:IV; and observations from 1994:I to 2010:I. For each, fit a common linear trend to logged output, consumption and investment; subtract the estimated trend;

TABLE 7.1  
Summary Statistics Estimated from VAR

Differenced data

$j$	$\sigma_j$	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\varphi_{j,y}(1)$
$y$	0.0101	1.00	0.41	1.00	0.41
$c$	0.0049	0.49	0.30	0.56	0.46
$i$	0.0511	5.06	0.18	0.92	0.26
$h$	0.0093	0.92	0.63	0.71	0.37

H-P filtered data

$j$	$\sigma_j$	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\varphi_{j,y}(1)$
$y$	0.0177	1.00	0.87	1.00	0.87
$c$	0.0078	0.44	0.84	0.82	0.75
$i$	0.0758	4.28	0.79	0.96	0.80
$h$	0.0190	1.07	0.90	0.83	0.61

B-P filtered data

$j$	$\sigma_j$	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\varphi_{j,y}(1)$
$y$	0.0200	1.00	0.95	1.00	0.95
$c$	0.0088	0.42	0.95	0.89	0.83
$i$	0.0834	4.01	0.93	0.97	0.93
$h$	0.0216	1.03	0.95	0.87	0.71

2. Estimate a VAR representation for the data using  $p = 8$ ; construct the companion matrix  $A$  as in (7.37); and then construct  $\Gamma(0)$  and  $\Gamma(1)$  using (7.38)–(7.40) (use the procedures contained in `var.spc` for this step).

3. Obtain  $[\sigma_j, \frac{\sigma_j}{\sigma_y}, \varphi(1), \varphi_{j,y}(0), \varphi_{j,y}(1)]$  from the relevant entries of  $\Gamma(0)$  and  $\Gamma(1)$ .

Spectral representations of the data may also be constructed using VARs. This is accomplished via a multivariate analogue of (7.44). Setting  $\theta(L) = 1$ , (7.44) characterizes how the spectrum of a single variable is obtained using estimates of an AR( $p$ ) model. To obtain the multivariate analogue of (7.44), write the VAR of  $X_t$  as

$$X_t = \Upsilon_1 X_{t-1} + \Upsilon_2 X_{t-2} + \cdots + \Upsilon_p X_{t-p} + \varepsilon_t, \quad (7.45)$$

where  $\Upsilon_j$  is  $k \times k$ . Letting  $\Upsilon(L) = I - \Upsilon_1 L - \cdots - \Upsilon_p L^p$ , the spectrum of  $X_t$ , denoted as  $S(\omega)$ , is given by<sup>6</sup>

$$S(\omega) = \frac{1}{2\pi} [\Upsilon(e^{-i\omega})' \Sigma^{-1} \Upsilon(e^{-i\omega})]^{-1}. \quad (7.46)$$

<sup>6</sup>For a derivation, see Hannan (1970). The procedure `spec_var.prc` is available for use in `vars` (see Section 7.1).

The  $j^{\text{th}}$  diagonal element of  $S(\omega)$  contains the spectrum of the  $j^{\text{th}}$  variable of  $X_t$  at frequency  $\omega$ ,  $s_j(\omega)$ , and the  $(i,j)^{\text{th}}$  element of  $S(\omega)$  contains the cross spectrum between the  $i^{\text{th}}$  and  $j^{\text{th}}$  variables of  $X_t$ ,  $s_{i,j}(\omega)$ . Estimates of the spectra for the H-P filtered data produced using (7.46) closely mimic those depicted in figure 7.3, and thus are not reported here.

Finally, impulse response functions may also be constructed using VARs. This is most conveniently accomplished using the companion form of the VAR given by (7.37). Initializing  $z_{-1} = 0$ , this proceeds by constructing a nonzero initial value  $e_0$ , and then constructing  $\{z_t\}_{t=0}^{\infty}$  as  $z_0 = e_0$ ,  $z_1 = Ae_0$ ,  $z_2 = A^2e_0, \dots$ . The specification for  $e_0$  is typically chosen to simulate the impact on the system of a one-standard-deviation innovation to the  $j^{\text{th}}$  variable of the system. If the innovations were all uncorrelated, so that the covariance matrix  $\Sigma$  were diagonal, the specification for  $e_0$  would contain a zero in all but its  $[(j-1)p+1]^{\text{th}}$  row, which corresponds to the  $j^{\text{th}}$  equation of the VAR. This entry would be set to the square root of the  $j^{\text{th}}$  diagonal element of  $\Sigma$ . However, correlation among the innovations implies that an innovation to the  $j^{\text{th}}$  variable of the system coincides with innovations to additional variables in the system. To capture this,  $e_0$  must be constructed accordingly.

Recall that  $e_t$  relates to the system innovations as  $e_t = [e_{1,t} \ 0 \dots 0 \ e_{2,t} \ 0 \dots 0 \ e_{mt} \ 0 \dots 0]'$ ; for the moment, it will be convenient to work directly with the vector  $\varepsilon_t$ . To recap, the problem faced in constructing  $e_0$  is that the elements of  $\varepsilon_t$  are typically correlated, so that a movement in one component will coincide with movements in additional components. The leading approach to dealing with this involves working with orthogonalized innovations in place of  $\varepsilon_t$ . Orthogonalized innovations are innovations that are uncorrelated across VAR equations. Let  $v_t$  represent such innovations, so that  $E v_t v_t' = I$ . Defining a matrix  $P$  such that

$$P^{-1} \Sigma P'^{-1} = I, \quad (7.47)$$

which implies

$$\Sigma = PP', \quad (7.48)$$

$v_t$  may be constructed using

$$v_t = P^{-1} \varepsilon_t. \quad (7.49)$$

The only problem at this point is that there are many possible specifications of  $P^{-1}$  that can be constructed to satisfy (7.48). Here we discuss a leading specification: the Cholesky decomposition of  $\Sigma$ ; see Hamilton (1994) for a discussion of alternative decompositions.

## 7.2 Summary Statistics

The Cholesky decomposition of  $\Sigma$  is a lower-triangular matrix that satisfies (7.48), with diagonal elements containing the square root of the diagonal elements of  $\Sigma$  (i.e., the standard deviations of the elements of  $\varepsilon_t$ ). Consider the construction of  $v_0$  obtained by introducing into (7.49) a specification of  $\varepsilon_0$  containing 1 in its  $j^{\text{th}}$  row and zeros elsewhere. The resulting specification of  $v_0$  will contain the  $j^{\text{th}}$  column of  $P^{-1}$ . Using the Cholesky decomposition of  $\Sigma$  for  $P^{-1}$ ,  $v_0$  will contain zeros in the first  $j-1$  entries, the standard deviation of the  $j^{\text{th}}$  element of  $\varepsilon_t$  in the  $j^{\text{th}}$  row, and nonzero entries in the remaining rows. These latter entries represent the influence of an innovation to the  $j^{\text{th}}$  equation of the system on innovations to the  $(j+1)^{\text{st}}$  through  $m$  equations. This reflects the correlation among the innovations. Note however that under this construct the innovation to the  $j^{\text{th}}$  equation is prevented from influencing the 1<sup>st</sup> through  $(j-1)^{\text{st}}$  equations; thus the ordering of the variables of the system will influence impulse response functions constructed in this manner.

Impulse response functions constructed using the H-P filtered data are illustrated in figure 7.5.<sup>7</sup> The ordering of the variables used to construct the responses is  $(y, c, i, h)$ . Most striking is the magnitude of the responses of investment to each shock, clearly indicating the high volatility of this series in comparison with the others. Investment initially responds negatively to a positive consumption shock; otherwise, all variables exhibit positive covariation in response to the shocks. Finally, the cyclical patterns described above are once again in evidence here, as the variables tend to overshoot their targets in returning to preshock levels.

### 7.2.1 Determining Lag Lengths

As noted, each of the summary statistics we have discussed may be constructed from estimates of ARMA or VAR models, which can themselves be estimated using ML or OLS techniques. An issue involving these estimates regards the choice of lag lengths  $p$  and  $q$  for these specifications. Often this choice will depend on factors particular to the problem at hand. For example, if the number of available data points is relatively low, parsimony may be of primary concern. Alternatively, if the objective is to obtain a general feel for the data, as with the examples provided above, it may be preferable to work with relatively liberal specifications. Here, we briefly mention three leading approaches to lag-length specification; for an extended discussion, see Judge et al. (1985).

First is a general-to-specific approach. This involves the sequential testing of exclusion restrictions given an initial specification of a liberally

<sup>7</sup>A procedure for calculating impulse response functions is included in var.spc.

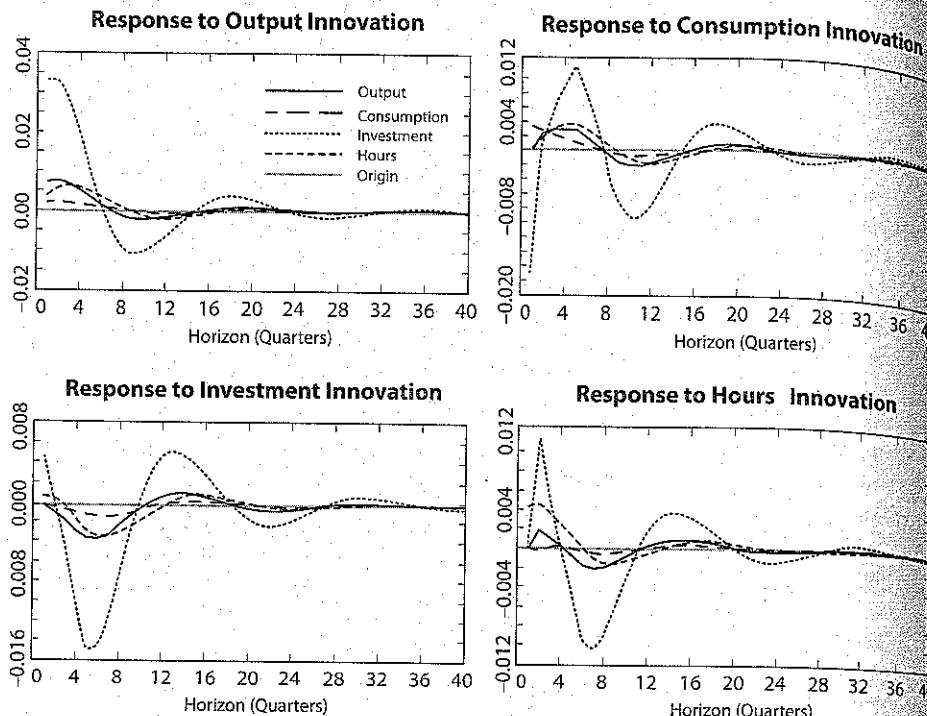


Figure 7.5 System Impulse Response Functions of H-P Filtered Data

parameterized model. For example, this could involve individual  $t$  tests of the null hypothesis that the  $q^{\text{th}}$  lag chosen for the MA component of an ARMA model, or the  $p^{\text{th}}$  lag chosen for the AR component, is zero. Alternatively, sets of exclusion restrictions could be evaluated using a likelihood ratio (LR) test. Letting  $L_u$  and  $L_r$  denote values of the unrestricted and restricted likelihood functions associated with the model being estimated, the LR test statistic is given by  $2[L_u - L_r]$ , which is asymptotically distributed as  $\chi^2(k)$ , where  $k$  denotes the number of restrictions imposed in calculating  $L_r$ .

Alternative approaches involve the use of selection criteria that explicitly incorporate penalties for selecting liberal parameterizations. Letting  $K$  denote the total number of parameters associated with a given model specification, two leading criteria are the Akaike Information Criterion (AIC) (Akaike, 1974), which delivers the  $K$  that minimizes

$$AIC = |\Sigma(K)| + \frac{2K}{T}; \quad (7.50)$$

and the Bayesian Information Criterion (BIC) (Schwarz, 1978), which delivers the  $K$  that minimizes

$$\text{BIC} = |\Sigma(K)| + \frac{K \log(T)}{T}. \quad (7.51)$$

The notation  $\Sigma(K)$  is used to indicate explicitly that the fit of the model, represented by  $|\Sigma(K)|$ , will improve as  $K$  increases:  $|\Sigma(K)|$  is decreasing in  $K$ . Of course, gains realized by increasing  $K$  are countered by increases in the penalty terms.

### 7.2.2 Characterizing the Precision of Measurements

We conclude this section with a note regarding measures of the precision with which summary statistics have been estimated. This provides a brief first glimpse of Monte Carlo methods, which are characterized in detail in Part IV of the text.

For concreteness, let  $\hat{g}(\vartheta)$  denote the estimate of a given function of the  $k \times 1$  vector of parameters  $\vartheta$ , which summarize a reduced-form model specified for the data. (The replacement of  $\hat{g}(\varphi)$  with  $\hat{g}(\mu)$ , so that the discussion is centered on a structural rather than a reduced-form model, yields an analogous discussion.) From a classical statistical perspective, under which parameters are interpreted as fixed and data as random, precision is conveyed by reporting the standard error associated with  $\hat{g}(\vartheta)$ . From a Bayesian perspective, under which the data are interpreted as fixed and parameters as random, precision is conveyed by reporting the posterior standard deviation associated with  $\hat{g}(\vartheta)$ .

Beginning with the former, since the data used to construct  $\hat{g}(\vartheta)$  are random, and represent one of many possible realizations that could have been obtained,  $\hat{g}(\vartheta)$  is also random. Its variance is referred to as the sampling variance of  $\hat{g}(\vartheta)$ , which must typically be estimated. The square root of this estimate is the standard error of  $\hat{g}(\vartheta)$ . In some cases analytical expressions for standard errors are available; often they are not. For example, if the vector  $\vartheta$  is being estimated using an OLS regression of the form  $y = X\vartheta + e$ , and  $\hat{g}(\vartheta)$  is a vector representing the OLS estimate of  $\vartheta$ , then the associated standard errors of the individual elements of  $\hat{g}(\vartheta)$  are the square roots of the diagonal elements of

$$\text{Var}(\hat{\vartheta}) = \hat{\sigma}^2 (X'X)^{-1}, \quad \hat{\sigma}^2 = \left( \frac{1}{T} \right) \sum_{t=1}^T \hat{e}_t^2, \quad \hat{e}_t = y - X\hat{\vartheta}.$$

When  $\hat{g}(\vartheta)$  represents a nonlinear function of  $\vartheta$ , analytical expressions for the standard error of  $\hat{g}(\vartheta)$  are generally unavailable. One remedy

for this is use of the Delta method. Consider a first-order Taylor Series approximation of  $\widehat{g}(\vartheta)$  around the true value  $\vartheta$ :

$$\widehat{g}(\vartheta) \approx g(\vartheta) + \left( \frac{\partial g(\vartheta)}{\partial \vartheta} \right)' (\widehat{\vartheta} - \vartheta). \quad (7.52)$$

If  $g(\vartheta)$  is a scalar,  $\left( \frac{\partial g(\vartheta)}{\partial \vartheta} \right)'$  is  $1 \times k$  and  $(\widehat{\vartheta} - \vartheta)$  is  $k \times 1$ . If  $g(\vartheta)$  is an  $\ell \times 1$  vector,  $\left( \frac{\partial g(\vartheta)}{\partial \vartheta} \right)'$  is an  $\ell \times k$  Jacobian matrix, with  $(i, j)^{\text{th}}$  element given by the derivative of the  $i^{\text{th}}$  row of  $g(\vartheta)$  with respect to the  $j^{\text{th}}$  element of  $\vartheta$ . By Slutsky's theorem (e.g., as presented in Greene, 2003), if  $\widehat{\vartheta}$  is a consistent estimate of  $\vartheta$ , then  $\widehat{g}(\vartheta)$  will be a consistent estimator of  $g(\vartheta)$ , with variance given by

$$\text{Var}[\widehat{g}(\vartheta)] \approx \left( \frac{\partial g(\vartheta)}{\partial \vartheta} \right)' \text{Var}(\widehat{\vartheta}) \left( \frac{\partial g(\vartheta)}{\partial \vartheta} \right). \quad (7.53)$$

The standard errors associated with the individual elements of  $\widehat{g}(\vartheta)$  are thus the square roots of the diagonal elements of  $\text{Var}[\widehat{g}(\vartheta)]$ .

Alternatively, if expressions for  $\frac{\partial g(\vartheta)}{\partial \vartheta}$  are difficult to obtain, standard errors may be calculated using numerical approximation methods, such as the method of Monte Carlo. This involves generating artificial sample drawings of data from the distribution implied by a parameterized version of their reduced-form model, calculating  $\widehat{g}(\vartheta)$  for each artificial drawing, and then computing the variance of  $\widehat{g}(\vartheta)$  from the resulting collection of drawings. Once again, square roots of the diagonal elements of the estimated variance of  $\widehat{g}(\vartheta)$  serve as estimates of the standard errors of the individual elements of  $\widehat{g}(\vartheta)$ .

As an example, consider the construction of estimates of  $\sigma_j$ ,  $\varphi_j(1)$ , and  $\varphi_{i,j}(s)$  via the use of a VAR specified for  $X_t$ , as demonstrated in table 7.1. In this case, a parameterized version of (7.34), or equivalently (7.37), serves as the model used to generate artificial realizations of  $\{X_t\}_{t=1}^T$ ; this model is known as the data generation process (DGP) in the experiment. Consider the use of (7.37) for this purpose. Using  $p$  initial values of  $\{X_t\}$  as starting values, implying a specification  $z_0$ , a drawing of  $\{X_t\}_{t=1}^T$  is obtained using (7.37) first by obtaining a drawing of  $\{e_t\}_{t=1}^T$  from a specified distribution, inserting  $e_1$  into (7.37) to obtain  $z_1 = Az_0 + e_0$ , and selecting the  $1^{\text{st}}, (p+1)^{\text{th}}, \dots, [(m-1)p+1]^{\text{th}}$  elements from  $z_1$  to obtain  $X_1$ . Repeating this process  $T$  times yields  $\{X_t\}_{t=1}^T$ . This realization is then used to estimate  $A$  and  $\Sigma$ , from which  $\Gamma(0)$  and  $\Gamma(1)$  are constructed using (7.38)–(7.40). Finally, extraction of  $\sigma_j$ ,  $\varphi_j(1)$ , and  $\varphi_{i,j}(s)$  from  $\Gamma(0)$  and  $\Gamma(1)$  yields a Monte Carlo drawing of these statistics.

## 7.2 Summary Statistics

Denote the  $i^{\text{th}}$  drawing of these statistics as  $\widehat{g}(\vartheta)^i$ . The mean and variance of  $\widehat{g}(\vartheta)$  calculated over the realization of  $N$  Monte Carlo drawings are given by

$$\overline{g(\vartheta)} = \frac{1}{N} \sum_{i=1}^N \widehat{g}(\vartheta)^i; \quad (7.54)$$

$$\text{Var}(g(\vartheta)) = \frac{1}{N} \sum_{i=1}^N [\widehat{g}(\vartheta)^i - \overline{g(\vartheta)}]^2. \quad (7.55)$$

The square root of the diagonal elements of  $\text{Var}(g(\vartheta))$  provide a single estimate of the standard error of  $\widehat{g}(\vartheta)$ . Denote this estimate as  $s.e.[g(\vartheta)]$ . Replication of this process using a total of  $J$  Monte Carlo experiments yields  $J$  drawings of  $s.e.[g(\vartheta)]$ . Letting the  $j^{\text{th}}$  drawing be given by  $s.e.[g(\vartheta)]^j$ , a natural estimator of  $s.e.[g(\vartheta)]$  is the average computed over the  $J$  experiments:

$$\overline{s.e.[g(\vartheta)]} = \frac{1}{J} \sum_{j=1}^J s.e.[g(\vartheta)]^j. \quad (7.56)$$

The variance of this estimator is calculated as

$$\text{Var}[\overline{s.e.[g(\vartheta)]}] = \frac{1}{J} \sum_{j=1}^J \left\{ s.e.[g(\vartheta)]^j - \overline{s.e.[g(\vartheta)]} \right\}^2, \quad (7.57)$$

and the standard deviation  $s.e.\overline{s.e.[g(\vartheta)]}$  is once again the square root of the diagonal elements of the variance.

An assessment of the accuracy of  $s.e.[g(\vartheta)]$  is provided by its numerical standard error (NSE). The NSE associated with any Monte Carlo estimator is given by the ratio of its standard error to the number of Monte Carlo replications (e.g., see Rubinstein, 1981). The NSE of  $s.e.[g(\vartheta)]$  is therefore

$$\text{NSE}\left\{\overline{s.e.[g(\vartheta)]}\right\} = \frac{s.e.\overline{s.e.[g(\vartheta)]}}{\sqrt{J}}. \quad (7.58)$$

A critical step in obtaining Monte Carlo estimates involves the design of the experiment. In the present context there are three components to the design: parameterization of the DGP (including the distribution chosen for  $e_t$ ); the specification for  $z_0$ ; and the specification of the artificial sample size  $T$ . Choices for these components in general will depend upon the particular objective of the experiment. But given the goal of constructing standard errors for point estimates of statistics constructed from the

parameters of a reduced-form specification, a natural starting point for parameterizing the DGP is to employ point estimates obtained using the actual sample. Also, natural choices for  $z_0$  include the unconditional mean of  $z$  (zero in this case), or the initial value for  $z_0$  obtained for the actual sample. And a natural choice for  $T$  is the actual sample size. Variations along all three dimensions are useful for assessing the sensitivity of results.

#### Exercise 7.4

Construct a Monte Carlo estimate of standard errors associated with estimates of the spectrum of a single variable  $y_t$  obtained using (7.44). Do so using the following steps.

1. Construct an ARMA(1,1) DGP for  $y_t$ , using  $T = 100$ ,  $y_0 = 0$ ,  $\rho = 0.8$ ,  $\theta = -0.3$ , and  $\epsilon_t \sim N(0, 1)$ .
2. Let  $g(\vartheta)$  represent the  $(n+1) \times 1$  vector of values of the spectrum  $s_y(\omega)$ ,  $\omega = 0, \pi/n, 2\pi/n, \dots, \pi$ ,  $n = 40$ .
3. For each of 100 realizations of  $\{y_t\}$  obtained from the DGP you constructed, estimate an ARMA(2,2) specification, and construct  $g(\vartheta)$  by inserting the parameter estimates you obtain in (7.44).
4. Calculate the standard deviation of each element of  $g(\vartheta)$  over the 100 drawings you obtained: this yields a single estimate of  $s.e.[s_y(\omega)]$ ,  $\omega = 0, \pi/n, 2\pi/n, \dots, \pi$ .
5. Repeat steps 3 and 4 100 times, and use the resulting 100 drawings of  $s.e.[s_y(\omega)]$  to construct  $s.e.[g(\vartheta)]$ ,  $s.e.[s.e.[g(\vartheta)]]$ , and  $nse[s.e.[g(\vartheta)]]$ .

Under the Bayesian perspective,  $\vartheta$  (and thus  $(g(\vartheta))$ ) is interpreted as random; inferences regarding  $g(\vartheta)$  involve calculations of conditional probabilities associated with alternative values of  $g(\vartheta)$ . Conditional probabilities are assigned by the posterior distribution associated with  $g(\vartheta)$ . The posterior distribution reflects the combined influence of a prior distribution specified by the researcher over  $g(\vartheta)$ , and the conditional likelihood function associated with  $g(\vartheta)$  (where conditioning is with respect to the observed data). Point estimates  $g(\vartheta)$  are typically given by means or modes of the posterior distribution obtained for  $g(\vartheta)$ ; and the precision of these estimates is typically summarized using posterior standard deviations. Details regarding the Bayesian perspective are provided in chapter 14.

### 7.3 Obtaining Theoretical Predictions of Summary Statistics

To this point the discussion has centered on summary statistics designed to characterize the time-series behavior of the collection of observable

variables  $X_t$ . However, the same collection of statistics may be used to characterize the time-series behavior of model variables. Here we provide a brief introduction relevant for log-linear model representations, which serves as foreshadowing for the detailed presentations of chapters 11 and 12.

Recall that log-linear model representations are given by given by

$$x_t = Fx_{t-1} + Gv_t, \quad (7.59)$$

$$Gv_t = e_t, \quad (7.60)$$

$$E(e_t e_t') = Q, \quad (7.61)$$

where  $x_t$  is the  $n \times 1$  collection of model variables that map into the observables via

$$X_t = H'x_t + u_t. \quad (7.62)$$

Note the similarity of this specification relative to the companion form (7.37) of the VAR specified for the  $(mp) \times 1$  vector

$$\begin{aligned} z_t = & [X_{1t} \ X_{1t-1} \ \dots \ X_{1t-(p+1)} \ X_{2t} \ X_{2t-1} \ \dots \ X_{2t-(p+1)} \ \dots \\ & X_{mt} \ X_{mt-1} \ \dots \ X_{mt-(p+1)}]', \end{aligned}$$

repeated here for convenience:

$$z_t = Az_{t-1} + e_t, \quad E(e_t e_t') = \Sigma.$$

Aside from the slight abuse of notation involving  $e_t$ , the parallel structures of (7.59) and (7.37) render as trivial the calculation of summary statistics associated with the specification of the theoretical model. In particular, each of the summary statistics that can be constructed from VAR estimates can also be constructed for  $x_t$  by replacing  $A$  and  $\Sigma$  with  $F$  and  $Q$ . Moreover, since the relationship between model variables and observables is given by the simple linear mapping, it is straightforward to align summary statistics obtained from reduced-form specifications with those obtained from structural specifications.

Despite the straightforward nature of alignment, a subtlety regarding the interpretation of impulse response functions obtained from structural as opposed to reduced-form specifications bears emphasis. To begin, note that the mechanics of their construction are identical across specifications: setting either  $x_{-1} = 0$  in (7.59) or  $z_{-1} = 0$  in (7.37), impulse response functions are obtained by specifying a nonzero specification for  $e_0$ , restricting subsequent realizations  $e_j = 0$ ,  $j > 0$ , and constructing either  $\{x_t\}$  or  $\{z_t\}$  via recursive substitution.

The difference across specifications lies in the interpretation of  $e_0$ . Returning to figure 7.5, recall that the responses depicted in each panel

represent responses to an innovation in a single equation in isolation. What we lack in interpreting these responses is the underlying source of the innovation: Did it arise from a shock to TFP? A surprise change in interest rates? A combination thereof? In short, what we lack is a structural interpretation of the innovation.

In contrast, in working with the structural model (7.59),  $e_0$  is constructed as

$$e_0 = Gv_0.$$

Typically, all but one of the elements of  $v_0$  are set to zero in constructing  $e_0$ , so that the impact of the exceptional element can be assessed in isolation. Since each of the elements of  $v_t$  has a clear structural interpretation, so too does the interpretation of the resulting responses. In section 3 of chapter 12, we shall return to this issue in characterizing a method of reverse-engineering structural innovations from VAR innovations. And in the following chapter, we describe how the time-series behavior of  $\{v_t\}_{t=1}^T$  can be inferred given the observations  $\{X\}_{t=1}^T$ , via a process known as filtering.

We conclude this chapter by demonstrating the use of model impulse response functions as an important tool for helping convey intuition regarding the mechanics of the model under investigation. We do so for the RBC model characterized in section 1 of chapter 3. Figure 7.6 illustrates responses to a one-standard-deviation innovation to TFP obtained under two model parameterizations: one set generated using a relatively small specification for the CRRA parameter  $\phi$  (0.5), and a second generated using a relatively large value (2). The remaining parameter values are those specified under the calibration exercise presented in section 4 of chapter 11; see table 11.4 for details.

Responses are measured as percentage deviations from steady state. Note that with the exception of investment, the vertical axes in each panel are identical; the scale of investment is expanded to capture the full extent of its relatively volatile response.

Under both model parameterizations, the response of output to the TFP shock is positive. Beyond the obvious direct impact of the shock, this reflects the agent's decision to substitute time away from leisure and towards labor in order to exploit the shock's positive impact on the marginal product of labor. The increase in output enables increases in both consumption and investment; the latter contributes towards the propagation of the impact of shock over time by generating a prolonged increase in the stock of capital. Finally, comparing responses across parameterizations, note that the response profiles of all variables are relatively flat under the parameterization  $\phi = 2$ . This reflects the relatively strong incentive for

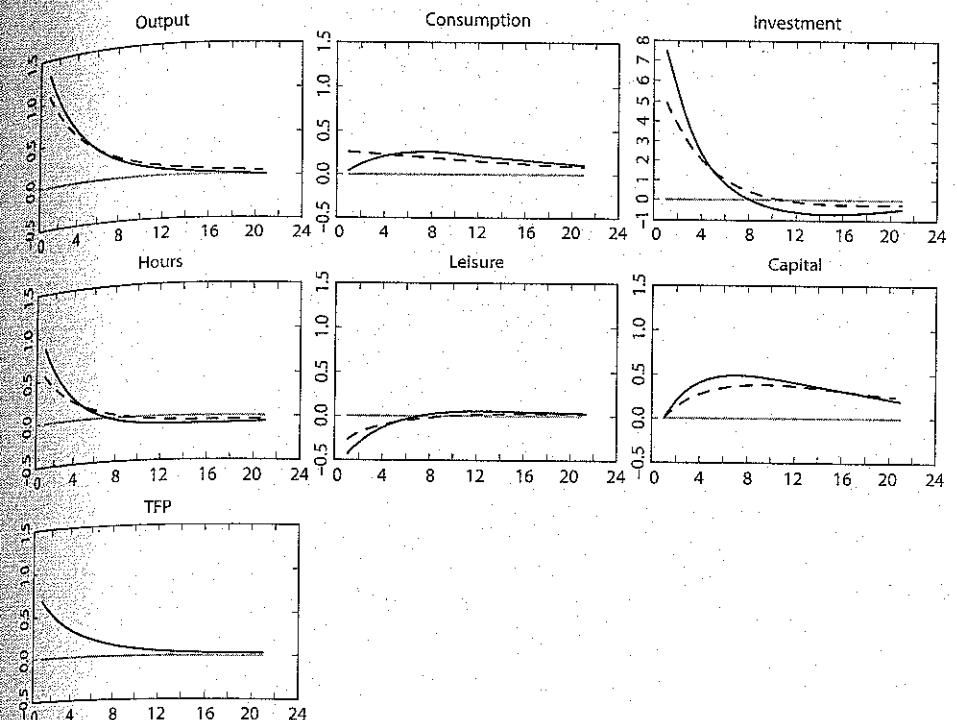


Figure 7.6 Model Impulse Response Functions. Solid:  $\phi = 0.5$ ; dashes:  $\phi = 2$ .

maintaining smooth intertemporal instantaneous utility profiles imparted by increases in  $\phi$ .

### Exercise 7.5

Generate additional sets of impulse responses by experimenting with alternative parameterizations of the RBC model. Provide intuition for the impact on responses resulting from changes in each parameter.

# Chapter 8 - Structural Macroeconomics

## State-Space Representations

State-space representations are highly flexible models that characterize the dynamic interaction of a collection of time-series variables, some of which is unobservable. The observable series may either be observed directly, or subject to measurement error. The flexibility of these models has led to their use in an extensive range of applications in multiple disciplines. For textbook characterizations of the application of state-space representations in disciplines including engineering, environmental biology, meteorology, and physics; and in problems including the detection of solar systems, the design of control systems, and the tracking of non-maneuvering, maneuvering, and stealthy targets, for example, see Dmicek, de Freitas, and Gordon (2001); Ristic, Arulampalam, and Gordon (2004); and Friedland (2005). In economics, examples of reduced-form state-space models include autoregressive moving average (ARMA) specification, time-varying parameters models, stochastic volatility models, regime switching models, and dynamic factor models. For textbook references see for example Harvey (1993), and Hamilton (1994).

Of course, our main interest here is in the characterization of DSGE models as state-space representations. Such characterizations are necessary for achieving likelihood-based inference, which we describe in detail in chapters 13 and 14. Here we provide an introduction to state-space representations; characterize the mapping of DSGE models into state-space representations; provide an overview of filtering, smoothing, and likelihood evaluation; describe the Kalman filter as a means to achieving likelihood evaluation and smoothing in linear/Gaussian state-space representations; and provide examples of reduced-form state-space representations.

### 8.1 Introduction

Let  $y_t$  be a  $n \times 1$  vector of observable variables, and denote  $\{y_j\}_{j=1}^n$  as  $y_t$ . Likewise, let  $s_t$  be a  $m \times 1$  vector of unobserved ("latent") state variables, and denote  $\{s_j\}_{j=1}^m$  as  $S_t$ . State-space representations can be fully

#### 8.1.1 Introduction

characterized in terms of three probability distribution functions (densities or pdfs, for short). The first is the state-transition density

$$f(s_t | s_{t-1}, Y_{t-1}).$$

This corresponds with the state-transition equation

$$s_t = \gamma(s_{t-1}, Y_{t-1}, v_t), \quad (8.1)$$

where  $v_t$  is a vector of innovations with respect to  $(s_{t-1}, Y_{t-1})$ . The second is the observation (or measurement) density

$$f(y_t | s_t, Y_{t-1}).$$

This corresponds with the observation (or measurement) equation

$$y_t = \delta(s_t, Y_{t-1}, u_t), \quad (8.2)$$

where  $u_t$  is a vector of innovations with respect to  $(s_t, Y_{t-1})$ . As in previous chapters, we refer to  $v_t$  as structural shocks, and  $u_t$  as measurement errors. The third is the initial (or unconditional) distribution of the state

$$f(s_0) \equiv f(s_0 | Y_0).$$

The inclusion of lagged  $y$ 's as conditioning variables in the transition and observation densities is somewhat atypical, but is useful for the following presentation of the filtering and likelihood evaluation. It also embodies a level of generality sufficient, for example, to allow for Granger (1969) causality running from observables to state variables.

It shall prove convenient to express models that are linear in both (8.1) and (8.2) in the form

$$x_t = Fx_{t-1} + e_t, \quad Ee_t e_t' = Q, \quad (8.3)$$

$$X_t = H'x_t + u_t, \quad Eu_t u_t' = \Sigma_u. \quad (8.4)$$

Note that this is precisely the form used to present linear solutions to DSGE models introduced in chapter 4, where in that context  $x_t$  represents deviations of variables from steady state values.

#### 8.1.1 ARMA Models

ARMA models were introduced in chapter 7 as convenient reduced-form specifications for individual time series. Here we digress briefly to illustrate that they may be recast as state-space representations in the form of (8.3) and (8.4). Then in section 5 we present additional examples of prominent

reduced-form models that can be cast as state-space representations in the form of either (8.1) and (8.2) or (8.3) and (8.4).

To begin, consider a simple model featuring an AR(2) law of motion for the univariate state variable  $s_t$ :

$$s_t = \rho_1 s_{t-1} + \rho_2 s_{t-2} + v_t, \quad v_t \sim iidN(0, \sigma_v^2), \quad (8.5)$$

and a degenerate measurement equation for the observable variable  $y_t$ :

$$y_t = s_t. \quad (8.6)$$

This model maps into (8.3) and (8.4) as follows:

$$\begin{aligned} x_t &= \begin{bmatrix} s_t \\ s_{t-1} \end{bmatrix}, & F &= \begin{bmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{bmatrix}, & e_t &= \begin{bmatrix} v_t \\ 0 \end{bmatrix}, \\ Ee_t e_t' &= \begin{bmatrix} \sigma_v^2 & 0 \\ 0 & 0 \end{bmatrix}, & H &= \begin{bmatrix} 1 \\ 0 \end{bmatrix}, & X_t &= y_t, & u_t &= \Sigma_u = 0. \end{aligned}$$

Generalizing slightly, and mapping in the opposite direction, consider the following state-space representation:

$$x_t = \begin{bmatrix} s_{1t} \\ s_{2t} \end{bmatrix} = \begin{bmatrix} \rho_1 & \rho_2 \\ 1 & 0 \end{bmatrix} x_{t-1} + \begin{bmatrix} v_t \\ 0 \end{bmatrix}, \quad v_t \sim iidN(0, \sigma_v^2), \quad (8.7)$$

$$X_t = y_t = [1 \ \theta] x_t, \quad u_t = \Sigma_u = 0. \quad (8.8)$$

From the observation equation, we have

$$y_t = s_{1t} + \theta s_{2t},$$

as well as

$$\rho_1 y_{t-1} = \rho_1 s_{1t-1} + \rho_1 \theta s_{2t-1},$$

$$\rho_2 y_{t-2} = \rho_2 s_{1t-2} + \rho_2 \theta s_{2t-2}.$$

Then quasi-differencing, we obtain

$$\begin{aligned} y_t - \rho_1 y_{t-1} - \rho_2 y_{t-2} &= (s_{1t} - \rho_1 s_{1t-1} - \rho_2 s_{1t-2}) \\ &\quad + \theta(s_{2t} - \rho_1 s_{2t-1} - \rho_2 s_{2t-2}) \\ &= v_t + \theta v_{t-1}, \end{aligned}$$

where the second equality is implied by the law of motion for  $x_t$ . Thus we see in this case that an ARMA(2,1) specification for the observable time series  $y_t$  may be recast as a state-space representation in the form of (8.3) and (8.4). This result generalizes to the ARMA( $p,q$ ) case, as the following exercise demonstrates.

### Exercise 8.1

Show that the ARMA( $p,q$ ) representation

$$y_t - \rho_1 y_{t-1} - \cdots - \rho_p y_{t-p} = v_t + \theta_1 v_{t-1} + \cdots + \theta_q v_{t-q}$$

can be cast in the form of (8.3) and (8.4). In so doing, derive explicit expressions for  $(x_t, F, e_t, Q, X_t, H, u_t, \Sigma_u)$ .

## 8.2 DSGE Models as State-Space Representations

We have already demonstrated that (8.3) and (8.4) can be established as representing the transition and measurement equations of a state-space representation for the special case in which (8.1) and (8.2) are linear. We have also noted that (8.3) and (8.4) are in the form used to represent linear solutions to DSGE models.

Specific characterizations of nonlinear solutions to DSGE models as state-space representations are case-specific, but the procedure for constructing such representations is fairly general. Here we demonstrate the mapping of two DSGE models from their original representation as systems of nonlinear expectational difference equations into state-space representations, and then sketch a general algorithm.

Recall the RBC model of chapter 3, section 1, repeated here for convenience:

$$\left(\frac{1-\varphi}{\varphi}\right) \frac{c_t}{l_t} = (1-\alpha) z_t \left(\frac{k_t}{n_t}\right)^\alpha \quad (8.9)$$

$$c_t^\kappa l_t^\lambda = \beta E_t \left\{ c_{t+1}^\kappa l_{t+1}^\lambda \left[ \alpha z_{t+1} \left(\frac{n_{t+1}}{k_{t+1}}\right)^{1-\alpha} + (1-\delta) \right] \right\} \quad (8.10)$$

$$\zeta_t = z_t k_t^\alpha n_t^{1-\alpha} \quad (8.11)$$

$$\zeta_t = c_t + i_t \quad (8.12)$$

$$k_{t+1} = i_t + (1-\delta) k_t \quad (8.13)$$

$$l_t = n_t + l_t \quad (8.14)$$

$$\ln z_t = (1-\rho) \ln (\bar{z}) + \rho \ln z_{t-1} + \varepsilon_t, \quad (8.15)$$

where

$$\kappa = \varphi(1-\phi) - 1$$

and

$$\lambda = (1-\varphi)(1-\phi).$$

Application of any of the solution methods described in chapter 5 yields policy functions for the controls as functions of the states ( $k_t, z_t$ ); the policy functions are expressed as

$$c(k_t, z_t) \quad i(k_t, z_t), \quad \zeta(k_t, z_t), \quad n(k_t, z_t), \quad l(k_t, z_t).$$

Substituting  $i(k_t, z_t)$  for  $i_t$  in (3.19), we obtain the state-transition equations

$$\begin{aligned} k_{t+1} &= i(k_t, z_t) + (1 - \delta)k_t \\ \ln z_t &= (1 - \rho) \ln(z_0) + \rho \ln z_{t-1} + \varepsilon_t, \end{aligned}$$

which are in the form of (8.1) with  $s_t = (k_t \quad \ln z_t)'$ , and  $v_t = \varepsilon_t$ . Then working with  $y_t = (\zeta_t \quad i_t \quad n_t)'$  as the observables, (8.2) is given by

$$\begin{aligned} \zeta_t &= z_t k_t^\alpha n(k_t, z_t)^{1-\alpha} + u_{\zeta t} \\ i_t &= i(k_t, z_t) + u_{it} \\ n_t &= n(k_t, z_t) + u_{nt}, \end{aligned}$$

with, for example,  $u_t \sim N(0, \Sigma_u)$ .

Next, recall the single-asset pricing model of chapter 3, section 3.1:

$$p_t = \beta E_t \left[ \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} (d_{t+1} + p_{t+1}) \right] \quad (8.16)$$

$$c_t = d_t + q_t \quad (8.17)$$

$$d_t = \bar{d} e^{\omega_{dt}}, \quad \omega_{dt} = \rho_d \omega_{dt-1} + \varepsilon_{dt}, \quad (8.18)$$

$$q_t = \bar{q} e^{\omega_{qt}}, \quad \omega_{qt} = \rho_q \omega_{qt-1} + \varepsilon_{qt}. \quad (8.19)$$

Here,  $s_t = (\ln d_t \quad \ln q_t)'$ , and  $v_t = (\varepsilon_{dt} \quad \varepsilon_{qt})'$ . The policy function for  $c_t$  is given trivially by the equilibrium condition (8.17), and a policy function  $p(d_t, q_t)$  can be obtained as illustrated in chapter 5. The state-transition equations are given by

$$\ln d_t = (1 - \rho_d) \ln \bar{d} + \rho_d \ln d_{t-1} + \varepsilon_{dt},$$

$$\ln q_t = (1 - \rho_q) \ln \bar{q} + \rho_q \ln q_{t-1} + \varepsilon_{qt},$$

and working with  $y_t = (c_t \quad p_t \quad d_t \quad q_t)'$ , the measurement equations are given by

$$\begin{aligned} c_t &= d_t + q_t + u_{ct} \\ p_t &= p(d_t, q_t) + u_{pt} \\ d_t &= d_t + u_{dt}. \end{aligned}$$

In both examples, we have represented all observable variables as being subject to nondegenerate measurement error. However, this assumption is not necessary: in general, any or all observations can be assumed as

being made in the absence of error. In such cases, the stochastic behavior of the observables is attributed entirely to the realization of the structural shocks  $v_t$ . The only requirement is that the combined dimension of  $v_t$  and nondegenerate measurement errors be at least equal to the dimension of the observables  $y_t$ . This requirement is needed to avoid a problem known as stochastic singularity: with fewer sources of stochastic innovation than observables, various linear combinations of the observables will be predicted to be deterministic; departures from this prediction will render likelihood analysis as untenable (for details, see Ingram, Kocherlakota, and Savin, 1994).

These examples suggest the following general algorithm for mapping nonlinear representations of DSGE models into state-space representations:

- Cast the DSGE model under consideration as a system of nonlinear expectational difference equations. Do so by combining laws of motion of the states (e.g., (3.19) and (3.21)) and equations representing intratemporal constraints (e.g., (3.17) and (3.18)) with first-order conditions for the choice of control variables (e.g., (3.15) and (3.16)).
- Obtain policy functions for the controls via the application of a nonlinear solution method.
- Substitute for any control variables appearing in state-transition equations with their corresponding policy functions. This yields state-transition equations in the form of (8.1).
- Match control variables (or functions thereof) deemed as observable with corresponding policy functions (or functions thereof); optionally, attribute differences in these matches as resulting from observation errors.
- Match state variables (or functions thereof) deemed as observable with their theoretical counterparts (or functions thereof); optionally, attribute differences in these matches as resulting from observation errors.

### Exercise 8.2

Cast the New Keynesian model presented in chapter 3, section 2 in the form of a state-space representation.

## 8.3 Overview of Likelihood Evaluation and Filtering

Collecting the parameters of the state-space representation under investigation in the vector  $\mu$ , the likelihood function is denoted as

$$T(Y_{\text{obs}})$$

Due to the recursive structure of the model, the likelihood function factors sequentially as

$$L(\Upsilon_T | \mu) = \prod_{t=1}^T f(y_t | \Upsilon_{t-1}), \quad (8.20)$$

where  $f(y_t | \Upsilon_{t-1})$  denotes the time- $t$  likelihood,  $f(y_1 | \Upsilon_0) = f(y_1)$ , and the dependence of the individual likelihoods on  $\mu$  is suppressed for ease of notation.

Note that while the measurement density  $f(y_t | s_t, \Upsilon_{t-1})$  assigns probabilities to specific realizations of  $y_t$  conditionally upon both  $s_t$  and  $\Upsilon_{t-1}$ , the time- $t$  likelihood function  $f(y_t | \Upsilon_{t-1})$  is unconditional upon  $s_t$ . Unconditionality along this dimension is achieved by marginalizing over  $s_t$ :

$$f(y_t | \Upsilon_{t-1}) = \int f(y_t | s_t, \Upsilon_{t-1}) f(s_t | \Upsilon_{t-1}) ds_t, \quad (8.21)$$

where the predictive density  $f(s_t | \Upsilon_{t-1})$  indicates probabilities of particular realizations of  $s_t$  given the observables  $\Upsilon_{t-1}$ . Likewise, note that while the state-transition density  $f(s_t | s_{t-1}, \Upsilon_{t-1})$  assigns probabilities to particular realizations of  $s_t$  conditionally upon both  $s_{t-1}$  and  $\Upsilon_{t-1}$ , the predictive density  $f(s_t | \Upsilon_{t-1})$  is unconditional upon  $s_{t-1}$ . Unconditionality along this dimension is achieved by marginalizing over  $s_{t-1}$ :

$$f(s_t | \Upsilon_{t-1}) = \int f(s_t | s_{t-1}, \Upsilon_{t-1}) f(s_{t-1} | \Upsilon_{t-1}) ds_{t-1}, \quad (8.22)$$

where  $f(s_{t-1} | \Upsilon_{t-1})$  is the time- $(t-1)$  filtering density.

The filtering density is often a critical target for inference: advancing the time subscript by one period,  $f(s_t | \Upsilon_t)$  indicates probabilities of particular realizations of  $s_t$  given  $\Upsilon_t$ . It is obtained via an application of Bayes' Theorem:

$$f(s_t | \Upsilon_t) = \frac{f(y_t, s_t | \Upsilon_{t-1})}{f(y_t | \Upsilon_{t-1})} = \frac{f(y_t | s_t, \Upsilon_{t-1}) f(s_t | \Upsilon_{t-1})}{f(y_t | \Upsilon_{t-1})}. \quad (8.23)$$

The construction of likelihood and filtering densities is achieved by calculating (8.21) and (8.22) sequentially from periods 1 to  $T$ , taking as an input in period  $t$  the filtering density constructed in period  $(t-1)$ . In period 1 the filtering density corresponds with the known density  $f(s_0)$ , and thus (8.22) specializes to

$$f(s_1 | \Upsilon_0) = \int f(s_1 | s_0, \Upsilon_0) f(s_0) ds_0.$$

Note that both components of the integrand are known in this case, rendering the calculation of  $f(s_1 | \Upsilon_0)$  as feasible (either analytically or via

a numerical approximation method). Having obtained  $f(s_1 | \Upsilon_0)$ , both components of the integrand in (8.21) are known; thus the calculation of the likelihood function  $f(y_1 | \Upsilon_0)$  is feasible. Finally, given  $f(s_1 | \Upsilon_0)$  and  $f(y_1 | \Upsilon_0)$ , and with  $f(y_1 | s_1, \Upsilon_0)$  given by the model specification, the construction of  $f(s_1 | \Upsilon_1)$  obtains directly from (8.23). The same sequence of steps is executed at time  $t$ , with  $f(s_{t-1} | \Upsilon_{t-1})$  initiating the sequence.

Filtering entails the approximation of the conditional (upon  $\Upsilon_t$ ) expectation of some function  $h(s_t)$ . Given the construction of the filtering density  $f(s_t | \Upsilon_t)$ , this expectation is given by

$$E_t(h(s_t) | \Upsilon_t) = \int h(s_t) f(s_t | \Upsilon_t) ds_t. \quad (8.24)$$

Relatedly, smoothing entails the approximation of the conditional expectation of  $h(s_t)$  conditional on the full set of observations  $\Upsilon_T$ :

$$E_T(h(s_t) | \Upsilon_T) = \int h(s_t) f(s_t | \Upsilon_T) ds_t, \quad (8.25)$$

where  $f(s_t | \Upsilon_T)$  is known as a smoothing density.

For state-space representations featuring linear state-transition and measurement equations, and Normally distributed stochastic innovations and measurement errors, the construction of likelihood, filtering, and smoothing densities can be accomplished analytically using the Kalman filter; details are provided in the following section. Given departures from linearity and/or Normality, the integrals represented in (8.21), (8.22), and (8.24) become intractable analytically, and therefore must be approximated using numerical methods; moreover, in such cases the calculation of (8.25) is generally infeasible. Numerical integration techniques are introduced in chapter 9, and their application to likelihood evaluation and filtering in state-space representations is described in chapter 10.

## 8.4 The Kalman Filter

### 8.4.1 Background

Recall that an  $r$ -dimensional vector  $y$  distributed as Normal with mean  $\mu$  and covariance matrix  $\Sigma$  has a pdf given by

$$f(y | \mu, \Sigma) = (2\pi)^{-r/2} |\Sigma|^{-1/2} \exp \left[ -\frac{1}{2}(y - \mu)' \Sigma^{-1} (y - \mu) \right]. \quad (8.26)$$

Recall also that  $(\mu, \Sigma)$  represent sufficient statistics for characterizing this pdf. In working with linear/Normal state-space representations, all

densities—those inherited from the model specification, and those constructed in the sequential sequence described above—are distributed as Normal. Thus a characterization of the Kalman filter reduces to a characterization of the sequential construction of the means and covariance matrices of the likelihood, predictive, and filtering densities given in (8.21)–(8.23).<sup>1</sup>

To provide intuition for the calculations entailed in the construction of the means and covariances we seek, three background items on linear projections are useful (for greater detail, see Sargent, 1987a, or Hamilton, 1994). First, the projection of an  $n \times 1$  vector  $x$  onto an  $m \times 1$  vector  $X$  is defined as

$$\begin{aligned} P(x|X) &= a'X, \\ a &= \arg \min E[(x - a'X)^2], \end{aligned}$$

where  $a$  is an  $m \times n$  matrix of parameters. The first-order conditions for choosing  $a$  are alternatively known as Normal equations or orthogonality conditions (the variables  $(X, Y)$  are said to be orthogonal if  $E[X \cdot Y] = 0$ ). The conditions are given by

$$E[(x - a'X)X'] = 0, \quad (8.27)$$

indicating that the  $n$  errors in predicting the individual elements of  $x$  are orthogonal to the each of the  $m$  elements of  $X$ . Solving for  $a'$ , we obtain

$$a' = (E[xX'])(E[XX'])^{-1}, \quad (8.28)$$

and thus

$$P(x|X) = (E[xX'])(E[XX'])^{-1}X. \quad (8.29)$$

Second, the mean squared error (MSE) associated with the prediction of  $x$  is defined as

$$E[(x - a'X)(x - a'X)'].$$

Expanding yields

$$\begin{aligned} E[xx' - a'Xx' - (x - a'X)X'a] \\ = E[xx'] - a'E[Xx'] \\ = E[xx'] - (E[xX'])(E[XX'])^{-1}E[Xx'], \end{aligned}$$

<sup>1</sup> The GAUSS procedures `kalman.prc` and `kalmann.prc` are available for performing the calculations described in this section; the latter is for models featuring measurement errors.

#### 8.4 The Kalman Filter

where the first equality results from (8.27), and the second results from the definition of  $a$  given by (8.28). Thus the MSE associated with the prediction of  $x$  is given by

$$E[(x - a'X)(x - a'X)'] = E[xx'] - (E[xX'])(E[XX'])^{-1}E[Xx']. \quad (8.30)$$

Third, updating from  $P(x|\{X_{t-1}, \dots\})$  to  $P(x|\{X_t, X_{t-1}, \dots\})$  can be accomplished via the recursion

$$\begin{aligned} P(x|\{X_t, X_{t-1}, \dots\}) &= \underbrace{P(x|\{X_{t-1}, \dots\})}_{\text{original forecast}} \\ &\quad + P\left(\frac{x - P(x|\{X_{t-1}, \dots\})}{\text{forecast error}} \mid \frac{X_t - P(X_t|\{X_{t-1}, \dots\})}{\text{new information}}\right), \end{aligned} \quad (8.31)$$

wherein the original forecast is modified using the projection of its associated *ex post* error on the new information available in  $X_t$  relative to  $\{X_{t-1}, \dots\}$ . Equations (8.29), (8.30), and (8.31) serve as building blocks in the Kalman filter.

#### Exercise 8.3

Establish the following properties regarding projections:

- For  $(X_1, X_2)$  orthogonal,  $P(X_1|X_2) = P(X_2|X_1) = 0$ .
- Again for  $(X_1, X_2)$  orthogonal,  $P(Y|X_1, X_2) = P(Y|X_1) + P(Y|X_2)$ .
- With  $(a, b)$  representing constant parameters,  $P(aY_1 + bY_2|X) = aP(Y_1|X) + bP(Y_2|X)$ .
- $P(P(Y|\Omega, \Psi)|\Omega) = P(Y|\Omega)$ . This is the Law of Iterated Projections.

#### 8.4.2 The Sequential Algorithm

Proceeding with notation (designed following Hamilton, 1994), the conditional means and covariance matrices associated with the predictive and filtering densities of  $x_t$  are defined as

$$\begin{aligned} x_{t|t-j} &= E[x_t|\{X_1, \dots, X_{t-j}\}], \\ P_{t|t-j} &= E[(x_t - x_{t|t-j})(x_t - x_{t|t-j})'], \end{aligned}$$

$j = 0, 1$ . Thus  $(x_{t|t-1}, P_{t|t-1})$  characterizes the predictive density (8.22), and  $(x_{t|t}, P_{t|t})$  the filtering density (8.23). Likewise,  $(X_{t|t-1}, \Omega_{t|t-1})$  are the mean and covariance matrix associated with the likelihood function (8.21).

The Kalman filter is initialized by establishing  $(x_{1|0}, P_{1|0})$ . Taking the unconditional expectation of (8.3), we obtain

$$x_{1|0} \equiv Ex_1 = F \cdot Ex_0.$$

Then since  $\{x_t\}$  is covariance-stationary,  $Ex_1 = Ex_0$ , thus  $x_{1|0}$  solves

$$(I - F)x_{1|0} = 0,$$

implying

$$x_{1|0} = 0.$$

Then the expression for  $P_{1|0}$  is simply

$$\begin{aligned} P_{1|0} &= E[x_1 x_1'] \\ &= E[(Ex_0 + e_1)(Ex_0 + e_1)'] \\ &= F \cdot E[x_0 x_0'] \cdot F' + Q \\ &= F \cdot P_{1|0} \cdot F' + Q, \end{aligned}$$

where the fourth equality results from the stationarity of  $\{x_t\}$ ; the solution of  $P_{1|0}$  is given by

$$\text{vec}(P_{1|0}) = [I - F \otimes F]^{-1} \text{vec}(Q).$$

Thus, the Kalman filter is initialized using the period-0 predictive density summarized by

$$x_{1|0} = 0 \quad (8.32)$$

$$\text{vec}(P_{1|0}) = [I - F \otimes F]^{-1} \text{vec}(Q). \quad (8.33)$$

Following initialization, the period- $t$  segment of the Kalman filter contains three steps: forecasting (involving the construction of the period- $t$  likelihood function); updating (involving the construction of the period- $t$  filtering density); and prediction (involving the construction of the period- $t$  predictive density). The steps take as input the period- $(t-1)$  predictive density, initialized using  $(x_{1|0}, P_{1|0})$ .

Regarding the forecasting step, and taking as given  $(x_{t|t-1}, P_{t|t-1})$ , substituting  $x_{t|t-1}$  into (8.4) and taking expectations yields

$$X_{t|t-1} = H' x_{t|t-1}. \quad (8.34)$$

Then writing the corresponding forecast error as

$$X_t - X_{t|t-1} = H'(x_t - x_{t|t-1}) + u_t,$$

#### 8.4 The Kalman Filter

the associated mean squared error is given by

$$\begin{aligned} \Omega_{t|t-1} &= E[(H'(x_t - x_{t|t-1}) + u_t)(H'(x_t - x_{t|t-1}) + u_t)'] \\ &= E[H'(x_t - x_{t|t-1})(x_t - x_{t|t-1})' H + u_t u_t'] \\ &= H' P_{t|t-1} H + \Sigma_u. \end{aligned} \quad (8.35)$$

Regarding the updating step, from the updating equation (8.31) we obtain

$$x_{t|t} = x_{t|t-1} + E[(x_t - x_{t|t-1})|(X_t - X_{t|t-1})],$$

where in light of (8.29),

$$\begin{aligned} E[(x_t - x_{t|t-1})|(X_t - X_{t|t-1})] &= \underbrace{E[(x_t - x_{t|t-1})(X_t - X_{t|t-1})']}_{P_{t|t-1} H} \cdot \\ &\quad \underbrace{(E[(X_t - X_{t|t-1})(X_t - X_{t|t-1})'])^{-1}}_{\Omega_{t|t-1}^{-1}} \cdot \underbrace{(X_t - X_{t|t-1})}_{X_t - H' x_{t|t-1}} \\ &= P_{t|t-1} H \cdot \Omega_{t|t-1}^{-1} \cdot (X_t - H' x_{t|t-1}), \end{aligned}$$

and thus

$$x_{t|t} = x_{t|t-1} + P_{t|t-1} H \cdot \Omega_{t|t-1}^{-1} \cdot (X_t - H' x_{t|t-1}). \quad (8.36)$$

The associated mean squared error is given by

$$\begin{aligned} P_{t|t} &= E[(x_t - x_{t|t})(x_t - x_{t|t})'] \\ &= P_{t|t-1} - P_{t|t-1} H \cdot \Omega_{t|t-1}^{-1} \cdot H' P_{t|t-1}, \end{aligned} \quad (8.37)$$

which is obtained by substituting for  $x_{t|t}$  using (8.36), expanding the quadratic form, and simplifying (or alternatively, by direct application of (8.30)).

Regarding the prediction step, advancing (8.3) by one period, substituting  $x_{t+1|t}$  for  $x_t$  on the right-hand side, and taking expectations, we obtain

$$\begin{aligned} x_{t+1|t} &= Fx_{t|t} \\ &= Fx_{t|t-1} + FP_{t|t-1} H \cdot \Omega_{t|t-1}^{-1} \cdot (X_t - H' x_{t|t-1}). \end{aligned} \quad (8.38)$$

Finally, the associated mean squared error is given by

$$\begin{aligned} P_{t+1|t} &= E[(x_{t+1} - x_{t+1|t})(x_{t+1} - x_{t+1|t})'] \\ &= E[(Fx_t + e_{t+1} - Fx_{t|t})(Fx_t + e_{t+1} - Fx_{t|t})'] \\ &= FE[(x_t - x_{t|t})(x_t - x_{t|t})']F' + E[e_{t+1}e_{t+1}'] \\ &= FP_{t|t}F' + Q. \end{aligned}$$

In sum, taking as input  $(x_{t|t-1}, P_{t|t-1})$ , initialized as

$$\begin{aligned} x_{1|0} &= 0 \\ \text{vec}(P_{1|0}) &= [I - F \otimes F]^{-1} \text{vec}(Q), \end{aligned}$$

the period- $t$  likelihood function is constructed as

$$\begin{aligned} X_{t|t-1} &= H'x_{t|t-1} \\ \Omega_{t|t-1} &= H'P_{t|t-1}H + \Sigma_u; \end{aligned}$$

the period- $t$  filtering density is constructed as

$$\begin{aligned} x_{t|t} &= x_{t|t-1} + P_{t|t-1}H \cdot \Omega_{t|t-1}^{-1} \cdot (X_t - H'x_{t|t-1}) \\ P_{t|t} &= P_{t|t-1} - P_{t|t-1}H \cdot \Omega_{t|t-1}^{-1} \cdot H'P_{t|t-1}; \end{aligned}$$

and the period- $t$  predictive density is constructed as

$$\begin{aligned} x_{t+1|t} &= Fx_{t|t-1} + FP_{t|t-1}H \cdot \Omega_{t|t-1}^{-1} \cdot (X_t - H'x_{t|t-1}) \\ P_{t+1|t} &= FP_{t|t}F' + Q. \end{aligned}$$

Completion of these steps from periods 1 through  $T$  yields the sequence of likelihood functions  $\{N(X_{t|t-1}, \Omega_{t|t-1})\}_{t=1}^T$  and filtering densities  $\{N(x_{t|t}, P_{t|t})\}_{t=1}^T$ . The latter could be used to infer the time-series behavior of the unobservables, but in this case it is possible do so using smoothed values of the unobservables, which take into account the information available in the entire sample  $\{X_j\}_{j=1}^T$ , rather than through period  $t$ . Smoothed values of the unobservables are denoted as  $\{x_{t|T}\}_{t=1}^T$ .

### 8.4.3 Smoothing

To initiate the process of smoothing, consider the impact on  $x_{t|t}$  of observing  $x_{t+1}$ . Of course, we cannot actually observe  $x_{t+1}$ , a fact we shall confront below using the Law of Iterated Projections established in

### 8.4 The Kalman Filter

Exercise 8.3. But supposing for now that we could, in light of the updating equation (8.31), the impact of doing so can be expressed as

$$E[x_t | \{X_1, \dots, X_t\}, x_{t+1}] = x_{t|t} + E[(x_t - x_{t|t})(x_{t+1} - x_{t+1|t})].$$

From (8.29), the second term on the right-hand side of this expression is given by

$$\begin{aligned} E[(x_t - x_{t|t})(x_{t+1} - x_{t+1|t})] &= E[(x_t - x_{t|t})(x_{t+1} - x_{t+1|t})'] \\ &\quad \cdot (E[(x_{t+1} - x_{t+1|t})(x_{t+1} - x_{t+1|t})'])^{-1} \\ &\quad \cdot (x_{t+1} - x_{t+1|t}). \end{aligned}$$

Now, the first term in this expression simplifies as

$$\begin{aligned} E[(x_t - x_{t|t})(x_{t+1} - x_{t+1|t})'] &= E[(x_t - x_{t|t})(Fx_t + v_t - Fx_{t|t})'] \\ &= E[(x_t - x_{t|t})(x_t - x_{t|t})']F' \\ &= P_{t|t}F'; \end{aligned}$$

further, the second term simplifies as

$$(E[(x_{t+1} - x_{t+1|t})(x_{t+1} - x_{t+1|t})'])^{-1} = P_{t+1|t}^{-1}.$$

Thus the updating equation may be written as

$$\begin{aligned} E[x_t | \{X_1, \dots, X_t\}, x_{t+1}] &= x_{t|t} + P_{t|t}F'P_{t+1|t}^{-1} \cdot (x_{t+1} - x_{t+1|t}) \\ &\equiv x_{t|t} + J_t \cdot (x_{t+1} - x_{t+1|t}), \end{aligned}$$

where we have defined

$$J_t = P_{t|t}F'P_{t+1|t}^{-1}.$$

Finally, note that additional observations of  $X_{t+j}$  beyond period  $t$  do not alter inferences on  $x_t$  once  $x_{t+1}$  is known; thus the final expression for the updating expression we seek is given by

$$E[x_t | \{X_1, \dots, X_T\}, x_{t+1}] = x_{t|t} + J_t \cdot (x_{t+1} - x_{t+1|t}). \quad (8.39)$$

To operationalize (8.39) for the purpose of smoothing, we must eliminate the presence of the unobservable  $x_{t+1}$ . This is accomplished via the Law of Iterated Projections, which for our purposes here implies

$$\begin{aligned}
x_{t|T} &\equiv E[x_t | \{X_1, \dots, X_T\}] \\
&= E[E[x_t | \{X_1, \dots, X_T\}, x_{t+1}] | \{X_1, \dots, X_T\}] \\
&= E[x_{t|t} + J_t \cdot (x_{t+1|t} - x_{t+1|t})] | \{X_1, \dots, X_T\}] \\
&= x_{t|t} + J_t \cdot E[x_{t+1} | \{X_1, \dots, X_T\}] - J_t \cdot x_{t+1|t} \\
&= x_{t|t} + J_t \cdot (x_{t+1|T} - x_{t+1|t}),
\end{aligned}$$

where the fourth equality stems from the linearity property of projections established in step 3 of Exercise 8.3, coupled with the fact that  $(x_{t|t}, x_{t+1|t})$  are linear functions of  $\{X_1, \dots, X_T\}$ , and  $J_t$ .

To summarize, we have established that

$$x_{t|T} = x_{t|t} + J_t \cdot (x_{t+1|T} - x_{t+1|t}) \quad (8.40)$$

$$J_t = P_{t|t} F' P_{t+1|t}^{-1}. \quad (8.41)$$

Taking as input  $\{x_{t|t}, x_{t+1|t}, P_{t|t}, P_{t+1|t}\}_{t=1}^T$  from the filtering stage, and noting that the smoothed estimate  $x_{T|T}$  is simply the last entry of  $\{x_{t|t}\}_{t=1}^T$ , we obtain  $\{x_{t|T}\}_{t=1}^T$  via backward recursion on (8.40), beginning with

$$x_{T-1|T} = x_{T-1|T-1} + J_{T-1}(x_{T|T} - x_{T|T-1}). \quad (8.42)$$

Regarding the construction of smoothed covariance matrices  $\{P_{t|T}\}_{t=1}^T$ , these are also obtained via backward recursion. To establish the recursive formula, begin by subtracting from  $x_t$  both sides of (8.40), and then rearrange to obtain

$$x_t - x_{t|T} + J_t x_{t+1|T} = x_t - x_{t|t} + J_t x_{t+1|t}. \quad (8.43)$$

Next, multiply both sides of (8.43) by its transpose and take expectations to obtain

$$\begin{aligned}
&E[(x_t - x_{t|T})(x_t - x_{t|T})'] + J_t E[x_{t+1|T} \cdot x'_{t+1|T}] J'_t \\
&= E[(x_t - x_{t|t})(x_t - x_{t|t})'] + J_t E[x_{t+1|t} \cdot x'_{t+1|t}] J'_t,
\end{aligned}$$

where cross-product terms on both sides have been eliminated due to the orthogonality of projection errors. Employing the definitions of  $(P_{t|T}, P_{t|t})$  and rearranging, we have

$$\begin{aligned}
P_{t|T} &= P_{t|t} + J_t [-E[x_{t+1|T} \cdot x'_{t+1|T}] + E[x_{t+1|t} \cdot x'_{t+1|t}] J'_t] \\
&= P_{t|t} + J_t (P_{t+1|T} - P_{t+1|t}) J'_t,
\end{aligned} \quad (8.44)$$

where the derivation of the second equality is left as an exercise.

In parallel to the sequential construction of smoothed means, taking as input  $\{P_{t|t}, P_{t+1|t}\}_{t=1}^T$  from the filtering state, and noting that the

smoothed estimate  $P_{T|T}$  is simply the last entry of  $\{P_{t|t}\}_{t=1}^T$ , we obtain  $\{P_{t|T}\}_{t=1}^T$  via backward recursion on (8.44), beginning with

$$P_{T-1|T} = P_{T-1|T-1} + J_{T-1}(P_{T|T} - P_{T|T-1}) J'_T.$$

#### 8.4.4 Serially Correlated Measurement Errors

As an extension of the linear state-space representation we have considered, suppose that the measurement errors in (8.4) are serially correlated, obeying:

$$u_t = \Gamma u_{t-1} + \xi_t \quad (8.45)$$

$$E(\xi_t \xi_t') = \Sigma_\xi. \quad (8.46)$$

Sargent (1989) demonstrated the evaluation of DSGE models featuring diagonal specifications for the  $m \times m$  matrices  $\Gamma$  and  $\Sigma_\xi$ , and Ireland (2004b) demonstrated an extension to the nondiagonal case. In either case, the evaluation of the likelihood function associated (8.3) and (8.4) turns out to be identical to the case in which the system has no measurement error. To see why, consider an augmentation of the state system given by

$$\zeta_t = \begin{bmatrix} x_t \\ u_t \end{bmatrix} \quad \text{and} \quad \gamma_t = \begin{bmatrix} e_t \\ \xi_t \end{bmatrix}. \quad (8.47)$$

Given this augmentation, the state-space representation may be written as

$$\zeta_t = \Theta_0 \zeta_{t-1} + \gamma_t \quad (8.48)$$

$$X_t = \Theta_1 \zeta_t \quad (8.49)$$

$$\Theta_0 = \begin{bmatrix} F & 0 \\ 0 & \Gamma \end{bmatrix}, \quad \Theta_1 = \begin{bmatrix} H \\ I \end{bmatrix}' \quad (8.50)$$

$$\tilde{Q} = E(\gamma_t \gamma_t') = \begin{bmatrix} Q & 0 \\ 0 & \Sigma_\xi \end{bmatrix}, \quad (8.51)$$

which is exactly in the form of (8.3) and (8.4), but with  $u_t$  absent from (8.4). Thus replacing  $F$  and  $H$  with  $\Theta_0$  and  $\Theta_1$  in section 4.2 yields the likelihood, predictive, and filtering densities associated with this extended model. Likewise, this replacement enables the calculation of smoothed estimates of  $\zeta_t$ .

As a preview of issues to be discussed in chapter 13, we conclude this section by noting that in the absence of restrictions on the matrices  $(F, G, H, Q)$ , the likelihood function can take on the same value for more than one set of values for  $\mu$ . This is known as an identification problem. There are several ways to address this problem, one of which involves

restricting  $(F, G, H, Q)$  on the basis of theoretical restrictions on  $\mu$ , which then map into restrictions on  $(F, G, H, Q)$ . Beyond this issue, we also note that in addition to likelihood evaluation and smoothing, the Kalman filter can be used to conduct a battery of model diagnostics. We discuss these and related issues in chapter 13, paying particular attention to peculiarities that arise in the context of the analysis of DSGE models.

## 8.5 Examples of Reduced-Form State-Space Representations

Chapter 7 presented several useful reduced-form models in which all variables are observable. We conclude this chapter with additional examples of reduced-form models featuring unobservable variables.

### 8.5.1 Time-Varying Parameters

The use of time-varying parameter (TVP) models in macroeconomics dates at least to Cooley and Prescott (1973), who motivated them in part as a means of assessing Lucas' (1976) critique of econometric policy evaluation. Following Cogley and Sargent (2001, 2005), they have reemerged as a popular means of capturing changes in the dynamic behavior of aggregate economic activity. A simple yet flexible example of a TVP model is given by

$$y_t = Z_t \theta_t + \varepsilon_t, \quad (8.52)$$

$$\theta_t = F \theta_{t-1} + v_t, \quad (8.53)$$

where  $y_t$  is either a single variable or vector of observable dependent variables,  $Z_t$  is a collection of explanatory variables,  $\theta_t$  is a collection of parameters interpreted as unobservable states, and  $(\varepsilon_t, v_t)$  are stochastic innovations. Assuming  $(\varepsilon_t, v_t)$  as Normal, the linear structure of this model renders it amenable to analysis using the Kalman filter.

As an example of a very simple version of this model, Stock and Watson (2007) specify  $y_t$  as inflation, set  $Z_t = F = 1$ , and specify  $(\varepsilon_t, \eta_t)$  as serially uncorrelated, thus interpreting observations of  $y_t$  as transitory deviations from the stochastic trend  $\theta_t$ . As an example of a more elaborate version, Cogley and Sargent (2001, 2005), Primiceri (2005), and Koop and Potter (2009) specify (8.52) as a VAR specified over unemployment, short-term interest rates, and inflation. In addition, they explore the imposition of restrictions on (8.53) requiring that the roots of the VAR parameters contained in  $\theta_t$  lie outside the unit circle.

To briefly illustrate smoothing, consider a modification of Stock and Watson's model applied to the study of inflation  $\pi_t$ . Here we define  $\theta_t$  as deviations of the fundamental component of inflation  $\pi_t^f$  from the sample average  $\bar{\pi}$ , and assume this evolves as an AR(1) process:

$$\pi_t^f - \bar{\pi} = \rho(\pi_{t-1}^f - \bar{\pi}) + v_t. \quad (8.54)$$

By defining

$$x_t = \pi_t^f - \bar{\pi}, \quad v_t = \varepsilon_t,$$

we map (8.54) into (8.3), with  $F = \rho$ ,  $Q = \sigma_v^2$ . Then defining  $Z_t = 1$ ,  $y_t = \pi_t - \bar{\pi}$ , and  $\varepsilon_t = u_t$ , (8.52) is mapped into (8.4) with  $H' = 1$ ,  $\Sigma_u = \sigma_\varepsilon^2$ . Figure 8.1 illustrates  $\{\pi_t, \pi_t^f\}$  obtained using two specifications of the model. In both,  $\rho$  was set to 0.8. In the top panel,  $(Q, \Sigma_u) = (0.5^2, 1.5^2)$ ; in the bottom panel,  $(Q, \Sigma_u) = (1, 1)$  (the sample average of the standard deviation of  $\pi_t$  is roughly 2). The actual inflation series was constructed by converting monthly observations on the CPI (all urban consumers) into

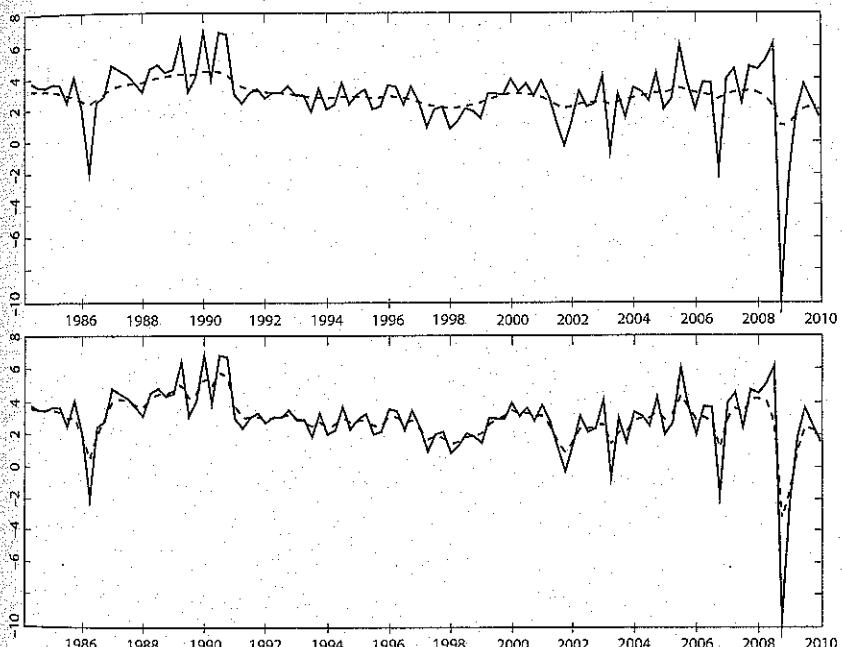


Figure 8.1 Inflation: Observed and Smoothed (dashes)

quarterly observations by averaging over the months in the quarter, and then computing

$$\pi_t = 400 \ln \left( \frac{CPI_t}{CPI_{t-1}} \right).$$

Note that when the standard deviation of the stochastic innovations  $v_t$  is relatively restricted, the inferred trajectory of the fundamental component  $\pi_t^f$  is relatively smooth. Differences in the two series are most notable during periods of high volatility.

Consider now the actual specification implemented by Stock and Watson:

$$\pi_t = \theta_t + \varepsilon_t, \quad (8.55)$$

$$\theta_t = \theta_{t-1} + v_t. \quad (8.56)$$

Despite its simplicity, implementation of this specification is complicated by the unit root implicit in the law of motion of  $\theta_t$ , rendering  $\{\theta_t, \pi_t\}$  nonstationary and thus invalidating the application of the Kalman filter. Differencing both series and combining equations yields

$$\begin{aligned} \Delta \pi_t &= v_t + \varepsilon_t - \varepsilon_{t-1} \\ &\equiv \tilde{u}_t, \end{aligned} \quad (8.57)$$

which serves to eliminate the unit root problem but introduces serial correlation in  $\tilde{u}_t$ . The correlation pattern is given by

$$E[\tilde{u}_t^2] = \sigma_v^2 + 2\sigma_\varepsilon^2, \quad (8.58)$$

$$\frac{E[\tilde{u}_t \tilde{u}_{t-1}]}{E[\tilde{u}_t^2]} = \frac{-\sigma_\varepsilon^2}{\sigma_v^2 + 2\sigma_\varepsilon^2} \quad (8.59)$$

$$\frac{E[\tilde{u}_t \tilde{u}_{t-j}]}{E[\tilde{u}_t^2]} = 0, \quad j > 1, \quad (8.60)$$

which equals that of an MA(1) specification of the form

$$u_t = \iota_t + \phi \iota_{t-1}, \quad (8.61)$$

since

$$E[u_t^2] = (1 + \phi^2)\sigma_i^2, \quad (8.62)$$

$$\frac{E[u_t u_{t-1}]}{E[u_t^2]} = \frac{\phi}{1 + \phi^2}, \quad (8.63)$$

$$\frac{E[u_t u_{t-j}]}{E[u_t^2]} = 0, \quad j > 1. \quad (8.64)$$

Matching terms in (8.58)–(8.59) with (8.62)–(8.63) and solving for  $(\phi, \sigma_i^2)$  yields

$$\phi = \frac{-b \pm \sqrt{b^2 - 4}}{2}, \quad b = \frac{1 + 2\sigma_\varepsilon^2/\sigma_v^2}{\sigma_\varepsilon^2/\sigma_v^2} \quad (8.65)$$

$$\sigma_i^2 = \frac{-1}{\phi} \sigma_\varepsilon^2, \quad (8.66)$$

where choice of the larger of the roots in (8.65) is necessary and sufficient for ensuring  $-1 < \phi < 0$ . Thus we can rewrite (8.57) as

$$\Delta \pi_t = \iota_t + \phi \iota_{t-1}, \quad (8.67)$$

with  $(\phi, \sigma_i^2)$  obeying (8.65)–(8.66). Thus in transforming the model via differencing, we have obtained an MA(1) specification for  $\Delta \pi_t$ . Use of this specification in constructing  $\pi_t^f \equiv \theta_t$  is demonstrated in the following exercise.

### Exercise 8.4

Using (8.7)–(8.8) for guidance, construct a state-space representation for  $\Delta \pi_t$  in (8.67). Obtaining the smoothed values  $\{\iota_{t|T}\}_{t=1}^T$  via the Kalman filter, and matching  $(\iota_{t-1|T}, \iota_{t|T})$  in (8.67) with  $(v_t, \varepsilon_t, \varepsilon_{t-1})$  in (8.57), construct the smoothed values  $\{v_{t|T}\}_{t=1}^T, \{\varepsilon_{t|T}\}_{t=1}^T$ . Finally, insert these values into (8.55) to obtain smoothed values of  $\{\theta_{t|T}\}_{t=1}^T \equiv \{\pi_{t|T}^f\}_{t=1}^T$ . Compare the resulting series with those illustrated in figure 8.1 for alternative specifications of  $(\sigma_\varepsilon^2, \sigma_v^2)$ .

Returning to the generic TVP specification, we conclude by noting that neither (8.52) nor (8.53) need be restricted as linear; a generalization admitting nonlinear relationships is given by

$$y_t = g(Z_t, \theta_t, \varepsilon_t), \quad (8.68)$$

$$\theta_t = s(\theta_{t-1}, \eta_t). \quad (8.69)$$

### 8.5.2 Stochastic Volatility

In chapter 7, section 1.2, we introduced ARCH and GARCH models as alternative means of capturing heteroskedasticity in the innovations of time-series models. The stochastic volatility (SV) model was introduced by Taylor (1986) as an additional, relatively flexible alternative for doing so. Specifically, Taylor's interest was in capturing the time-varying and

persistent volatility exhibited by financial returns data, in addition to fat-tailed behavior. The stochastic volatility (SV) model is given by

$$y_t = u_t \beta \exp(s_t/2) \quad (8.70)$$

$$s_{t+1} = \phi s_t + v_t, \quad (8.71)$$

where  $u_t$  and  $v_t$  are independent Gaussian random variables with variances 1 and  $\sigma^2$ ;  $\beta$  represents modal volatility; and  $\phi$  and  $\sigma^2$  determine the persistence and variance of volatility shocks. As this model is nonlinear, characterization of its empirical implementation is deferred to chapter 10.

### 8.5.3 Regime Switching

Markov processes were introduced in chapter 5 as a means of modeling the stochastic evolution of state variables that can take on a finite number of values; here we characterize these processes using state-space representations.

Consider a univariate variable  $s_t$  that can take on one of  $N$  values:  $\{1, 2, \dots, N\}$ . Let  $x_t$  be an  $N \times 1$  vector that equals the  $j^{\text{th}}$  column of the  $N$ -dimensional identity matrix  $I_N$  when  $s_t = j$ . For  $s_t = i$ , the probability that  $x_{t+1}$  takes on the value  $j$  is given by  $p_{ij}$ , which is Markov in the sense that

$$\begin{aligned} p_{ij} &= \Pr[s_{t+1} = j | s_t = i, s_{t-1} = k, \dots] \\ &= \Pr[s_{t+1} = j | s_t = i]. \end{aligned}$$

The Markov property ensures that conditional expectations regarding the next realization of the state depend only on the current value of the state:

$$E[x_{t+1} | s_t = i] = \begin{bmatrix} p_{i1} \\ p_{i2} \\ \vdots \\ p_{iN} \end{bmatrix}. \quad (8.72)$$

Let  $P$  denote the matrix containing the complete set of state-transition probabilities for  $s_t$ : its  $(i, j)^{\text{th}}$  element is  $p_{ij}$ , and its  $i^{\text{th}}$  column is the right-hand side of (8.72). Since the  $i^{\text{th}}$  row of  $x_t$  is 1 given  $s_t = i$ , (8.72) implies

$$E[x_{t+1} | x_t] = Px_t;$$

further, defining

$$v_{t+1} = x_{t+1} - E[x_{t+1} | x_t],$$

we obtain the state-transition equation

$$x_{t+1} = Px_t + v_{t+1}. \quad (8.73)$$

Hamilton (1989) used a two-state Markov specification to characterize asymmetry in the mean growth of GDP over phases of the business cycle, launching a large literature. For a recent example of an extended regime-switching model, including an overview of related literature, see DeJong, Liesenfeld, and Richard (2005). Hamilton's model combined (8.73) with an AR(4) specification of GDP growth of the form

$$\begin{aligned} g_{yt} - \mu_t &= \sum_{i=1}^4 \phi_i (g_{y(t-i)} - \mu_{t-i}) + \varepsilon_t, \\ \mu_t &= \mu' x_t, \\ \mu &= \begin{bmatrix} \mu_L \\ \mu_H \end{bmatrix}, \quad \mu_L < \mu_H. \end{aligned} \quad (8.74)$$

### 8.5.4 Dynamic Factor Models

As illustrated in section 1 of chapter 7, vector autoregressions (VARs) provide flexible yet heavily parameterized characterizations of the time-series interactions of the collection of observable variables  $\{X_t\}_{t=1}^T$ . In many contexts, these interactions likely result from the realization of a relatively small number of stochastic driving processes. For example, the RBC model introduced in section 1 of chapter 3 characterizes fluctuations in output, consumption, investment, and labor hours as resulting from innovations to a single driving process: total factor productivity. In such cases, dynamic factor models (DFMs) provide an attractive alternative to VARs, retaining their flexibility while economizing significantly on parameterization.

DFMs were developed by Geweke (1977); their application to macroeconomic time series dates to Sargent and Sims (1977). They have been used in pursuit of a wide range of empirical objectives, including the attribution of domestic and international business-cycle dynamics to a small set of common factors (e.g., Sargent and Sims, 1977; Kose, Otrok, and Whiteman, 2003); forecasting, particularly when the dimensionality of  $X_t$  is large relative to sample size (e.g., Stock and Watson, 2002a, 2002b); formal tests for common factors (developed, e.g., by Bai and Ng, 2002); and the identification of structural shocks using factor-augmented VARs (e.g., Bernanke, Boivin, and Eliasz, 2005). For surveys, see Stock and Watson (2006, 2010) and Bai and Ng (2008).

Following the notation of Stock and Watson (2010), let  $X_t$  denote an  $m \times 1$  vector of observable variables, and  $f_t$  denote a  $q \times 1$  vector of latent factors. Then a generic representation of a DFM is of the form

$$X_t = \Lambda(L)f_t + \xi_t, \quad (8.75)$$

$$\xi_t = \Theta(L)\xi_{t-1} + \varepsilon_t, \quad (8.76)$$

$$f_t = \Psi(L)f_{t-1} + \eta_t, \quad (8.77)$$

where  $(\Lambda(L), \Theta(L), \Psi(L))$  are  $(m \times q, m \times m, q \times q)$  matrices of polynomials in the lag operator  $L$ . The  $(i, j)^{\text{th}}$  element of  $\Lambda(L)$  is denoted as

$$\lambda_{i,j}(L) = \sum_{\ell=1}^{p_k} \lambda_{i,j,\ell} L^\ell;$$

the elements of  $\Theta(L)$  and  $\Psi(L)$  are denoted analogously. Typically, it is assumed that  $\{\varepsilon_t\}$  and  $\{\eta_t\}$  are independent *iid* processes; under the additional assumption of Normality, (8.75)–(8.77) constitute a linear-Normal system amenable to analysis using the Kalman filter.

It is straightforward to map (8.75)–(8.77) into the form

$$x_t = Fx_{t-1} + e_t, \quad Ee_t e_t' = Q,$$

$$X_t = H'x_t,$$

where  $(F, Q, H)$  serve as inputs to the Kalman filter, as outlined in section 4. As a simple example of this mapping, and anticipating the modest empirical application that follows, let  $X_t = (c_t \ i_t \ n_t)'$ , denoting the quarterly observations of consumption, investment, and labor hours introduced

**TABLE 8.1**  
Parameter Estimates, DFM

Coefficient	$\mu_0$	Estimate	Std. err.
$\Psi_1$	1.2	1.7373	0.0277
$\Psi_2$	-0.3	-0.8640	0.0194
$\theta_c$	0.85	0.6824	0.0515
$\theta_i$	0.85	0.4820	0.0515
$\theta_n$	0.85	0.6271	0.0339
$\lambda_{c,1}$	1	0.2905	0.0251
$\lambda_{c,2}$	0.5	-0.1616	0.0115
$\lambda_{i,1}$	1	4.9152	0.0256
$\lambda_{i,2}$	0.5	-3.6925	0.0176
$\lambda_{n,1}$	1	0.9137	0.0105
$\lambda_{n,2}$	0.5	-0.4014	0.5782
$\sigma_\eta$	0.001	0.0063	0.0003
$\sigma_c$	0.001	0.0039	0.0002
$\sigma_i$	0.001	0.0323	0.0013
$\sigma_n$	0.001	0.0024	0.0001
$\log L$	2413.29		

in chapter 6. Further, let  $f_t$  be one-dimensional, let  $(p_\lambda, p_\theta, p_\psi) = (2, 1, 2)$ , and let  $\theta_{i,j}(L) = 0$  for  $i \neq j$ . Then the mapping we seek can be achieved via the specification

$$\begin{bmatrix} f_t \\ f_{t-1} \\ \xi_{c,t} \\ \xi_{i,t} \\ \xi_{n,t} \end{bmatrix}_{x_t} = \begin{bmatrix} \psi_1 & \psi_2 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & \theta_c & 0 & 0 \\ 0 & 0 & 0 & \theta_i & 0 \\ 0 & 0 & 0 & 0 & \theta_n \end{bmatrix}_F \begin{bmatrix} f_{t-1} \\ f_{t-2} \\ \xi_{c,t-1} \\ \xi_{i,t-1} \\ \xi_{n,t-1} \end{bmatrix}_{x_{t-1}} + \begin{bmatrix} \eta_t \\ 0 \\ \varepsilon_{c,t} \\ \varepsilon_{i,t} \\ \varepsilon_{n,t} \end{bmatrix}_{e_t},$$

$$\begin{bmatrix} c_t \\ i_t \\ n_t \end{bmatrix}_{X_t} = \begin{bmatrix} \lambda_{c,1} & \lambda_{c,2} & 1 & 0 & 0 \\ \lambda_{i,1} & \lambda_{i,2} & 0 & 1 & 0 \\ \lambda_{n,1} & \lambda_{n,2} & 0 & 0 & 1 \end{bmatrix}_{H'} \begin{bmatrix} f_t \\ f_{t-1} \\ \xi_{c,t} \\ \xi_{i,t} \\ \xi_{n,t} \end{bmatrix}_{x_t}.$$

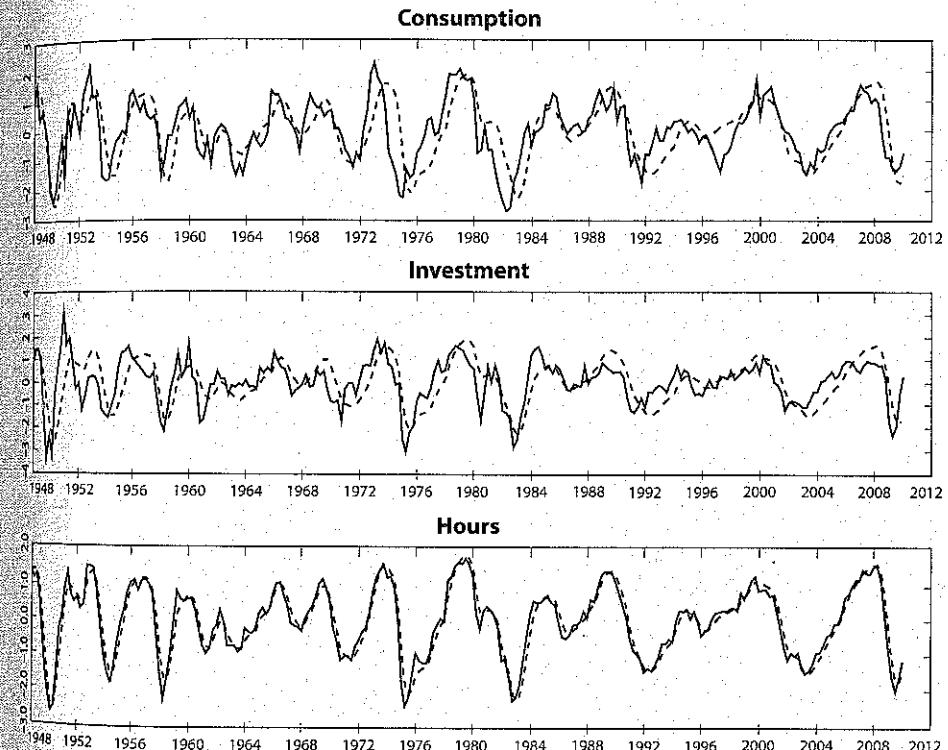


Figure 8.2 Macroeconomic Variables and Their Dynamic Factor (dashes: factor)

Regarding the application, we estimated the foregoing model via maximum likelihood using HP-filtered variables, and then obtained the smoothed time series  $\{\hat{f}_t\}_{t=1}^T$ . (See chapter 13 for details regarding ML estimation.) Parameter estimates (along with starting values, denoted as  $\mu_0$ ) are reported in table 8.1, and the relationship between the HP-filtered series and  $\hat{f}_t$  are illustrated in figure 8.2. The series in the figure are represented in terms of own-standard-deviation units.

With the exception of  $\lambda_{n,2}$ , each parameter is estimated as statistically significant, indicating nontrivial factor loadings for all three variables. Underscoring this result, figure 8.2 illustrates close correspondence between each series and the smoothed factor, with the factor leading consumption and investment, and varying contemporaneously with hours. Correlations with  $\hat{f}_t$  are  $(0.71, 0.64, 0.96)$  for  $(c_t, i_t, n_t)$ ;  $(0.53, 0.37, 0.81)$  for  $(c_{t-1}, i_{t-1}, n_{t-1})$ ; and  $(0.81, 0.79, 0.97)$  for  $(c_{t+1}, i_{t+1}, n_{t+1})$ . Thus the estimated factor accounts for a substantial component of the stochastic variation observed for these series.

## Part IV

### Monte Carlo Methods

## *Chapter 9*

### *Monte Carlo Integration: The Basics*

#### **9.1 Motivation and Overview**

In pursuing the empirical implementation of DSGE models, the need to calculate integrals arises in four distinct contexts, two of which we have already encountered. First, recall from chapter 5 that in obtaining nonlinear approximations of policy functions mapping state to control variables, the approximations we seek using projection methods are derived as the solution of nonlinear expectational difference equations. Generically, with a DSGE model represented as

$$\Gamma(E_t z_{t+1}, z_t, v_{t+1}) = 0, \quad (9.1)$$

the solution we seek is of the form

$$c_t = c(s_t), \quad (9.2)$$

$$s_t = f(s_{t-1}, v_t); \quad (9.3)$$

the policy function  $c(s_t)$  is derived as the solution to

$$F(c(s)) = 0, \quad (9.4)$$

where  $F(\cdot)$  is an operator defined over function spaces. With  $c(s_t)$  approximated by a polynomial in  $s_t$  parameterized by  $\chi - \widehat{c}(s, \chi)$  – we choose  $\chi$  to “nearly” satisfy

$$F(\widehat{c}(s, \chi)) = 0. \quad (9.5)$$

It is the need to evaluate (9.5) that entails the calculation of integrals in this context. For example, consider a simplification of the single-asset pricing environment presented in chapter 3, section 3.1, in which dividends serve as the lone stochastic process. In this case, with  $c_t = d_t$  in equilibrium, (9.5) is given by

$$\begin{aligned} & u'(d_t) \widehat{p}(d_t, \chi) - \beta E_t(u'(d_{t+1}) (\widehat{p}(d_{t+1}, \chi) + d_{t+1})) \\ &= u'(d_t) \widehat{p}(d_t, \chi) - \beta \int u'(d_{t+1}) (\widehat{p}(d_{t+1}, \chi) + d_{t+1}) dd_{t+1} = 0. \end{aligned} \quad (9.6)$$

The second context arises in seeking to accomplish likelihood evaluation and filtering given the characterization of a DSGE model as a state-space

representation. As we saw in chapter 8, with the model cast in the form of state-transition and measurement equations

$$s_t = \gamma(s_{t-1}, \gamma_{t-1}, v_t), \quad (9.7)$$

$$y_t = \delta(s_t, \gamma_{t-1}, u_t), \quad (9.8)$$

with corresponding densities given respectively by

$$f(s_t | s_{t-1}, \gamma_{t-1})$$

$$f(y_t | s_t, \gamma_{t-1}),$$

likelihood evaluation and filtering entails the sequential calculation of the period- $t$  likelihood function

$$f(y_t | \gamma_{t-1}) = \int f(y_t | s_t, \gamma_{t-1}) f(s_t | \gamma_{t-1}) ds_t, \quad (9.9)$$

where the predictive density  $f(s_t | \gamma_{t-1})$  is given by

$$f(s_t | \gamma_{t-1}) = \int f(s_t | s_{t-1}, \gamma_{t-1}) f(s_{t-1} | \gamma_{t-1}) ds_{t-1}, \quad (9.10)$$

and the filtering density  $f(s_{t-1} | \gamma_{t-1})$  is obtained via an application of Bayes' Theorem:

$$f(s_t | \gamma_t) = \frac{f(y_t, s_t | \gamma_{t-1})}{f(y_t | \gamma_{t-1})} = \frac{f(y_t | s_t, \gamma_{t-1}) f(s_t | \gamma_{t-1})}{f(y_t | \gamma_{t-1})}. \quad (9.11)$$

The third context arises in seeking to translate the stochastic behavior of  $\{s_t\}_{t=1}^T$  into unconditional expectations of functions  $f(X_t)$  of the observable variables. This context typically arises in pursuing empirical objectives using calibration and moment-matching procedures, the subjects of chapters 11 and 12. With the mapping from states to observables denoted by

$$X_t = g(s_t, u_t), \quad (9.12)$$

and given the law of motion (9.7), this translation entails the calculation of

$$\begin{aligned} Ef(X_t) &= \int \int \cdots \int f(X_t) \cdot p(v_t, v_{t-1}, \dots, v_1, s_0, u_t, u_{t-1}, \dots, u_1) \\ &\quad \times dv_t dv_{t-1} \cdots dv_1 ds_0 du_t du_{t-1} \cdots du_1, \end{aligned} \quad (9.13)$$

where  $p()$  is the joint pdf of  $\{v_t, v_{t-1}, \dots, v_1, s_0, u_t, u_{t-1}, \dots, u_1\}$ .

The final context arises in seeking to conduct posterior inference regarding functions of model parameters  $g(\mu)$ , which is the subject of chapter 14. Briefly, given the model likelihood function  $L(X|\mu)$  and prior distribution  $\pi(\mu)$ , the corresponding posterior distribution is given by

$$P(\mu|X) = L(X|\mu)\pi(\mu). \quad (9.14)$$

Posterior inference entails the calculation of the conditional (with respect to  $X$ ) expected value of a function of the parameters  $g(\mu)$ :

$$E[g(\mu)] = \int g(\mu) P(\mu|X) d\mu. \quad (9.15)$$

In special cases, the integrals characterized above can be calculated analytically. For example, as we saw in section 4 of chapter 8, the Kalman filter is available for calculating (9.9) and (9.10) in working with linear/Normal state-space representations. But in general, the integrals are intractable analytically; thus we must rely on numerical techniques to construct approximations.

Quadrature methods were introduced in chapter 5, section 1.6 as one means of constructing numerical approximations; when applicable, these methods provide fast and accurate approximations. Unfortunately, their applicability is limited to a relatively narrow range of functional forms for targeted integrands, and to problems of relatively low dimension.

In contrast, Monte Carlo integration techniques are highly flexible and applicable to a broad range of problems, including the full range of those sketched above. Coined by Metropolis and Ulam (1949), the use of Monte Carlo methods in macroeconomic applications dates to Kloek and van Dijk (1978); by now they have long become established as central components of practitioners' toolkits.

Monte Carlo integration entails the construction of approximations using pseudo-random numbers. By pseudo-random, we mean numbers generated by an algorithm designed to mimic those generated by a particular pdf. For example, Rubinstein (1981) describes algorithms for generating simulated realizations from a wide range of pdfs, and such algorithms are routinely available in software packages. Hereafter we shall refer to simulated realizations simply as random numbers. For textbook characterizations of Monte Carlo integration techniques, particularly in the context of conducting Bayesian inference, see Bauwens, Lubrano, and Richard (1999), Robert and Casella (1999), Koop (2003), and Geweke (2005).

In what follows, we shall characterize the approximation of integrals of the form

$$G(Y) = \int \varphi(\theta; Y) d\theta, \quad (9.16)$$

considering as a special case

$$G(\Upsilon) = \int \phi(\theta; \Upsilon) p(\Upsilon|\theta) d\theta. \quad (9.17)$$

Use of this notation represents a purposeful deviation from the form of the specific integrals introduced above, but each can be seen as encompassed by either (9.16) or (9.17) as a special case.

## 9.2 Direct Monte Carlo Integration

In the simplest case, the targeted integrand can be factorized as in (9.17), and it is possible to simulate random drawings of  $\theta$  directly from  $p(\Upsilon|\theta)$ . In this case, (9.17) can be approximated via direct Monte Carlo integration. Let  $\theta_i$  denote the  $i^{\text{th}}$  of a sequence of  $N$  draws obtained from  $p(\Upsilon|\theta)$ . Then by the law of large numbers (e.g., as in Geweke, 1989), the Monte Carlo estimate

$$\overline{G(\Upsilon)}_N = \frac{1}{N} \sum_{i=1}^N \phi(\theta_i; \Upsilon) \quad (9.18)$$

converges in probability to  $G(\Upsilon)$ . In addition, so long as the expectation of  $\phi(\theta; \Upsilon)^2$  is finite,  $\overline{G(\Upsilon)}_N$  is approximately distributed as

$$\frac{\overline{G(\Upsilon)}_N - G(\Upsilon)}{\sigma(G(\Upsilon))/\sqrt{N}} \sim N(0, 1), \quad (9.19)$$

where  $\sigma(G(\Upsilon))$  is the standard deviation of  $G(\Upsilon)$ , and

$$\frac{\sigma(G(\Upsilon))}{\sqrt{N}} \equiv s.e.(\overline{G(\Upsilon)}_N) \quad (9.20)$$

is known as the numerical standard error (NSE) associated with  $\overline{G(\Upsilon)}_N$ . Defining  $\overline{\sigma}_N(G(\Upsilon))$  as the sample standard deviation of  $G(\Upsilon)$ , given by

$$\overline{\sigma}_N(G(\Upsilon)) = \left[ \left( \frac{1}{N} \sum_i^N \phi(\theta_i; \Upsilon)^2 \right) - \overline{G(\Upsilon)}_N^2 \right]^{1/2}, \quad (9.21)$$

$s.e.(\overline{G(\Upsilon)}_N)$  is estimated in practice using

$$s.e.(\overline{G(\Upsilon)}_N) = \frac{\overline{\sigma}_N(G(\Upsilon))}{\sqrt{N}}. \quad (9.22)$$

A typical choice for  $N$  is 10,000, in which case  $1/\sqrt{N} = 1\%$ .

Note carefully the interpretation of NSE: this indicates the size of potential approximation errors associated with  $\overline{G(\Upsilon)}_N$ . It is closely associated with the notion of numerical efficiency, which pertains to the size of the artificial sample  $N$  needed to achieve a particular level of numerical precision, quantified using NSE. As we shall see, the NSE associated with the direct Monte Carlo estimator  $\overline{G(\Upsilon)}_N$  in (9.18) represents a best-case scenario in approximating (9.17).

As a simple but important example of direct Monte Carlo integration, consider the context under which (9.17) specializes to (9.15), and consider the objective of approximating the marginal pdf of  $g(\mu)$  via direct Monte Carlo integration. This can be accomplished via the construction of a histogram using the following steps. First, partition the range of  $g(\mu)$  into equally spaced intervals, and let  $p_k$  denote the probability assigned by the actual pdf to a realization of  $g(\mu)$  within the  $k^{\text{th}}$  interval. Next, let  $\iota_k(g(\mu))$  denote an indicator function that equals 1 when  $g(\mu)$  falls within the  $k^{\text{th}}$  interval and 0 otherwise. Then approximate  $p_k$  using

$$\overline{p_k} = \left( \frac{1}{N} \right) \sum_{i=1}^N \iota_k(g(\mu_i)); \quad (9.23)$$

the assignment of  $\overline{p_k}$  as the height of the histogram over bin  $k$  yields the desired approximation.<sup>1</sup> Moreover, interpreting each drawing  $\mu_i$  as a Bernoulli trial with success defined as  $\iota_k(g(\mu_i)) = 1$ , so that the probability of success is given by  $p_k$ , Geweke (1989) notes that the standard error associated with  $\overline{p_k}$  is given by

$$s.e.(\overline{p_k}) = \left( \frac{p_k(1-p_k)}{N} \right)^{1/2}; \quad (9.24)$$

thus the standard error of  $\overline{p_k}$  given  $N = 10,000$  is no greater than 0.005.

### Exercise 9.1

Let  $\mu$  consist of a single element,  $P(\mu|X)$  be a  $N(0, 1)$  pdf, and  $g(\mu) = \mu^2$ . Using a histogram consisting of 11 equally spaced bins over the range  $[0, 8]$ , approximate the pdf of  $g(\mu)$  for  $N = 100, 1,000, 10,000$ . Compare your results with a  $\chi^2(1)$  distribution: this is what you are approximating! Repeat this exercise using histograms consisting of 21, 51, and 101 bins. (Be sure to incorporate the normalizing factor  $K/R$  mentioned in footnote 1 to ensure your approximations integrate to 1.) Note how the quality of the

<sup>1</sup> Denoting the number of bins comprising the approximating histogram by  $K$ , and the range spanned by the histogram by  $R$ , it is necessary to adjust  $\overline{p_k}$  by the factor  $K/R$  to ensure that the histogram integrates to 1.

approximations varies as a function of  $N$  and the number of bins you use; why is this so?

Section 2 of chapter 14 describes an important class of empirical applications under which integrals of interest can be approximated via direct Monte Carlo integration. Under this class of applications, structural models are used as sources of prior information to be combined with likelihood functions associated with reduced-form models, typically in pursuit of forecasting applications. Judicious choices of prior distributions in this case can yield posterior distributions from which sample realizations of  $\mu$  can be simulated.

Another important class of applications amenable to direct Monte Carlo integration involves the simulation of data using a DSGE model as a data generation process (DGP). This is done for the purpose of approximating unconditional expectations of functions  $f(X_t)$  of the observable variables of the DSGE model, as in (9.13). As this objective remains central to the bulk of empirical work conducted using DSGE models, we elaborate on this special case in the following subsection.

### 9.2.1 Model Simulation

The goal of model simulation is to approximate  $Ef(X_t)$  based on a simulated realization of  $\{X_t\}_{t=1}^T$  constructed from drawings  $\{v_t, v_{t-1}, \dots, v_1, s_0, u_t, u_{t-1}, \dots, u_1\}_{t=1}^T$  obtained from  $p(v_t, v_{t-1}, \dots, v_1, s_0, u_t, u_{t-1}, \dots, u_1)$  in (9.13). Since the model is stationary by construction, sample averages

$$\overline{f(X)}_T = \frac{1}{T} \sum_{t=1}^T f(X_t)$$

can be used to approximate the unconditional moment  $Ef(X_t)$ . Thus so long as

$$p(v_t, v_{t-1}, \dots, v_1, s_0, u_t, u_{t-1}, \dots, u_1)$$

can be used to generate drawings

$$\{v_t, v_{t-1}, \dots, v_1, s_0, u_t, u_{t-1}, \dots, u_1\}_{t=1}^T,$$

model simulation is a particular application of direct Monte Carlo integration.

The recursive nature of DSGE models renders the task of model simulation straightforward; for details regarding the accuracy of simulations involving DSGE models, see Santos and Peralta-Alva (2005).

### 9.2 Direct Monte Carlo Integration

The list of inputs to the simulation process includes a specification of the initial state vector  $s_0$ , which can be taken either as known or as a realization obtained from the initializing distribution  $f(s_0)$  inherent in the specification of a state-space representation. The stationary nature of the model renders this distinction as unimportant in simulations designed to characterize the asymptotic properties of the model. Additional inputs include the law of motion (5.10) for  $s_t$ , the approximated policy function  $\tilde{\pi}(s_t)$ , the mapping (9.12) from model variables  $x_t$  to observables  $X_t$ , and a random number generator used to obtain artificial drawings of stochastic shocks (and if appropriate, measurement errors) from their underlying distributions. The parameterization of these inputs is determined by a specific specification chosen for the structural parameters  $\mu$ .

Given these inputs, model simulation proceeds as follows. From (5.10),  $s_0$  and  $\{v_t\}_{t=1}^T$  yield  $\{s_t\}_{t=1}^T$  directly. Using this series as an input to the policy function yields  $\{c_t\}_{t=1}^T$ . Then  $\{s_t\}_{t=1}^T$  and  $\{c_t\}_{t=1}^T$  combine to form  $\{x_t\}_{t=1}^T$  ( $x_t \equiv (s'_t, c'_t)$ ), which combined with  $\{u_t\}_{t=1}^T$  maps into  $X_t$  as in (9.12). Functions of interest can then be computed using the simulated realizations of  $\{X_t\}_{t=1}^T$ .

If  $s_0$  is not obtained as a drawing from  $f(s_0)$ , but is instead assigned an arbitrary value, the influence of its selection can be eliminated using what is known as a “burn-in” phase. This merely involves discarding the first, say, 1,000 artificial realizations obtained for  $\{s_t\}_{t=1}^T$ . Finally, numerical standard errors can be virtually eliminated using a sufficiently large number of artificial realizations.

A novelty in this case arises when using a Markov process to represent the evolution of a stochastic process. For example, recall from the specification of the optimal growth model described in chapter 5 that a Markov process was used to represent the behavior of total factor productivity  $\{a_t\}_{t=1}^T$ . A simple algorithm for generating artificial realizations of  $\{a_t\}_{t=1}^T$  for the two-state case is as follows (extension to the  $n$ -state case is straightforward). Begin with a particular specification of  $\xi_0$ , the  $2 \times 1$  vector with a 1 in the first row and a zero in the second row given state one (e.g., the low state for  $a_t$ ), and a 0 in the first row and a 1 in the second row given state two (e.g., the high state). Set  $a_0$  accordingly (e.g., either  $a_0 = aL$  or  $a_0 = aH$ ). With  $P$  denoting the state-transition matrix, the probabilities of realizing states one and two next period are given by  $P\xi_0 = [p_1, p_2]'$ . To determine the outcome of this random event, draw a random variable  $u$  distributed uniformly over the  $[0, 1]$  interval (e.g., using the GAUSS command `rndu`), and infer the realization of state one if  $p_1 > u$  and state two otherwise. Update  $\xi_1$  accordingly and repeat until  $T$  realizations of  $a_t$  are obtained.

To calibrate the Markov process to match an AR(1) specification

$$a_t = (1 - \rho)\bar{a} + \rho a_{t-1} + \varepsilon_t, \quad s.e.(\varepsilon) = \sigma_\varepsilon, \quad (9.25)$$

set the diagonal elements of  $P$  as

$$p_{11} = p_{22} = \frac{\rho + 1}{2}, \quad (9.26)$$

and the values for  $a_t$  in states one and two as

$$a_1 = \bar{a} - \sigma_a, \quad a_2 = \bar{a} + \sigma_a, \quad (9.27)$$

$$\sigma_a = \sigma_\varepsilon \sqrt{\left(\frac{1}{1 - \rho^2}\right)}. \quad (9.28)$$

Under this specification,  $a_t$  bounces one standard deviation above and below its mean value  $\bar{a}$ .

### SIMULATING THE OPTIMAL GROWTH MODEL

To explore potential differences in model inferences that can arise in working with nonlinear rather than linear model approximations, table 9.1 presents a collection of moments calculated using linear and nonlinear approximations of the stochastic growth model presented in chapter 5. Both sets of simulations are based on the use of steady state values for  $s_0$ , 1,000 burn-in drawings, and 10,000 retained drawings. Simulations of the linear approximation were based on the use of an AR(1) specification for  $a_t$ , under the assumption of Normality for its innovations  $v_t$ . Simulations of

TABLE 9.1  
Model Simulations

	Linear approximation				Nonlinear approximation			
$j$	$\sigma_j$	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\sigma_j$	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$
$y$	0.0153	1.00	0.90	1.00	0.0151	1.00	0.89	1.00
$c$	0.0110	0.72	0.97	0.95	0.0109	0.72	0.96	0.95
$i$	0.0330	2.16	0.80	0.94	0.0330	2.19	0.79	0.94
$k$	0.0189	1.24	0.99	0.86	0.0184	1.22	0.99	0.76
$a$	0.0112	0.73	0.80	0.93	0.0112	0.74	0.80	0.94

Note:  $\varphi(1)$  denotes first-order serial correlation;  $\varphi_{j,y}(0)$  denotes contemporaneous correlation between variables  $j$  and  $y$ . Model moments based on the parameterization

$$\mu = [\alpha \beta \delta \phi \rho \sigma]' = [0.33 0.96 0.1 2.0 0.8 0.0067]'$$

### 9.2 Direct Monte Carlo Integration

the nonlinear approximation were based on the use of a two-state Markov process, as described above. In both cases, moments were calculated for logged deviations of the model variables from steady state values.

In this case, the moment conditions turn out to be quite similar across solution methods. The one notable difference is in the correlation between  $k$  and  $y$ , which is relatively low under the nonlinear approximation (0.76 versus 0.86 in the linear approximation). This comparison indicates that, as parameterized, the RBC model is well-approximated as linear. Exploration of this result is the subject of the following exercise.

#### Exercise 9.2

As an alternative to the two-state Markov process specified for  $a_t$  under the nonlinear approximation, reconstruct table 9.1 using an AR(1) specification, parameterized as indicated in table 9.1, along with the associated policy function approximation constructed using `ara.prg`. Under this exercise, you should be able to match closely the calculations reported in the table. Next, as noted, the model is well-approximated as linear under the parameterization employed in the table. Explore the sensitivity of this result to alternative parameterizations.

#### Exercise 9.3

Consider the problem of approximating the integral

$$\int u'(d_{t+1}) (\hat{p}(d_{t+1}, x) + d_{t+1}) dd_{t+1}$$

in (9.6). Taking  $d_t$  as given, along with the law of motion

$$d_{t+1} = (1 - \rho)\bar{d} + \rho d_t + \sigma \varepsilon_{t+1}, \quad \varepsilon_{t+1} \sim N(0, \sigma_\varepsilon^2),$$

sketch an algorithm for achieving this approximation via direct Monte Carlo integration.

#### 9.2.2 Posterior Inference via Direct Monte Carlo Integration

We conclude this section by noting the feasibility of conducting posterior inference via direct Monte Carlo integration. Feasibility in this context attains in working with linear/Normal model representations, under which likelihood evaluation is available analytically via the Kalman filter. (Recall that given departures from linearity and Normality, the densities  $f(s_t | Y_{t-1})$  and  $f(s_{t-1} | Y_{t-1})$  appearing in (9.9) and (9.10) do not have analytical expressions, rendering likelihood evaluation via direct Monte Carlo integration as infeasible.)

Recall that linear and log-linear model approximations are of the form

$$\begin{aligned}x_{t+1} &= F(\mu)x_t + G(\mu)v_{t+1}, \\e_{t+1} &= G(\mu)v_{t+1}, \\Q(\mu) &= E(e_{t+1}e'_{t+1}), \\X_t &= H(\mu)'x_t + u_t, \\E(u_tu'_t) &= \Sigma_u.\end{aligned}$$

Defining the collection of likelihood parameters as

$$\Lambda(\mu) \equiv \{F(\mu), G(\mu), Q(\mu), H(\mu), \Sigma_u\},$$

note that the likelihood function appearing in (9.15) is defined over  $\mu$  indirectly:

$$L(X|\mu) = L(X|\Lambda(\mu)),$$

where the mapping from  $\mu$  to  $\Lambda$  is achieved via the solution techniques outlined in chapter 4.

Rewriting (9.15) as

$$\begin{aligned}E[g(\mu)] &= \int g(\mu)P(\mu|X)d\mu \\&= \int g(\mu)L(X|\Lambda(\mu))\pi(\mu)d\mu,\end{aligned}\tag{9.29}$$

so long as the prior distribution  $\pi(\mu)$  can be used to obtain drawings sample drawings  $\mu_i$ , it can be implemented to approximate  $E[g(\mu)]$  as

$$\overline{g(\mu)}_N = \frac{1}{N} \sum_{i=1}^N g(\mu_i)L(X|\Lambda(\mu_i)).\tag{9.30}$$

However, since the likelihood function is parameterized directly over  $\Lambda$ , while the prior is specified in terms of  $\mu$ , it is not possible to implement the posterior  $P(\mu|X)$  as a sampler for generating drawings  $\mu_i$ . Moreover, for reasons described below, absent near proportionality between  $\pi(\mu)$  and  $g(\mu)L(X|\Lambda(\mu))$ , the approximation  $\overline{g(\mu)}_N$  will be numerically inefficient. Details follow.

### 9.3 Importance Sampling

Importance sampling is closely related to direct Monte Carlo integration. In approximating integrals of the form

$$G(Y) = \int \varphi(\theta; Y)d\theta,\tag{9.31}$$

the idea behind this approach is to recast the problem in a form amenable to direct Monte Carlo integration. This is achieved by augmenting (9.31) to include a density  $g(\theta; a)$  that is used to generate sample drawings  $\{\theta_i\}$ . Although the probability of a given draw  $\theta_i$  is influenced directly by the choice of  $g(\theta; a)$ , each draw is weighted so that this influence is eliminated. The density  $g(\theta; a)$  is called an importance sampler, where  $a$  represents its parameterization.

The importance sampler is introduced into the problem via the augmentation

$$\begin{aligned}G(Y) &= \int \frac{\varphi(\theta; Y)}{g(\theta; a)}g(\theta; a)d\theta \\&\equiv \int \omega(\theta)g(\theta; a)d\theta,\end{aligned}\tag{9.32}$$

where the weight function  $\omega(\theta)$  is defined as

$$\omega(\theta, Y) = \frac{\varphi(\theta; Y)}{g(\theta; a)}.\tag{9.33}$$

Note that (9.32) is a special case of (9.17), which as we have seen is amenable to direct Monte Carlo integration. This is also the case here: so long as the support of  $g(\theta; a)$  spans that of  $\varphi(\theta; Y)$ , and  $G(Y)$  exists and is finite, then the law of large number implies once again that  $G(Y)$  can be approximated as

$$\overline{G(Y)}_{N,g} = \frac{1}{N} \sum_{i=1}^N \omega(\theta_i, Y),\tag{9.34}$$

$$\omega(\theta_i, Y) = \frac{\varphi(\theta_i; Y)}{g(\theta_i; a)};\tag{9.35}$$

for details, see, for example, Geweke (1989). The presence of  $g(\theta; a)$  in  $\omega(\theta, Y)$  serves to downweight relatively high-probability draws from the perspective of  $g(\theta; a)$ , and upweight relatively low-probability draws, thus offsetting the direct influence of  $g(\theta; a)$  in generating  $\{\theta_i\}$ .

The numerical standard error associated with  $\overline{G(\theta, Y)}_{N,g}$  is given by

$$s.e.(\overline{G(Y)}_{N,g}) = \frac{\bar{\sigma}_{N,g}(\omega(\theta, Y))}{\sqrt{N}},\tag{9.36}$$

where  $\bar{\sigma}_{N,g}(\omega(\theta))$  denotes the Monte Carlo estimate of the standard deviation of  $\omega(\theta, Y)$ :

$$\bar{\sigma}_{N,g}(\omega(\theta, Y)) = \left[ \left( \frac{1}{N} \sum \omega(\theta_i, Y)^2 \right) - \overline{G(Y)}_{N,g}^2 \right]^{1/2}.\tag{9.37}$$

Note that variability in  $\omega(\theta; \Upsilon)$  reflects a mismatch between the contours of the importance sampler  $g(\theta; \alpha)$  and the targeted integrand  $p(\theta; \Upsilon)$ . In turn, this variability translates into increases in numerical standard errors, or correspondingly, deteriorations in numerical efficiency. Methods for achieving improvements in the efficiency of  $g(\theta; \alpha)$  are discussed below.

To build further intuition behind importance sampling, consider the special case in which the integrand factorizes as  $\phi(\theta; \Upsilon)p(\theta| \Upsilon)$ , and  $p(\theta| \Upsilon)$  is interpreted as a posterior distribution. In this case the augmented integral we face can be written as

$$\begin{aligned} G(\Upsilon) &= \int \phi(\theta; \Upsilon) \frac{p(\theta| \Upsilon)}{g(\theta; \alpha)} g(\theta; \alpha) d\theta \\ &\equiv \int \phi(\theta; \Upsilon) \bar{w}(\theta, \Upsilon) g(\theta; \alpha) d\theta, \end{aligned} \quad (9.38)$$

where

$$\bar{w}(\theta, \Upsilon) = \frac{p(\theta| \Upsilon)}{g(\theta; \alpha)}. \quad (9.39)$$

The corresponding Monte Carlo approximation of  $G(\Upsilon)$  is given by

$$\overline{G(\Upsilon)}_{N,g} = \frac{1}{N} \sum_{i=1}^N \phi(\theta_i; \Upsilon) \bar{w}(\theta_i, \Upsilon), \quad (9.40)$$

$$\bar{w}(\theta_i, \Upsilon) = \frac{p(\theta_i| \Upsilon)}{g(\theta_i; \alpha)}. \quad (9.41)$$

The corresponding NSE of  $\overline{G(\Upsilon)}_{N,g}$  is given by (9.36) and (9.37), with

$$\omega(\theta_i, \Upsilon) = \phi(\theta_i; \Upsilon) \bar{w}(\theta_i, \Upsilon). \quad (9.42)$$

Finally, the sample standard deviation of  $G(\Upsilon)$  is given by

$$\overline{\sigma}_{N,g}(G(\Upsilon)) = \left[ \left( \frac{1}{N} \sum_i^N \phi(\theta_i; \Upsilon)^2 \bar{w}(\theta_i, \Upsilon) \right) - \overline{G(\Upsilon)}_{N,g}^2 \right]^{1/2}. \quad (9.43)$$

Note in this case that  $g(\theta; \alpha)$  is being used as a stand-in for  $p(\theta| \Upsilon)$ , and that a mismatch between the two will be reflected by variability in  $w(\theta_i; \Upsilon)$ . Given perfect fit, this variability is eliminated,  $w(\theta_i; \Upsilon) = 1 \forall i$ , and we have attained the best-case scenario of direct Monte Carlo integration.

In practice, off-the-rack samplers will rarely deliver anything close to this best-case scenario: a good sampler requires tailoring. The consequence of

working with a poor sampler is slow convergence to the targeted integral, or equivalently, NSEs that decay slowly as  $N$  increases. The problem in this context is that if the importance sampler is ill-fitted to the posterior, enormous weights will be assigned occasionally to particular drawings, and a large share of drawings will receive virtually no weight. Fortunately, it is easy to diagnose an ill-fitted sampler: rather than being approximately uniform, the weights it generates will be concentrated over a small fraction of drawings, a phenomenon easily detected via the use of a histogram plotted for the weights.

A formal diagnostic measure, proposed by Geweke (1989), is designed to assess the relative numerical efficiency (RNE) of the sampler. This is given by the ratio of the squared numerical standard errors of the optimal and actual samplers:

$$RNE = \frac{(\text{optimal NSE})^2}{(\text{actual NSE})^2}. \quad (9.44)$$

Recall that the optimal NSE (i.e., that obtained under direct Monte Carlo integration) is given by

$$s.e.(\overline{G(\Upsilon)}_{N,g}) = \frac{\overline{\sigma}_{N,g}(G(\Upsilon))}{\sqrt{N}},$$

and thus

$$RNE = \frac{(\overline{\sigma}_{N,g}(G(\Upsilon)))^2}{(s.e.(\overline{G(\Upsilon)}_{N,g}))^2}. \quad (9.45)$$

Regarding the actual NSE, note from (9.42) that in the optimal case of  $\bar{w}(\theta_i, \Upsilon) = 1 \forall i$ , we obtain  $\omega(\theta_i, \Upsilon) = \phi(\theta_i; \Upsilon)$ . Thus from (9.36) and (9.37), the NSE of  $\overline{G(\Upsilon)}_{N,g}$  is seen as identical to that corresponding with its direct Monte Carlo counterpart, and RNE equals 1. More generally, solving for the NSE of  $\overline{G(\Upsilon)}_{N,g}$ , we obtain

$$s.e.(\overline{G(\Upsilon)}_{N,g}) = \frac{\overline{\sigma}_{N,g}(G(\Upsilon))}{\sqrt{N \cdot RNE}}. \quad (9.46)$$

Comparing  $s.e.(\overline{G(\Upsilon)}_{N,g})$  in (9.46) with  $s.e.(\overline{G(\Upsilon)}_N)$  in (9.22), note that  $\sqrt{RNE \cdot N}$  has replaced  $\sqrt{N}$  as the denominator term. Thus RNE indicates the extent to which the realization of unequally distributed weights has reduced the effective number of drawings used to compute  $\overline{G(\Upsilon)}_{N,g}$ .

As a final note regarding the special case in which the targeted integrand factorizes as  $\phi(\theta; \Upsilon)p(\theta| \Upsilon)$ , and  $p(\theta| \Upsilon)$  is interpreted as

a posterior distribution, it is possible to ignore integrating constants in constructing  $(\overline{G(\Upsilon)})_{N,g}, \overline{\sigma}_{N,g}(G(\Upsilon))$ . This is facilitated by modifying (9.40) and (9.43) as

$$\overline{G(\Upsilon)}_{N,g} = \left( \frac{1}{\sum \bar{w}(\theta_i, \Upsilon)} \right) \sum_{i=1}^N \phi(\theta_i; \Upsilon) \bar{w}(\theta_i, \Upsilon) \quad (9.47)$$

$$\overline{\sigma}_{N,g}(G(\Upsilon)) = \left[ \left( \left( \frac{1}{\sum \bar{w}(\theta_i, \Upsilon)} \right) \sum_{i=1}^N \phi(\theta_i; \Upsilon)^2 \bar{w}(\theta_i, \Upsilon) \right) - \overline{G(\Upsilon)}_{N,g}^2 \right]^{1/2} \quad (9.48)$$

Note that the impact of ignoring integrating constants is eliminated by the inclusion of the accumulated weights in the denominators of these expressions.

### 9.3.1 Achieving Efficiency: A First Pass

As noted, in order for  $g(\theta; \alpha)$  to serve as a valid sampler for approximating  $G(\Upsilon)$ , two modest conditions must be met: its support must span that of the targeted integrand  $\varphi(\theta; \Upsilon)$ , and  $G(\Upsilon)$  must exist and be finite. However, for  $g(\theta; \alpha)$  to serve as an effective sampler in practice, it must be tailored to the contours of  $\varphi(\theta; \Upsilon)$  in order to improve numerical efficiency.

To underscore the importance of tailoring, consider a simple example involving the calculation of

$$E[g(\mu)] = \int g(\mu) P(\mu|X) d\mu$$

using an importance sampler  $g(\theta; \alpha)$ . Specifically, let  $P(\mu|X)$  be distributed as  $N(\mu, \Sigma)$ , with

$$\mu = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{bmatrix}, \quad \text{corr}(\Sigma) = \begin{bmatrix} 1 & 0.6 & 0 & 0 & 0 \\ 0.6 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & -0.8 \\ 0 & 0 & 0 & -0.8 & 1 \end{bmatrix},$$

$$\text{sqrt}(\text{diag}(\Sigma))' = [ 2 \ 0.2 \ 5 \ 1 \ 0.1 ],$$

where  $\text{sqrt}(\text{diag}(\Sigma))$  is the square root of the diagonal elements of  $\Sigma$  (i.e., the standard deviations implicit in  $\Sigma$ ). Further, let  $g(\mu)$  represent three specific functions:

$$g(\mu) = \left( \mu \ \sum_{i=1}^5 \mu_i \ \prod_{i=1}^5 \mu_i \right)'.$$

TABLE 9.2  
Monte Carlo Experiment

	$\overline{G(\Upsilon)}_N$	$\overline{\sigma}_N(G(\Upsilon))$	$s.e.(G(\Upsilon)_N)$	$\overline{G(\Upsilon)}_{N,g}$	$\overline{\sigma}_{N,g}(G(\Upsilon))$	$s.e.(\overline{G(\Upsilon)}_{N,g})$	RNE
$E\mu_1$	1.0077	1.9664	0.0197	1.1646	1.5831	0.4140	0.0023
$E\mu_2$	2.0031	0.2007	0.0020	1.9884	0.1721	0.0572	0.0012
$E\mu_3$	2.9157	4.9899	0.0499	3.1161	3.5692	0.6526	0.0058
$E\mu_4$	4.0041	0.9793	0.0098	4.2545	0.8731	0.2691	0.0013
$E\mu_5$	4.9993	0.0986	0.0010	4.9947	0.0803	0.0173	0.0033
$E \sum_{i=1}^5 \mu_i$	14.9299	5.4883	0.0549	15.5184	3.9212	0.8116	0.0023
$E \prod_{i=1}^5 \mu_i$	103.7081	533.3792	5.3338	168.4666	471.0707	64.8759	0.0053

Table 9.2 presents two sets of estimates of these functions obtained using  $N = 10,000$ : one set obtained via direct Monte Carlo integration, and a second set obtained using the importance sampler  $g(\theta; \alpha)$  specified as  $N(\mu_g, \Sigma_g)$ , with

$$\begin{aligned} \mu_g &= \mu + 1.5 \cdot \text{sqrt}(\text{diag}(\Sigma)); \\ \text{corr}(\Sigma_g) &= 0 \quad \forall i, j; \quad \text{sqrt}(\text{diag}(\Sigma_g)) = \text{sqrt}(\text{diag}(\Sigma)). \end{aligned}$$

In short,  $g(\theta; \alpha)$  is mildly right-shifted relative to  $P(\mu|X)$ , has identically thick tails, and fails to capture correlation among the individual elements of  $\mu$ .

The RNE statistics indicate that  $g(\theta; \alpha)$  is highly inefficient. Specifically, they indicate that  $N$  must be scaled by a factor ranging from approximately 200 to 850 in order to attain the numerical accuracy associated with the direct sampler. As a further indication of inefficiency, figure 9.1 plots the weights  $\bar{w}(\theta_i, \Upsilon)$  assigned to the drawings  $\{\theta_i\}$  generated by  $g(\theta; \alpha)$ , sorted from lowest to highest in moving from left to right. The top panel plots all 10,000 draws; the second panel plots the top 1,000 draws, and the third panel plots the top 100 draws. As the figure illustrates, only the top 30 draws receive discernible weight in this case. In light of the simplicity of this example problem, the inefficiency of  $g(\theta; \alpha)$  is striking.

### Exercise 9.4

Replicate the results of table 9.2. In so doing, use the following GAUSS commands for obtaining drawings from an  $N(\mu, \text{sig})$  distribution:

- `swish = chol(sig)'`; (`swish` is the lower-diagonal Cholesky decomposition of `sig`).

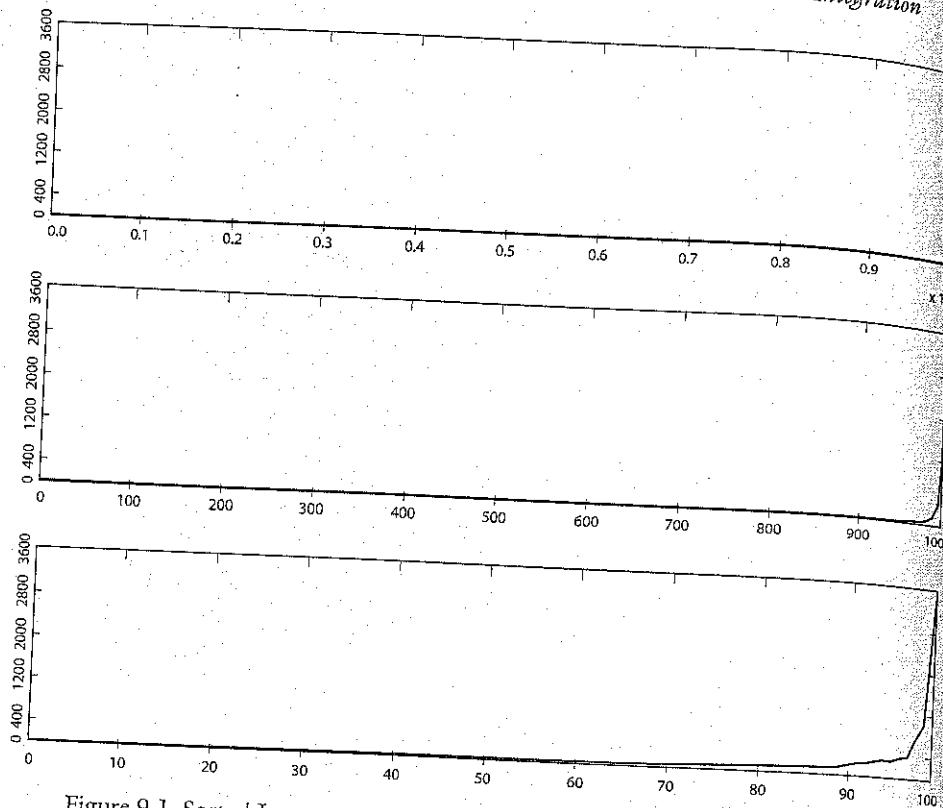


Figure 9.1 Sorted Importance Sampling Weights

- draw = mu + swish\*rndn(k, 1); ( $k$  is the dimensionality of  $\mu$ ).

### Exercise 9.5

Reapproximate the statistics  $g(\mu)$  of table 9.2 by scaling  $\Sigma_g$  by factors of 2, 5, and 10. What happens to RNE as you do so? What is the intuition behind this result?

Having illustrated the importance of working with tailored samplers, we note two key characteristics of effective samplers: their tails should be fat relative to the targeted integrand (to ensure adequate coverage of the sample space), and their center and rotation must not be too unlike the targeted integrand. In a variety of applications of Bayesian analyses to the class of state-space representations of interest here, multivariate- $t$  densities have proven to be effective along these dimensions (e.g., see DeJong, Ingram,

and Whiteman, 2000a, 2000b; DeJong and Ingram, 2001; and DeJong and Ripoll, 2004). Thus in this subsection we discuss the issue of tailoring with specific reference to multivariate- $t$  densities.

The parameters of the multivariate- $t$  are given by

$$\theta = (\gamma, V, \nu).$$

For a  $k$ -variate specification, the kernel is given by

$$g(x; \gamma, V, \nu) \propto [\nu + (x - \gamma)' V (x - \gamma)]^{-(k+\nu)/2}, \quad (9.49)$$

its mean and second-order moments are  $\gamma$  and  $(\frac{\nu}{\nu-2})V^{-1}$ ; and the integrating constant is given by

$$\frac{\nu^{\nu/2} \Gamma((\nu+k)/2) |V|^{1/2}}{\pi^{k/2} \Gamma(\nu/2)},$$

where  $\Gamma()$  denotes the gamma function.<sup>2</sup>

Regarding the first key characteristic, tail behavior, it is a simple matter to specify a multivariate- $t$  density with fatter tails than a Normal density. All that is needed is the assignment of small values to  $\nu$ . For example, figure 9.2 illustrates the relationship between a  $N(0, 1)$  density and a  $t(0, 1, 2)$  density.

Given  $(\gamma, V, \nu)$ , an algorithm for obtaining sample drawings  $\mu_i$  from the multivariate- $t$  is as follows. Begin by drawing a random variable  $s_i$  from a  $\chi^2(\nu)$  distribution (i.e., construct  $s_i$  as the sum of  $\nu$  squared  $N(0, 1)$  random variables). Use this to construct

$$\sigma_i = (s_i/\nu)^{-1/2},$$

which serves as a scaling factor for the second-moment matrix. Finally, obtain a  $k \times 1$  drawing of  $N(0, 1)$  random variables  $w_i$ , and convert this into a drawing of  $\mu_i$  as follows:

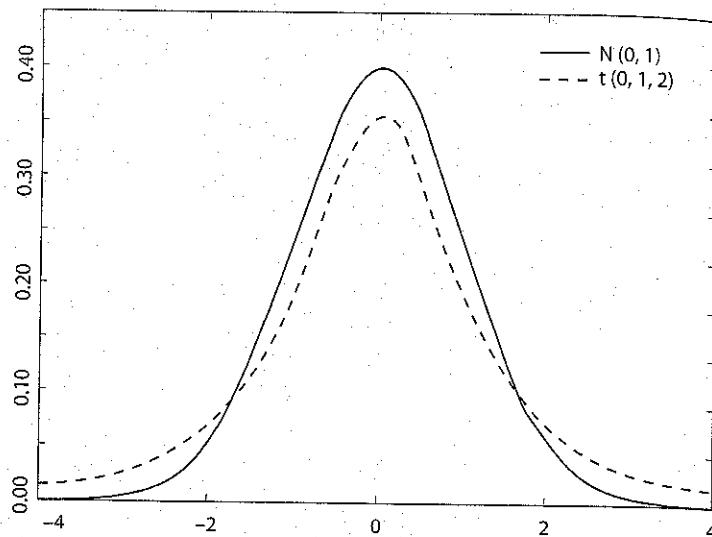
$$\mu_i = \gamma + \sigma_i V^{-1/2} w_i, \quad (9.50)$$

where  $V^{-1/2}$  is the  $k \times k$  lower triangular Cholesky decomposition of  $V^{-1}$ . As noted by Geweke (1988), it is beneficial to complement  $\mu_i$  with its antithetic counterpart<sup>3</sup>

$$\mu_i = \gamma - \sigma_i V^{-1/2} w_i.$$

<sup>2</sup> The GAUSS command `gamma(x)` facilitates the evaluation of  $\Gamma(x)$ .

<sup>3</sup> The GAUSS procedure `multit.prc` is available for use in generating sample drawings from multivariate- $t$  distributions.

Figure 9.2 Comparison of  $N(0, 1)$  and  $t(0, 1, 2)$  Densities**Exercise 9.6**

Reapproximate the statistics  $g(\mu)$  of table 9.2 using a multivariate-t distribution specified for  $g(\theta; a)$ , specifying  $\gamma$  as  $\mu_g$ , and  $V^{-1}$  as  $\Sigma_g$ . Study the behavior of RNE as a function of  $v$ , and compare the results you obtain with those of exercise 9.5.

Regarding center and rotation, these of course are manipulated via  $(\gamma, V)$ . One means of tailoring  $(\gamma, V)$  to mimic the targeted integrand is sequential. In the context of targeting a posterior distribution, a simple algorithm is as follows. Beginning with an initial specification  $(\gamma_0, V_0)$ , obtain an initial sequence of drawings and estimate the posterior mean and covariance matrix of  $\mu$ . Use the resulting estimates to produce a new specification  $(\gamma_1, V_1)$ , with  $\gamma_1$  specified as the estimated mean of  $\mu$ , and  $V_1$  as the inverse of the estimated covariance matrix. Repeat this process until posterior estimates converge. Along the way, in addition to histograms plotted for the weights and the RNE measure (9.45), the quality of the importance sampler can be monitored using

$$\varpi = \max_j \left[ \frac{\bar{w}(\mu_j, \Upsilon)}{\sum \bar{w}(\mu_i, \Upsilon)} \right], \quad (9.51)$$

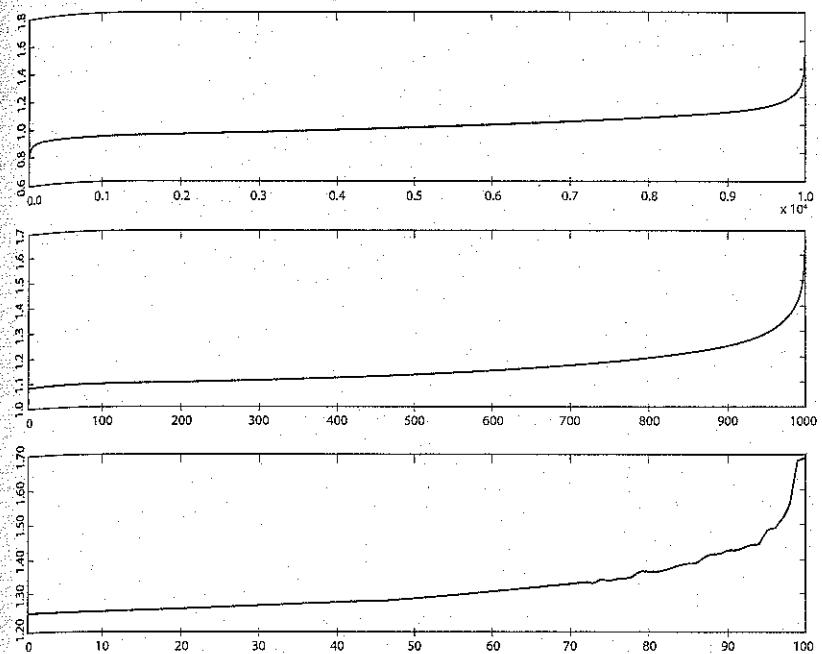


Figure 9.3 Sorted Importance Sampling Weights, Tailored Sampler

that is, the largest weight relative to the sum of all weights. Weights in the neighborhood of 2% are indicative of approximate convergence.

Applying this algorithm to the example presented above (reverting to the use of a Normal distribution for  $g(\theta; a)$ ), we obtain RNEs in the neighborhood of 90% within three iterations. Weights associated with the draws generated by the resulting sampler are illustrated in figure 9.3, with the panels once again illustrating the top 10,000, 1,000, and 100 draws.

While figure 9.3 illustrates a high payoff to the implementation of this sequential algorithm for this simple problem, success of this magnitude is by no means guaranteed in general. With this in mind, we now present a more systematic means of achieving tailoring: efficient importance sampling.

**9.4 Efficient Importance Sampling**

An alternative means of tailoring  $g(\theta; a)$  is provided by the efficient importance sampling (EIS) methodology of Richard and Zhang (1997, 2007).

The goal of EIS is to specify the parameters  $\alpha$  to minimize the NSE associated with the approximation of

$$G(\Upsilon) = \int_{\Theta} \varphi(\theta; \Upsilon) d\theta.$$

Construction is achieved as follows. One initially selects a parametric class of auxiliary density kernels  $\mathcal{K} = \{k(\theta; \alpha); \alpha \in A\}$  amenable to Monte Carlo simulation and with known integrating constant  $\chi(\alpha)$ . The relationship between kernel and density is given by

$$g(\theta; \alpha) = \frac{k(\theta; \alpha)}{\chi(\alpha)}, \quad \chi(\alpha) = \int k(\theta; \alpha) d\theta. \quad (9.52)$$

The optimal value  $\hat{\alpha}(\Upsilon)$  obtains as the solution of the following least-squares problem:

$$(\hat{\alpha}(\Upsilon), \hat{c}(\Upsilon)) = \arg \min_{\alpha \in A, c \in \mathbb{R}} \sum_{i=1}^R [\ln \varphi(\theta_i; \Upsilon) - c - \ln k(\theta_i; \alpha)]^2. \quad (9.53)$$

Here,  $\{\theta_i\}_{i=1}^R$  denotes *iid* draws from  $g(\theta; \hat{\alpha}(\Upsilon))$ , with  $R$  typically set at 3 to 5 times the dimension of  $\alpha$  (and thus typically smaller than the number  $N$  of draws subsequently used for Monte Carlo approximation). The fact that these draws are conditional on  $\hat{\alpha}(\Upsilon)$  implies that the latter obtains as the fixed point solution of a sequence  $\{\hat{\alpha}_v(\Upsilon)\}_{v=1,2,\dots}$ , where  $\hat{\alpha}_v(\Upsilon)$  solves (9.53) under draws from  $g(\theta; \hat{\alpha}_{v-1}(\Upsilon))$ . In turn,  $c$  is a normalizing constant that accounts for factors in  $\varphi(\cdot)$  and  $k(\cdot)$  that do not depend upon  $\theta$ . The sequence can be initialized, for example, by centering  $g(\theta; \hat{\alpha}_0(\Upsilon))$  at the mode of  $\ln \varphi(\theta; \Upsilon)$ , and specifying sufficient diffusion to ensure adequate coverage of the sample space.

To achieve smooth fixed-point convergence, all draws under  $\hat{\alpha}_v(\theta)$  are obtained by transformation of a single set  $\{u_i\}_{i=1}^R$  of common random numbers (CRNs), where  $u_i$  denotes a draw from a canonical distribution—that is, one that does not depend on  $\alpha$  (typically  $U(0, 1)$  or  $N(0, 1)$ ). Under these conditions convergence is fast for well-behaved problems (typically less than five iterations). Most importantly, these iterations are the main driver for achieving optimal efficiency (within the class of densities  $\mathcal{K}$ ), as they progressively reposition  $k(\theta; \alpha)$  on the region of importance of  $\varphi(\theta; \Upsilon)$ . Fixed-point convergence can be assessed by monitoring  $\hat{\alpha}_v(\theta)$  across successive iterations  $v$ , and a stopping rule can be established using a relative-change threshold. Although there is no guarantee nor formal proof that the sequence  $\{\hat{\alpha}_v(\theta)\}$  converges for every possible pair  $(\varphi(\cdot), k(\cdot))$ , convergence typically obtains absent a complete mismatch between the target  $\varphi(\cdot)$  and the kernel  $k(\cdot)$ .

#### 9.4 Efficient Importance Sampling

unimodal), in which case failure to converge serves as a signal that the class  $\mathcal{K}$  needs to be adjusted.

Note that fixed-point convergence of  $\hat{\alpha}_v(\theta)$  is not necessary for guaranteeing convergence of the estimator

$$\overline{G(\Upsilon)}_{N,g} = \frac{1}{N} \sum_{i=1}^N \frac{\varphi(\theta_i; \Upsilon)}{g(\theta_i; \hat{\alpha}_v(\theta))}$$

to  $G(\Upsilon)$ . Instead, convergence continues to be guaranteed as described above: so long as the support of  $g(\theta; \hat{\alpha}_v(\theta))$  includes that of  $\varphi(\theta; \Upsilon)$ ,  $\overline{G(\Upsilon)}_{N,g}$  converges almost surely to  $G(\Upsilon)$  so long as  $G(\Upsilon)$  exists and is finite. Regarding the numerical efficiency of  $g(\theta; \hat{\alpha}_v(\theta))$ , this can be assessed using diagnostic measures such as the  $R^2$  statistic associated with (9.53), in addition to the diagnostic statistics described above.

In what follows, we characterize in general the construction of EIS samplers belonging to the exponential family of distributions, including the specialized characterization of applications involving Normal distributions (a particular member of the exponential family). In so doing, for ease of notation we denote the targeted integrand as  $\varphi(\theta; \Upsilon) \equiv \varphi(\theta)$ , and take  $\theta$  as  $k$ -dimensional, with individual elements  $(x_1, x_2, \dots, x_k)$ .

The exponential family of distributions is defined as having kernels that satisfy

$$\ln k(\theta; \alpha) = b(\theta) + \alpha' \cdot T(\theta), \quad (9.54)$$

where  $T(\theta)$  denotes a fixed-dimensional vector of sufficient statistics (for details, see, e.g., Lehmann, 1986). In the case of Normal distributions,  $T(\theta)$  is  $(k + \frac{k(k+1)}{2})$ -dimensional, and includes the individual elements of  $\theta$  and all unique cross-product terms:

$$x_1, \dots, x_k, x_1 x_2, x_1 x_3, \dots, x_{k-1} x_k, x_1^2, x_2^2, \dots, x_k^2$$

For distributions of this class, the minimization problem (9.53) specializes to an ordinary-least-squares problem involving the regression of  $\{\ln \varphi(\theta_i)\}_{i=1}^R$  on  $\{1, T(\theta_i)\}_{i=1}^R$ , the parameters of which map into  $(\hat{\alpha}_v(\theta), \hat{c}_v(\theta))$ . In the case of Normal distributions, the specific form of the problem is as follows:

$$\min_{\beta} (y - X\beta)^2,$$

with

$$y = \begin{pmatrix} \ln \varphi(\theta_1) \\ \vdots \\ \ln \varphi(\theta_R) \end{pmatrix}, \quad X = \begin{pmatrix} \kappa_1 \\ \vdots \\ \kappa_R \end{pmatrix},$$

$$\kappa_i = \prod_{j=1}^k \theta_j^{a_{ij}} vech(\theta_i \cdot \theta_i' Y_i).$$

where  $\text{vech}(\theta_i \cdot \theta'_i)'$  denotes the unique cross-product terms of  $\theta_i \cdot \theta'_i$ :

$$\text{vech}(\theta_i \cdot \theta'_i)' = (x_1 x_2 \ x_1 x_3 \ \cdots \ x_{k-1} x_k \ x_1^2 \ x_2^2 \ \cdots \ x_k^2).^4$$

The solution to the problem is simply

$$\hat{\beta} = (X'X)^{-1}(X'y),$$

where  $\hat{\beta}$  is  $(1 + k + \frac{k(k+1)}{2}) \times 1$ . Defining  $K = (1 + k + \frac{k(k+1)}{2})$  and letting  $(\hat{\mu}, \hat{\Sigma})$  denote the mean and covariance matrix of  $g(\theta; \hat{\alpha}(\theta))$ , the mapping from  $\hat{\beta}$  to  $(\hat{\mu}, \hat{\Sigma})$  begins by mapping the  $k+2$  through  $K$  elements of  $\hat{\beta}$  into a symmetric matrix  $\tilde{H}$ , with the  $j^{\text{th}}$  diagonal element corresponding to the coefficient associated with the squared value of the  $j^{\text{th}}$  element of  $\theta$ , and the  $(j, k)^{\text{th}}$  element corresponding to the coefficient associated with the product of the  $j^{\text{th}}$  and  $k^{\text{th}}$  elements of  $\theta$ .<sup>5</sup> Next, the elements of  $\tilde{H}$  are multiplied by  $-1$  to yield  $\tilde{H}$ , and then the diagonal elements of  $\tilde{H}$  are multiplied by  $2$ . The resulting matrix is the precision matrix  $\hat{H}$ ; inverting this matrix yields the desired approximation  $\hat{\Sigma}$ . Finally, post-multiplying  $\hat{\Sigma}$  by the  $2^{\text{nd}}$  through  $k+1$  coefficients of  $\hat{\beta}$  yields the desired approximation  $\hat{\mu}$ .

As a concrete example, consider a 3-dimensional problem where  $\theta = (x_1 \ x_2 \ x_3)'$ . The regression specification in this case is given by

$$\begin{aligned} \ln \varphi(\theta) = & \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 \\ & + \beta_4(x_1 x_2) + \beta_5(x_1 x_3) + \beta_6(x_2 x_3) \\ & + \beta_7(x_1^2) + \beta_8(x_2^2) + \beta_9(x_3^2). \end{aligned}$$

Having obtained least-squares estimates, the unique elements of the approximated precision matrix  $\hat{H}$  are given by

$$\begin{aligned} \hat{h}_{11} &= -2\hat{\beta}_7; \hat{h}_{22} = -2\hat{\beta}_8; \hat{h}_{33} = -2\hat{\beta}_9 \\ \hat{h}_{12} &= -\hat{\beta}_4; \hat{h}_{13} = -\hat{\beta}_5; \hat{h}_{23} = -\hat{\beta}_6. \end{aligned}$$

Then the covariance matrix is simply  $\hat{\Sigma} = \hat{H}^{-1}$ , and  $\hat{\mu}$  is given by

$$\hat{\mu} = \hat{\Sigma} \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix}.$$

When  $\theta_t$  is univariate, the least-squares problem reduces to the regression

$$\ln \varphi(\theta_t) = \beta_0 + \beta_1 \theta_t + \beta_2 \theta_t^2,$$

<sup>4</sup> In GAUSS, the command `vech(theta*theta')` returns precisely these elements.

<sup>5</sup> The GAUSS command `xpdn` accomplishes this mapping.

and  $(\hat{\mu}, \hat{\sigma}^2)$  are obtained using

$$\begin{aligned} \hat{\sigma}^2 &= -\frac{1}{2\hat{\beta}_2}, \\ \hat{\mu} &= -\frac{1}{2}\frac{\hat{\beta}_1}{\hat{\beta}_2}. \end{aligned}$$

### Exercise 9.7

Reapproximate the statistics  $g(\mu)$  of table 9.2 using a Normal importance sampler tailored using EIS. As an initial sampler, use the naive importance sampler used to obtain the results reported in table 9.2, with the covariance matrix scaled by a factor of 5 to ensure sufficient coverage of the parameter space. You should obtain convergence to the direct Monte Carlo sampler in one step.

## 9.5 Markov Chain Monte Carlo Integration

As its name suggests, the approximation of integrals of the form

$$E[g(\mu)] = \int g(\mu)P(\mu|X)d\mu$$

via a Markov Chain Monte Carlo (MCMC) method involves the construction of a Markov chain in  $\mu$  whose distribution converges to the posterior of interest  $P(\mu|X)$ . (For an introduction to Markov chains, see Hamilton, 1994; or Ljungqvist and Sargent, 2004; and for a detailed probabilistic characterization, see Shiryaev, 1995.) Applications of MCMC methods to Bayesian analyses of DSGE models have become widespread; prominent examples include Otrok (2001), Fernandez-Villaverde and Rubio-Ramirez (2004a), and Smets and Wouters (2003, 2007); for a survey, see An and Schorfheide (2007).

In short, recall that a stochastic process  $\{x_i\}$  has a Markov property if for all  $i$ ,

$$\Pr(x_{i+1}|x_i, x_{i-1}, x_{i-2}, \dots) = \Pr(x_{i+1}|x_i). \quad (9.55)$$

In a time-series context  $i$  corresponds to a time index; here  $i$  corresponds to an index of Monte Carlo replications. To characterize the stochastic behavior of  $x_i$  via a Markov chain, we suppose that  $x_i$  can take on one of  $n$  possible values that collectively define the state space over  $x$ . The movement of  $x_i$  over  $i$  is characterized via  $P$ , an  $n \times n$  transition matrix.

The  $(q, r)^{\text{th}}$  element of  $P$  is denoted as  $p_{qr}$ , and indicates the probability that  $x_i$  will transition to state  $r$  in replication  $i + 1$  given its current state  $q$ . Thus, the  $q^{\text{th}}$  row of  $P$  indicates the entire conditional probability distribution for  $x_{i+1}$  given the current state  $q$  in replication  $i$ .

The original algorithm used to accomplish numerical integration via Markov chain approximation was developed by Metropolis et al. (1953), and was subsequently refined by Hastings (1970). In these algorithms, sample drawings of  $\mu$  represent links in the chain; corresponding calculations of  $g(\mu)$  for each link serve as inputs in the approximation of  $E[g(\mu)]$ . Many extensions of these algorithms have been developed in subsequent work. Here we discuss three leading approaches to simulating Markov chains: the Gibbs sampler, and the independence and random walk variants of the Metropolis-Hastings Algorithm. Further details are provided by, for example, Geweke (1999a, 2005) and Robert and Casella (1999).

### 9.5.1 The Gibbs Sampler

The Gibbs sampler provides an alternative approach to that outlined in section 9.2 for obtaining drawings of  $\mu$  from the exact distribution  $P(\mu|X)$ . Although once again this will rarely be possible in applications of interest here, it is instructive to consider this case before moving on to more complicated scenarios.

Rather than obtaining drawings of  $\mu$  from the full distribution  $P(\mu|X)$ , with a Gibbs sampler one partitions  $\mu$  into  $m \leq k$  blocks,

$$\mu = (\mu^1 | \mu^2 | \dots | \mu^m),$$

and obtains drawings of the components of each block  $\mu^j$  from the conditional distribution  $P(\mu|X, \mu^{-j})$ ,  $j = 1 \dots m$ , where  $\mu^{-j}$  denotes the removal of  $\mu^j$  from  $\mu$ . This is attractive in scenarios under which it is easier to work with  $P(\mu|X, \mu^{-j})$  rather than  $P(\mu|X)$ . For example, a natural partition arises in working with the Normal/Inverted-Wishart distribution, described in chapter 14. Note that when  $m = 1$ , we are back to the case in which  $\mu$  is drawn directly from  $P(\mu|X)$ .

Proceeding under the case in which  $m = 2$ , the Gibbs sampler is initiated with a specification of initial parameter values

$$\mu_0 = (\mu_0^1 | \mu_0^2).$$

Thereafter, subsequent drawings are obtained from the conditional posteriors

$$\mu_{i+1}^1 \sim P(\mu^1|X, \mu_i^2) \quad (9.56)$$

$$\mu_{i+1}^2 \sim P(\mu^2|X, \mu_i^1). \quad (9.57)$$

This process generates a Markov chain  $\{\mu_i\}_{i=1}^N$  whose ergodic distribution is  $P(\mu|X)$ . To eliminate the potential influence of  $\mu_0$ , it is typical to discard the first  $S$  draws from the simulated sample, serving as another example of burn-in. Hereafter,  $N$  represents the number of post-burn-in draws; a common choice for  $S$  is 10% of  $N$ .

As with direct sampling, the approximation of  $E[g(\mu)]$  using

$$\bar{g}_N = \left( \frac{1}{N} \right) \sum_{i=1}^N g(\mu_i)$$

yields convergence at the rate  $1/\sqrt{N}$  (e.g., see Geweke, 1992). However, because the simulated drawings  $\mu_i$  are not serially independent, associated numerical standard errors will be higher than under direct sampling (wherein serial independence holds). Following Bauwens, Lubrano, and Richard (1999), letting  $\gamma_0$  denote the variance of  $g(\mu)$ , and  $\gamma_l$  the  $l^{\text{th}}$ -order autocovariance of  $g(\mu)$ , the numerical standard error of  $\bar{g}_N$  can be approximated using

$$\text{s.e.}(\bar{g}_N)_G = \left[ \frac{1}{N} \left( \bar{\gamma}_0 + 2 \sum_{l=1}^{N-1} \bar{\gamma}_l \frac{N-l}{N} \right) \right]^{1/2}, \quad (9.58)$$

where  $\bar{\gamma}_l$  denotes the numerical estimate of the  $l^{\text{th}}$ -order autocovariance of  $g(\mu)$  obtained from the simulated  $\mu_i$ 's.

Geweke (1992) suggests a convenient means of judging whether estimates obtained under this procedure have converged. Partition the simulated drawings of  $\mu_i$  into three subsets,  $I$ ,  $II$ , and  $III$  (e.g., into equal  $1/3s$ ), and let  $[\bar{g}_N]^J$  and  $[\text{s.e.}(\bar{g}_N)_G]^J$  denote the sample average and associated numerical standard error calculated from subset  $J$ . Then under a central limit theorem,

$$CD = \frac{[\bar{g}_N]^I - [\bar{g}_N]^{III}}{[\text{s.e.}(\bar{g}_N)_G]^I + [\text{s.e.}(\bar{g}_N)_G]^{III}} \quad (9.59)$$

will be distributed as  $N(0, 1)$  given the attainment of convergence. A test of the “null hypothesis” of convergence can be conducted by comparing this statistic with critical values obtained from the standard Normal distribution. The reason for separating the subsamples  $I$  and  $III$  is to help ensure their independence.

An additional diagnostic statistic, suggested by Yu and Mykland (1994), is based on a CUSUM statistic. Let  $\bar{g}_N$  and  $\bar{\sigma}_N(g(\mu))$  denote the sample mean and standard deviation of  $g(\mu)$  calculated from a chain of

length  $N$ . Then continuing the chain over  $t$  additional draws, the  $CS_t$  statistic is

$$CS_t = \frac{\bar{g}_t - \bar{g}_N}{\sigma_N(g(\mu))}. \quad (9.60)$$

Thus  $CS_t$  measures the deviation of  $\bar{g}_t$  from  $\bar{g}_N$  as a percentage of the estimated standard deviation of  $g(\mu)$ . Given convergence,  $CS_t$  should approach 0 as  $t$  increases; Bauwens and Lubrano (1998) suggest 5% as a criterion for judging convergence.

### 9.5.2 Metropolis-Hastings Algorithms

We now turn to the case in which  $P(\mu|X, \mu^{-j})$  is unavailable as a sampling distribution for  $\mu^j$ . In this case, it is still possible to generate a Markov chain  $\{\mu_i\}$  with ergodic distribution  $P(\mu|X)$ . There are many alternative possibilities for doing so; an overview is provided by Chib and Greenberg (1995), and a procedure specialized to the application of DSGE models is presented by Chib and Ramamurthy (2009). Here we focus on two important variants of the Metropolis-Hastings Algorithm: the independence chain and random-walk algorithms.

To begin, we must specialize the notation that characterizes the stand-in density used to obtain drawings  $\{\mu_i\}$ . In particular, we must distinguish between the parameters that determine its center and spread/rotation. We do so using  $\iota(\mu; m, H)$ , where  $m$  characterizes centering and  $H$  spread/rotation. For example, for  $\iota()$  distributed as  $N(m, \Sigma)$ ,  $H$  denotes the precision matrix  $H = \Sigma^{-1}$ ; and for  $\iota()$  distributed as multivariate- $t$  ( $\gamma, V, v$ ),  $H$  encompasses  $(V, v)$ .

Beginning with a characterization of the general Metropolis-Hastings Algorithm, we take the first ( $i-1$ ) Monte Carlo drawings as given, and consider the attainment of  $\mu_i$ . (Once again, the process is initialized with a starting specification  $\mu_0$ , the influence of which is eliminated via a burn-in phase.) The attainment of  $\mu_i$  is accomplished through the use of the stand-in density  $\iota(\mu; \mu_{i-1}, H)$ . Just as in the case of importance sampling,  $\iota(\mu; \mu_{i-1}, H)$  is ideally designed to have fat tails relative to the posterior, and center and rotation not unlike the posterior. Its implementation as a sampling distribution proceeds as follows.

Let  $\mu_i^*$  denote a drawing obtained from  $\iota(\mu; \mu_{i-1}, H)$  that serves as a candidate to become the next successful drawing  $\mu_i$ . Under the Metropolis-Hastings Algorithm, this will be the case according to the probability

$$q(\mu_i^* | \mu_{i-1}) = \min \left[ 1, \frac{P(\mu_i^* | X)}{P(\mu_{i-1} | X)} \frac{\iota(\mu_{i-1}; \mu_{i-1}, H)}{\iota(\mu_i^*; \mu_{i-1}, H)} \right]. \quad (9.61)$$

In practice, the outcome of this random event can be determined by comparing  $q(\mu_i^* | \mu_{i-1})$  with a sample drawing  $\varsigma$  obtained from a uniform distribution over  $[0, 1]$ :  $\mu_i^* = \mu_i$  if  $q(\mu_i^* | \mu_{i-1}) > \varsigma$ , else  $\mu_i^*$  is discarded and replaced with a new candidate drawing. Once a candidate drawing is accepted, the stand-in density is recentered at  $\iota(\mu; \mu_i, H)$ , and iteration continues.

Note that for the case in which

$$\iota(\mu; m, H) \equiv P(\mu | X),$$

the probability

$$q(\mu_i^* | \mu_{i-1}) = 1 \quad \forall \mu_i^*,$$

and we have reverted to the direct-sampling case.

Finally, letting  $\{\mu_i\}$  denote the sequence of accepted drawings,  $E[g(\mu)]$  is once again approximated via

$$\bar{g}_N = \frac{1}{N} \sum_1^N g(\mu_i),$$

with associated numerical standard error estimated as in (9.58).

Note the parallels between the role played by  $q(\mu_i^* | \mu_{i-1})$  in this algorithm, and the analogous weight function

$$\bar{w}(\mu, \Upsilon) = \frac{P(\mu | X)}{g(\mu | \alpha)}$$

of the importance-sampling algorithm. Both serve to eliminate the influence of their respective proposal densities in estimating  $E[g(\mu)]$ , and reserve this role to the posterior distribution. In this case,  $q(\mu_i^* | \mu_{i-1})$  will be relatively low when the probability assigned to  $\mu_i^*$  by  $\iota(\mu; \mu_{i-1}, H)$  is relatively high; and  $q(\mu_i^* | \mu_{i-1})$  will be relatively high when the probability assigned to  $\mu_i^*$  by  $P(\mu | X)$  is relatively high.

As with the Gibbs sampler, numerical standard errors, the CD statistic (9.59), and the  $CS_t$  statistic (9.60) can be used to check for the convergence of  $\bar{g}_N$ . In addition, a symptom of an ill-fitted specification of  $\iota(\mu; m, H)$  in this case is a tendency to experience high rejection frequencies.

Turning to the independence chain variant of this algorithm, the stand-in density  $\iota(\mu; \mu_{i-1}, H)$  specializes to

$$\iota(\mu; \mu_{i-1}, H) = \iota(\mu; m, H).$$

That is, the stand-in density is independent across Monte Carlo replications, precisely as with the case of the importance-sampling counterpart  $g(\mu; \alpha)$ .

Under the random-walk variant, candidate drawings  $\mu_i^*$  are obtained according to

$$\mu_i^* = \mu_{i-1} + \varepsilon_i, \quad (9.62)$$

where the density specified over  $\varepsilon$  is symmetric about 0, and its spread/rotation is determined by  $H$ . In this case, the stand-in density specializes to

$$\iota(\mu; \mu_{i-1}, H) = \iota(\mu - \mu_{i-1}; 0, H),$$

so that the center of  $\iota(\mu; \mu_{i-1}, H)$  evolves over Monte Carlo replications following a random walk.

The approaches to tailoring  $g(\theta; a)$  described above can be adopted to tailoring  $\iota(\mu; \mu_{i-1}, H)$  under either variant of the algorithm. For example, under the independence chain algorithm,  $(m, H)$  can be constructed to mimic posterior estimates of the mean and covariance of  $\mu$  obtained given an initial specification  $(m_0, H_0)$ , and updated sequentially. And a good initial specification can be fashioned after maximized values of the mean and covariance of  $\mu$  obtained via the application of a hill-climbing procedure to  $P(\mu|X)$ . (Hill-climbing procedures are described in detail in chapter 13.) Likewise, under the random-walk algorithm the rotation of  $\iota(\mu - \mu_{i-1}; H)$  can be constructed to mimic estimates of the covariance of  $\mu$  under  $P(\mu|X)$ . Finally, tailoring can be accomplished via efficient importance sampling; for details, see Liesenfeld and Richard (2010).

### Exercise 9.8

*Reapproximate the statistics  $g(\mu)$  of table 9.2 using the random-walk variant of the Metropolis-Hastings Algorithm. Do so by specifying  $\iota(\mu - \mu_{i-1}; 0, H)$  as Normal, and parameterize  $(\mu_0, H)$  to mimic the specification of the naive importance sampler used to produce the results reported in table 9.2.*

We conclude this chapter by noting that of the four contexts introduced in section 1 as motivating the implementation of Monte Carlo integration procedures, we have not characterized in this chapter that of achieving likelihood evaluation and filtering for state-space representations. Due to the appearance of analytically intractable densities in the targeted integrands that are encountered in this context, these objectives require special care. Details are provided in the following chapter.

## Chapter 10

### Likelihood Evaluation and Filtering in State-Space Representations Using Sequential Monte Carlo Methods

#### 10.1 Background

Chapter 8 demonstrated the characterization of DSGE models as state-space representations. This characterization facilitates pursuit of the dual objectives of likelihood evaluation and filtering (i.e., inferring the time-series behavior of the model's unobservable state variables, conditional on the observables). Both objectives entail the calculation of integrals over the unobservable state variables.

As we have seen, when models are linear and stochastic processes are Normally distributed, required integrals can be calculated analytically via the Kalman filter. Departures entail integrals that must be approximated numerically. Here we characterize methods for approximating such integrals; the methods fall under the general heading of sequential Monte Carlo (SMC) methods.

Dating at least to the sequential importance sampling (SIS) algorithm of Handschin and Mayne (1969) and Handschin (1970), SMC methods have long served as workhorses for achieving likelihood evaluation and filtering. Moreover, extensive efforts have been made to develop refinements of the baseline SIS algorithm with the goal of enhancing accuracy and numerical efficiency in the face of challenging scenarios that give rise to problems known as degeneracy and sample impoverishment (for extensive surveys, see Ristic, Arulampalan, and Gordon, 2004; and Cappé, Godsill, and Moulines, 2007). In the terminology of Pitt and Shephard (1999), the goal of these efforts is to achieve adaption.

The application of SMC methods to the evaluation of DSGE models has been demonstrated by Fernandez-Villaverde and Rubio-Ramirez (2004a, 2005, 2007). The specific filtering method they implemented was developed by Gordon, Salmond, and Smith (1993). Like the SIS algorithm, this filter employs discrete fixed-support approximations to unknown densities that appear in the predictive and updating stages of the filtering process. The discrete points that collectively provide density

approximations are known as particles; the approach is known as the bootstrap particle filter, and is a leading example of a sampling importance resampling (SIR) algorithm.

As we shall see, SIS and SIR algorithms are intuitive and straightforward to implement; however, they suffer two important shortcomings. First, because the density approximations they provide are discrete, associated likelihood approximations can feature spurious discontinuities (with respect to parameters), rendering as problematic the application of likelihood maximization procedures. Second, the supports upon which approximations are based are not adapted: period- $t$  approximations are based on supports that incorporate information conveyed by values of the observable variables available in period  $t - 1$ , but not period  $t$ . As characterized by Pitt and Shephard (1999), this renders the approximations as “blind.” This problem gives rise to numerical inefficiencies that can be acute when observable variables are highly informative with regard to state variables, particularly given the presence of outliers.

Numerous extensions of SIS and SIR algorithms have been proposed in attempts to address these problems. For examples, see Pitt and Shephard (1999); the collection of papers in Doucet, de Freitas, and Gordon (2001); Pitt (2002); and Ristic, Arulampalam, and Gordon (2004). While these efforts have produced algorithms that have proven to be effective in many applications, the use of discrete approximations to filtering densities entails a significant limitation: the support of the period- $(t - 1)$  approximations carried over for the construction of period- $t$  approximations can no longer be changed in period  $t$ . Only the weights attached to the period- $(t - 1)$  discrete elements of the state space (i.e., particles) can be adjusted in period  $t$ . To be sure, there exists a wide range of algorithms designed to optimize (adapt) these weights in light of new information available at time  $t$ . Nevertheless, optimality is restricted as being conditional on the individual particles established in period  $(t - 1)$ . That is, discrete approximations cannot achieve unconditional optimality, or in other words, full adaption.

Motivated by these shortcomings, DeJong et al. (2010) developed an approach to filtering based on the concept of efficient importance sampling (EIS), characterized in section 4 of chapter 9. Implementation entails the abandonment of discrete approximations of filtering densities in favor of continuous approximations constructed sequentially as optimal importance sampling densities based upon period- $t$  information, and absent fixed- and prespecified-support constraints. That is, the EIS filter is designed to achieve unconditional adaption. As introduced in chapter 9, optimality is associated with the minimization of the numerical standard error of the weights associated with the chosen importance sampler; applied to DSGE models, its pursuit takes the form of a sequence of least-squares regressions.

### 10.1 Background

In order to characterize discrete approximations of unknown densities, it is useful to introduce Dirac measures, and then develop some specialized notation. In general, denoting  $s$  as an element of the set  $\Sigma$ , and  $S$  as a subset of  $\Sigma$ , a Dirac measure is a probability measure defined as

$$\delta_s(S) = \begin{cases} 1, & s \in S \\ 0, & s \notin S \end{cases}$$

This measure specializes to the case in which  $S$  is the single element  $s$ :

$$\delta_s(s) = \begin{cases} 1, & s \in S \\ 0, & s \notin S \end{cases}$$

This is the measure of interest here;  $\delta_s(s)$  indicates the outcome  $s$  in the set  $\Sigma$ .

Regarding notation, let  $\{s^i\}_{i=1}^N$  denote a sequence of random draws obtained from an importance sampler  $g(s; \alpha)$  introduced to approximate  $f(s|\Upsilon)$ . The individual draws  $s^i$  are known as particles, and  $\{s^i\}_{i=1}^N$  is known as a particle swarm. The swarm can be used to approximate  $f(s|\Upsilon)$  as

$$\hat{f}(s|\Upsilon) = \sum_{i=1}^N \omega^i \delta_{s^i}(s), \quad (10.1)$$

where the weights  $\omega^i$  are defined as

$$\omega^i \propto \frac{f(s^i|\Upsilon)}{g(s^i; \alpha)},$$

and are normalized as

$$\sum_{i=1}^N \omega^i = 1.$$

Unadapted filters, as well as those designed to attain (approximate) conditional optimality (or in other words, partial adaption) implement approximations of unknown densities in the form of (10.1).

Before proceeding with the characterization of unadapted, conditionally optimal, and unconditionally optimal filters, we briefly resketch the sequential pursuit of likelihood evaluation and filtering for state-space representations. Recall that these representations are characterized in the form of state-transition and measurement equations

$$s_t = \gamma(s_{t-1}, \Upsilon_{t-1}, u_t), \quad (10.2)$$

$$y_t = \delta(s_t, \Upsilon_{t-1}, u_t), \quad (10.3)$$

with corresponding densities given respectively by

$$\begin{aligned} f(s_t | s_{t-1}, \mathcal{Y}_{t-1}), \\ f(y_t | s_t, \mathcal{Y}_{t-1}). \end{aligned}$$

Characterization is completed with the specification of the initial (or unconditional) distribution of the state

$$f(s_0) \equiv f(s_0 | \mathcal{Y}_0).$$

Taking as input the period- $(t-1)$  filtering density  $f(s_{t-1} | \mathcal{Y}_{t-1})$  (initialized as  $f(s_0)$ ), likelihood evaluation and filtering in period  $t$  begins with the calculation of the predictive density  $f(s_t | \mathcal{Y}_{t-1})$ , given by

$$f(s_t | \mathcal{Y}_{t-1}) = \int f(s_t | s_{t-1}, \mathcal{Y}_{t-1}) f(s_{t-1} | \mathcal{Y}_{t-1}) ds_{t-1}. \quad (10.4)$$

This serves as an input to the period- $t$  likelihood function, given by

$$f(y_t | \mathcal{Y}_{t-1}) = \int f(y_t | s_t, \mathcal{Y}_{t-1}) f(s_t | \mathcal{Y}_{t-1}) ds_t. \quad (10.5)$$

Finally, the period- $t$  filtering density is constructed via an application of Bayes' Theorem:

$$f(s_t | \mathcal{Y}_t) = \frac{f(y_t, s_t | \mathcal{Y}_{t-1})}{f(y_t | \mathcal{Y}_{t-1})} = \frac{f(y_t | s_t, \mathcal{Y}_{t-1}) f(s_t | \mathcal{Y}_{t-1})}{f(y_t | \mathcal{Y}_{t-1})}. \quad (10.6)$$

Having constructed  $f(s_t | \mathcal{Y}_t)$ , filtering is accomplished via the calculation

$$E_t(b(s_t) | \mathcal{Y}_t) = \int b(s_t) f(s_t | \mathcal{Y}_t) ds_t \quad (10.7)$$

for functions of interest  $b(s_t)$ . Finally,  $f(s_t | \mathcal{Y}_t)$  is passed as an input for accomplishing period- $(t+1)$  calculations. Having moved sequentially through the entire sample, the likelihood function is given by

$$L(\mathcal{Y}_T | \mu) = \prod_{t=1}^T f(y_t | \mathcal{Y}_{t-1}). \quad (10.8)$$

## 10.2 Unadapted Filters

The SIS filtering algorithm is initialized by obtaining a swarm of particles  $\{s_0^i\}_{i=1}^N$  directly from the unconditional density  $f(s_0)$ . Each element  $\omega_0^i$  in

the corresponding sequence of weights  $\{\omega_0^i\}_{i=1}^N$  equals  $1/N$ , since  $f(s_0)$  is both the importance sampling density and the density targeted for approximation. Upon initialization, the period- $t$  sequence of the filtering process is executed in four stages: propagation, likelihood integration, filtering, and continuation. The stages are characterized as follows.

**Propagation:** Inherit  $\{\omega_{t-1}^i, s_{t-1}^i\}_{i=1}^N$  from period  $(t-1)$ . Then for each particle  $s_{t-1}^i$ , draw a particle  $s_t^i$  from the transition density  $f(s_t | s_{t-1}^i, \mathcal{Y}_{t-1})$ . The corresponding approximation of the predictive density is given by the mixture of Dirac measures

$$\widehat{f}(s_t | \mathcal{Y}_{t-1}) = \sum_{i=1}^N \omega_{t-1}^i \cdot \delta_{s_t^i}(s_t). \quad (10.9)$$

**Likelihood integral:** Substitution of  $\widehat{f}(s_t | \mathcal{Y}_{t-1})$  for  $f(s_t | \mathcal{Y}_{t-1})$  in (10.5), followed by analytical integration in  $s_t$ , yields the following approximation for the likelihood integral:

$$\widehat{f}(y_t | \mathcal{Y}_{t-1}) = \sum_{i=1}^N \omega_{t-1}^i f(y_t | s_t^i, \mathcal{Y}_{t-1}). \quad (10.10)$$

**Filtering:** Equation (10.6) amounts to a direct application of Bayes' Theorem to the particles  $\{s_{t-1}^i\}_{i=1}^N$ , resulting in the following discrete approximation for the period- $t$  filtering density:

$$\widehat{f}(s_t | \mathcal{Y}_t) = \sum_{i=1}^N \omega_t^i \delta_{s_t^i}(s_t), \quad \omega_t^i = \frac{\omega_{t-1}^i \cdot f(y_t | s_t^i, \mathcal{Y}_{t-1})}{\sum_{j=1}^N \omega_{t-1}^j \cdot f(y_t | s_t^j, \mathcal{Y}_{t-1})}. \quad (10.11)$$

The filtered value in (10.7) is approximated by  $\widehat{b}_t = \sum_{i=1}^N \omega_t^i b(s_t^i)$ .

**Continuation:** Pass the weights and mass points  $\{\omega_t^i, s_t^i\}_{i=1}^N$  to the period- $(t+1)$  propagation step and proceed through period  $T$ .

SIR algorithms include a fifth stage to this process known as resampling. This is performed following the filtering state, and is characterized as follows.

**Resampling:** Draw from  $\{s_t^i\}_{i=1}^N$  with replacement, with the probability of selecting  $s_t^i$  given by  $\omega_t^i$ . Denote the resulting swarm as  $\{\omega_t^i, s_t^i\}_{i=1}^N$ , where following resampling  $\omega_t^i = 1/N \forall i$ .

Motivation for the resampling stage comes from Bobrovsky and Zakai (1975), who showed that as the sequential approximation of filtering densities advances through time, the variance of importance sampling weights can only increase. This problem has become known as degeneracy. Resampling was proposed originally by Rubin (1987) in the context of

achieving importance sampling; inclusion of the resampling step in the context of achieving filtering was first proposed by Gordon, Salmond, and Smith (1993). The specific resampling step described above is exactly theirs, and its inclusion yields the bootstrap particle filter. Variations on their resampling algorithm include those proposed by Kitagawa (1996), Del Moral (1996), Blake and Isard (1998), and Liu and Chen (1995).

A numerical algorithm for achieving resampling is as follows.

- Obtain  $N$  random draws from a  $U(0, 1)$  distribution, and sort in ascending order to obtain  $\{i^i\}_{i=1}^N$ . (Do this one time, prior to implementation of the SIR algorithm.)
- Inside the SIR algorithm, upon completion of the filtering step, construct the cumulative density function (cdf)  $\widehat{F}(s_t | \gamma_t)$  corresponding with  $\{\omega_t^i, s_t^i\}_{i=1}^N$  using<sup>1</sup>

$$\widehat{F}(s_t^i | \gamma_t) = \sum_{j=1}^i \omega_t^j.$$

- Using the initialization  $L^0 = \widehat{F}(s_t^0 | \gamma_t) = 0$ , for  $\ell = 1, \dots, N$ , determine the value of  $L^\ell$  such that the elements of  $\{j^j\}_{j=L^{\ell-1}+1}^{L^\ell}$  satisfy  $\widehat{F}(s_t^{\ell-1} | \gamma_t) < j^\ell \leq \widehat{F}(s_t^\ell | \gamma_t)$ ; include  $L^\ell - L^{\ell-1}$  values of  $s_t^\ell$  in the resampled swarm.

While SIR methods overcome the problem of degeneracy, they are prone to being numerically inefficient. The source of this problem is the propagation step: the candidate drawings  $s_t^i$  generated by  $f(s_t | s_{t-1}^i, \gamma_{t-1})$  do not take into account information regarding the plausibility of  $s_t$  in light of the observation  $y_t$ . It is in this sense that  $\{s_t^i\}_{i=1}^N$  is generated blindly. For cases in which  $f(s_t | s_{t-1}, \gamma_{t-1})$  is misaligned with  $f(y_t | s_t, \gamma_{t-1})$ , a relatively small number of candidate draws will be assigned appreciable weight by  $f(y_t | s_t, \gamma_{t-1})$ . In light of (10.7), the weights  $\{\omega_t^i\}_{i=1}^N$  produced in the filtering stage will be unevenly distributed, and thus in the resampling stage a relatively small number of unique particles will be resampled from the swarm  $\{s_t^i\}_{i=1}^N$ . Thus a manifestation of numerical inefficiency in this case is sample impoverishment.

There are two general cases under which misalignment between  $f(s_t | s_{t-1}, \gamma_{t-1})$  and  $f(y_t | s_t, \gamma_{t-1})$  can be acute. First,  $f(y_t | s_t, \gamma_{t-1})$  can be tightly distributed relative to  $f(s_t | s_{t-1}, \gamma_{t-1})$ , loosely referenced as a needle problem. Second, the realization of an outlier for  $y_t$  can result in the location of  $f(y_t | s_t, \gamma_{t-1})$  far in the tails of  $f(s_t | s_{t-1}, \gamma_{t-1})$ . Both cases can induce significant problems with sample impoverishment.

<sup>1</sup>In GAUSS, the command `cumsumc` accomplishes this step.

As a demonstration, we conduct a series of Monte Carlo experiments involving the repeated approximation of the sample log-likelihood  $\ln L(\gamma_T | \mu)$  for a given data set  $\gamma_T$ . Each sample log-likelihood is approximated using a distinct set of common random numbers (CRNs); the standard deviation in the approximations obtained across sets of CRNs provides an estimate of the numerical standard error (NSE) associated with the approximation of  $\ln L(\gamma_T | \mu)$ .

We conduct a series of such experiments using artificial data generated from a simple example model adapted from Fernandez-Villaverde and Rubio-Ramirez (2004b). Let  $s_t$  be single-valued, and let the transition and measurement equations be given by

$$s_t = \alpha + \beta \frac{s_{t-1}}{1 + s_{t-1}^2} + v_t, \quad (10.12)$$

$$y_t = s_t + \gamma \cdot u_t, \quad (10.13)$$

where  $v_{t+1} \sim N(0, \sigma_v^2)$  and  $u_t$  is  $t$ -distributed with  $v$  degrees of freedom:

$$f(u_t) \sim (\nu + u_t^2)^{-0.5(\nu+1)}, \quad \text{Var}(u_t) = \frac{\nu}{\nu - 2} \text{ for } \nu > 2.$$

Note that the expectation of  $s_{t+1} | s_t$  is highly nonlinear around  $s_t = 0$ , and becomes virtually constant for  $|s_t| > 10$ .

To parameterize the model, we fix  $\alpha$  and  $\beta$  at 0.5 (as adjustments to these settings have minimal impact on efficiency), and evaluate the numerical efficiency of the particle filter using eight combinations of specifications for  $(\sigma_v, \gamma)$ :  $\sigma_v = (0.5, 1)$ ,  $\gamma = (0.01, 0.05, 0.5, 1)$ . As an approximation of the unconditional distribution  $f(s_0)$ , we use a Normal distribution parameterized using sample summary statistics calculated using a lengthy simulation of  $\{s_t\}$ . Finally, we fix  $\nu = 4$ , and leave the study of implications in adjustments to  $\nu$  as an exercise. Using each parameterization, we generate a single realization  $\gamma_T$ ,  $T = 10$ , and calculate  $\ln L(\gamma_T | \mu)$  using 100 sets of CRNs with swarm sizes  $N = 10,000$ . The means and standard deviations of  $\ln L(\gamma_T | \mu)$  calculated across sets of CRNs are reported in table 10.1.

For a given specification of  $\sigma_v$ , note that as  $\gamma$  decreases the measurement density tightens and NSEs increase: for  $\sigma_v = 0.5$ , NSE increases by a factor of roughly 23 as  $\gamma$  falls from 1 to 0.01; for  $\sigma_v = 1$ , the factor is roughly 17. This reflects the needle phenomenon. In turn, note that NSEs increase as  $\sigma_v$ , (and hence the volatility of  $s_t$ ) increases. Thus we see that volatile states, coupled with tightly distributed measurements, pose relative challenges to the bootstrap particle filter.

### Exercise 10.1

Assess the impact of adjustments in  $\nu$  on the results reported in table 10.1.

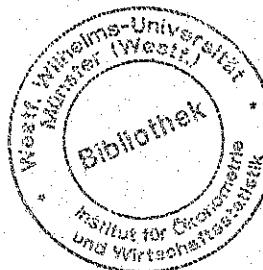


TABLE 10.1

Demonstration of the Numerical Efficiency of the Bootstrap Particle Filter

$\sigma_u$	0.5	1
0.01	-53.506, 0.2574	-60.437, 0.3764
0.05	-37.608, 0.1103	-44.434, 0.1584
0.5	-17.320, 0.0221	-22.542, 0.0410
1	-14.042, 0.0112	-17.317, 0.0222

Note: The entry of each cell contains a Monte Carlo mean and standard deviation.

### 10.3 Conditionally Optimal Filters

Conditionally optimal filters implement modifications of the propagation stage of unadapted filtering algorithms. Modifications seek to approximate the optimal (in terms of numerical efficiency) importance sampler used to generate proposal drawings of  $s_t^i$  given information contained in  $y_t$ , but conditional on the discrete approximation of the period- $(t-1)$  filtering density

$$\hat{f}(s_{t-1} | \mathcal{Y}_{t-1}) = \sum_{i=1}^N \omega_{t-1}^i \delta_{s_{t-1}^i}(s_{t-1}). \quad (10.14)$$

The conditionally optimal importance sampler is given by  $f(s_t | s_{t-1}, \mathcal{Y}_t)$ . Note the dependence of  $f(s_t | s_{t-1}, \mathcal{Y}_t)$  on  $\mathcal{Y}_t$ , rather than  $\mathcal{Y}_{t-1}$ : this serves to overcome the “blindness” problem faced by unadapted samplers. This density is not analytically tractable in general, and therefore must typically be approximated. However, it is useful to proceed with its characterization, which stems from a critical conditional independence assumption implicit in the specification of state-space representations.

The assumption is that  $y_t$  is independent of  $s_{t-1}$  given  $(s_t, \mathcal{Y}_{t-1})$ :

$$f(y_t | s_t, \mathcal{Y}_{t-1}) = f(y_t | s_t, s_{t-1}, \mathcal{Y}_{t-1}). \quad (10.15)$$

This assumption implies the attainment of  $f(s_t | s_{t-1}, \mathcal{Y}_t)$  via the factorization

$$f(y_t | s_t, \mathcal{Y}_{t-1}) \cdot f(s_t | s_{t-1}, \mathcal{Y}_{t-1}) = f(s_t | s_{t-1}, \mathcal{Y}_t) \cdot f(y_t | s_{t-1}, \mathcal{Y}_{t-1}). \quad (10.16)$$

If this factorization were tractable analytically, and  $f(s_t | s_{t-1}, \mathcal{Y}_t)$  could be implemented as an importance sampler, likelihood evaluation and

filtering could proceed using an alternative expression of the likelihood integral (10.5). Specifically, rewriting (10.5) by substituting for the predictive density  $f(s_t | \mathcal{Y}_{t-1})$  using (10.4), we obtain

$$f(y_t | \mathcal{Y}_{t-1}) = \int \int [f(y_t | s_t, \mathcal{Y}_{t-1}) f(s_t | s_{t-1}, \mathcal{Y}_{t-1})] f(s_{t-1} | \mathcal{Y}_{t-1}) ds_t ds_{t-1}. \quad (10.17)$$

Note that the term in brackets constitutes the left-hand side of (10.16). Substituting for this term yields

$$f(y_t | \mathcal{Y}_{t-1}) = \int \int f(y_t | s_{t-1}, \mathcal{Y}_{t-1}) \cdot [f(s_t | s_{t-1}, \mathcal{Y}_t) f(s_{t-1} | \mathcal{Y}_{t-1})] ds_t ds_{t-1}. \quad (10.18)$$

Then given that  $f(s_t | s_{t-1}, \mathcal{Y}_t)$  integrates to 1 in  $s_t$ , we obtain

$$f(y_t | \mathcal{Y}_{t-1}) = \int f(y_t | s_{t-1}, \mathcal{Y}_{t-1}) f(s_{t-1} | \mathcal{Y}_{t-1}) \cdot ds_{t-1}. \quad (10.19)$$

With  $\{\omega_{t-1}^i, s_{t-1}^i\}_{i=1}^N$  inherited from the period- $(t-1)$  stage of the filtering algorithm, (10.19) indicates that likelihood integration can be accomplished via direct importance sampling:

$$\hat{f}(y_t | \mathcal{Y}_{t-1}) = \sum_{i=1}^N \omega_{t-1}^i \cdot f(y_t | s_{t-1}^i, \mathcal{Y}_{t-1}). \quad (10.20)$$

Further, with  $f(s_t | s_{t-1}, \mathcal{Y}_t)$  available as an importance sampler, propagation can be achieved by drawing a particle  $s_t^i$  from  $f(s_t | s_{t-1}^i, \mathcal{Y}_t)$  for each particle  $s_{t-1}^i$ . In light of the bracketed term in (10.18), the weight associated with  $s_t^i$  is given by

$$\omega_t^i = \frac{\omega_{t-1}^i \cdot f(y_t | s_{t-1}^i, \mathcal{Y}_{t-1})}{\sum_{j=1}^N \omega_{t-1}^j \cdot f(y_t | s_{t-1}^j, \mathcal{Y}_{t-1})}, \quad (10.21)$$

which is analogous to the weight function (10.11) associated with unadapted filters. Given  $\{\omega_t^i, s_t^i\}_{i=1}^N$ , the filtering density is estimated as

$$\hat{f}(s_t | \mathcal{Y}_t) = \sum_{i=1}^N \omega_t^i \delta_{s_t^i}(s_t).$$

A summary of the period- $t$  algorithm is as follows.

**Propagation:** Inherit  $\{\omega_{t-1}^i, s_{t-1}^i\}_{i=1}^N$  from period  $(t-1)$ . Then for each particle  $s_{t-1}^i$ , draw a particle  $s_t^i$  from the conditionally optimal importance sampler  $f(s_t | s_{t-1}^i, \mathcal{Y}_t)$ .

**Likelihood integral:** Using  $\{\omega_{t-1}^i, s_{t-1}^i\}_{i=1}^N$ , estimate  $\hat{f}(y_t | \mathcal{Y}_{t-1})$  as in (10.20).

**Filtering:** Assign the weight  $\omega_t^i$  to  $s_t^i$  as in (10.21).

**Resampling:** Draw from  $\{s_t^i\}_{i=1}^N$  with replacement, with the probability of selecting  $s_t^i$  given by  $\omega_t^i$ . Denote the resulting swarm as  $\{\omega_t^i, s_t^i\}_{i=1}^N$ , where following resampling  $\omega_t^i = 1/N \forall i$ .

**Continuation:** Pass the weights and mass points  $\{\omega_t^i, s_t^i\}_{i=1}^N$  to the period- $(t+1)$  propagation step and proceed through period  $T$ .

The benefit of implementing  $f(s_t | s_{t-1}, \mathcal{Y}_t)$  as an importance sampler can be seen by comparing the weight functions in (10.11) and (10.21). Note that the latter does not depend on  $s_t^i$ , but only on  $s_{t-1}^i$ . By implication, its conditional variance is zero given  $\{s_{t-1}^i\}_{i=1}^N$ , motivating its characterization as an optimal sampler (following Zaritskii, Svetnik, and Shimelevich, 1975; and Akasaki and Kumamoto, 1977).

As noted, since the factorization in (10.16) is tractable analytically only in special cases, the conditionally optimal sampler represents a theoretical rather than an operational benchmark. The construction of approximations proceeds as follows.

Recall that (10.17) was obtained by substituting for the predictive density  $f(s_t | \mathcal{Y}_{t-1})$  in (10.5) using (10.4). Here, we instead substitute for  $f(s_t | \mathcal{Y}_{t-1})$  using

$$\hat{f}(s_t | \mathcal{Y}_{t-1}) = \sum_{i=1}^N \omega_{t-1}^i \cdot f(s_t | s_{t-1}^i, \mathcal{Y}_{t-1}), \quad (10.22)$$

which yields an approximation of the likelihood integral given by

$$\hat{f}(y_t | \mathcal{Y}_t) = \sum_{i=1}^N \int \omega_{t-1}^i \cdot f(y_t | s_t, \mathcal{Y}_{t-1}) \cdot f(s_t | s_{t-1}^i, \mathcal{Y}_{t-1}) ds_t. \quad (10.23)$$

Implementation of this approximation proceeds by interpreting the integrand in (10.23) as the kernel of a mixed density for the pair  $(s_t, k_t)$ , where  $k_t$  is a discrete random index variable distributed over  $\{1, \dots, N\}$ . Under this interpretation, the likelihood integral is evaluated via importance sampling using an auxiliary mixed kernel of the form

$$g_t(s, k) = \omega_{t-1}^k \cdot p_t(s, k) \cdot f(s | s_{t-1}^k, \mathcal{Y}_{t-1}), \quad (10.24)$$

where  $p(s, k)$  is introduced to minimize the Monte Carlo variance of the likelihood estimate conditionally on  $k$ .

Numerical tractability requires that  $p_t(s, k)$  be selected such that  $g_t(s, k)$  can be normalized into an operational importance sampler  $g_t(s, k)$ . Many

### 10.3 Conditionally Optimal Filters

specific approaches have been developed for selecting  $p_t(s, k)$ ; here we present a prominent example developed by Pitt and Shephard (1999), which yields what they term the Auxiliary Particle Filter (APF):

$$p_t(s, k) = f(y_t | \mu_t^k, \mathcal{Y}_{t-1}), \quad (10.25)$$

$$\mu_t^k = E(s_t | s_{t-1}^k, \mathcal{Y}_{t-1}). \quad (10.26)$$

The corresponding importance sampler is given by

$$g_t(s, k) = \pi_t^k \cdot f(s | s_{t-1}^k, \mathcal{Y}_{t-1}), \quad (10.27)$$

$$\pi_t^k = D_t^{-1} \cdot \omega_{t-1}^k \cdot f(y_t | \mu_t^k, \mathcal{Y}_{t-1}), \quad (10.28)$$

$$D_t = \sum_{j=1}^N \omega_{t-1}^j \cdot f(y_t | \mu_t^j, \mathcal{Y}_{t-1}). \quad (10.29)$$

Let  $\omega_t(s, k)$  denote the ratio between the integrand in (10.23) and  $g_t(s, k)$ :

$$\omega_t(s, k) = D_t \cdot \frac{f(y_t | s, \mathcal{Y}_{t-1})}{f(y_t | \mu_t^k, \mathcal{Y}_{t-1})}. \quad (10.30)$$

Also, let  $\{s_t^i, k_t^i\}$  denote  $N$  iid draws from  $g_t(s, k)$ . Specifically,  $k_t^i$  is drawn from the multinomial distribution  $MN(N; \{\pi_t^i\}_{i=1}^N)$ , and  $s_t^i$  from  $f(s_t | s_{t-1}^{k_t^i}, \mathcal{Y}_{t-1})$ . The corresponding estimate of the likelihood function is given by

$$\hat{f}_N(y_t | \mathcal{Y}_{t-1}) = \frac{1}{N} \sum_{i=1}^N \tilde{\omega}_t^i, \quad (10.31)$$

$$\tilde{\omega}_t^i \equiv \omega_t(s_t^i, k_t^i).$$

In turn, the filtering density is approximated as

$$\hat{f}_N(s_t | \mathcal{Y}_t) = \sum_{i=1}^N \omega_t^i \delta_{s_t^i}(s_t), \quad (10.32)$$

$$\omega_t^i \equiv \frac{\tilde{\omega}_t^i}{\sum_{j=1}^N \tilde{\omega}_t^j}.$$

An algorithmic summary of the period- $t$  APF algorithm is as follows.

**Propagation:** Inheriting  $\{\omega_{t-1}^i, s_{t-1}^i\}_{i=1}^N$  from period  $(t-1)$ , construct  $\{\pi_t^i\}_{i=1}^N$  as in (10.28). Next, obtain draws  $k_t^i$  by sampling with replacement from  $\{1, \dots, N\}$  with

$$\Pr(k_t^i = j) = \pi_t^j.$$

Finally, obtain draws  $s_t^i$  from  $f(s_t|s_{t-1}^{k_t^i}, Y_t)$ .

**Likelihood integral:** Using  $\{\omega_{t-1}^i, s_t^i, k_t^i\}_{i=1}^N$ , construct  $\{\omega_t^i\}_{i=1}^N$  as in (10.30). Then calculate  $\widehat{f}(y_t|Y_{t-1})$  as in (10.31).

**Filtering:** Assign the weight  $\omega_t^i$  to  $s_t^i$  as in (10.32).

**Resampling:** Draw from  $\{s_t^i\}_{i=1}^N$  with replacement, with the probability of selecting  $s_t^i$  given by  $\omega_t^i$ . Denote the resulting swarm as  $\{\omega_t^i, s_t^i\}_{i=1}^N$ , where following resampling  $\omega_t^i = 1/N \forall i$ .

**Continuation:** Pass the weights and mass points  $\{\omega_t^i, s_t^i\}_{i=1}^N$  to the period- $(t+1)$  propagation step and proceed through period  $T$ .

### Exercise 10.2

Reconstruct table 10.1 using the APF in place of the bootstrap particle filter. Note how gains in numerical efficiency vary as a function of  $\gamma$ : what is the intuition behind this variation?

Under special circumstances further adaption can be achieved. If  $f(s_t|s_{t-1}, Y_{t-1})$  belongs to a family of densities closed under multiplication, then selecting  $p_t(s, k)$  in (10.24) from that family can produce an operational and improved sampler. Its integrating constant obtains by first integrating the product  $p_t(s, k) \cdot f(s|s_{t-1}^k, Y_{t-1})$  with respect to  $s$ , then summing the remainder over  $k$ . Examples wherein  $f(s_t|s_{t-1}, Y_{t-1})$  is Normally distributed in  $s_t$  are discussed in Pitt and Shephard (1999) and Smith and Santos (2006). Using for  $p_t(s, k)$  a first-order Taylor Series expansion of  $\ln f(y_t|s_t, Y_{t-1})$  in  $s_t$  around  $\mu_t^k$  yields the adapted particle filter of Pitt and Shephard; implementation of a second-order expansion is demonstrated by Smith and Santos.

We conclude this section by reiterating that samplers of this class are only (approximately) optimal conditionally: a different weight is associated with each particle  $s_{t-1}^i$ , implying that while the conditional Monte Carlo variance of  $\omega_t^i$  is zero, its unconditional Monte Carlo variance is not. The limitation to conditional optimality is inherently linked to the use of mixtures of Dirac measures in representing filtering densities. Next, we discuss the attainment of (approximate) unconditional optimality, which is the target of the EIS filter.

### 10.4 Unconditional Optimality: The EIS Filter

Recall the characterization of the likelihood function  $f(y_t|Y_{t-1})$  as resulting from the nested integral

$$f(y_t|Y_{t-1}) = \int \int f(y_t|s_t, Y_{t-1}) \cdot f(s_t|s_{t-1}, Y_t) \cdot f(s_{t-1}|Y_{t-1}) ds_t ds_{t-1}.$$

Outside the linear/Normal class of models, all filtering algorithms operationalize the approximation of this integral by substituting for  $f(s_{t-1}|Y_{t-1})$  using an approximation  $\widehat{f}(s_{t-1}|Y_{t-1})$ :

$$f(y_t|Y_{t-1}) = \int \int f(y_t|s_t, Y_{t-1}) \cdot f(s_t|s_{t-1}, Y_{t-1}) \cdot \widehat{f}(s_{t-1}|Y_{t-1}) ds_{t-1} ds_t. \quad (10.33)$$

As we have discussed above, particle-based algorithms represent  $\widehat{f}(s_{t-1}|Y_{t-1})$  using the discrete approximation

$$\widehat{f}(s_{t-1}|Y_{t-1}) = \sum_{i=1}^N \omega_{t-1}^i \cdot \delta_{s_{t-1}^i}(s_{t-1}).$$

In contrast, the EIS filter seeks to attain unconditional optimality in approximating (10.33) via the use of continuous approximations  $\widehat{f}(s_{t-1}|Y_{t-1})$ .

The unconditionally optimal sampler for approximating  $f(y_t|Y_{t-1})$  is given by  $f(s_t, s_{t-1}|Y_t)$ . As is true for its conditionally optimal counterpart  $f(s_t|s_{t-1}, Y_t)$ , it is not tractable analytically, and therefore must be approximated. Nevertheless, it is useful to proceed with its characterization.

Denote the integrand in (10.33) as  $\varphi_t(s_{t-1}, s_t)$ , where the subscript  $t$  on  $\varphi$  indicates conditionality on  $Y_t$ , and consider the (theoretical) factorization given by

$$f(y_t|s_t, Y_{t-1}) \cdot f(s_t|s_{t-1}, Y_{t-1}) \cdot \widehat{f}(s_{t-1}|Y_{t-1}) = f(s_t, s_{t-1}|Y_t) \cdot f(y_t|Y_{t-1}). \quad (10.34)$$

If analytically tractable,  $f(s_t, s_{t-1}|Y_t)$  would clearly be the unconditionally optimal importance sampler for approximating (10.33), as a single draw from it would produce an estimate of  $f(y_t|Y_{t-1})$  with zero Monte Carlo variance. The period- $t$  filtering density would then obtain by marginalization with respect to  $s_{t-1}$ :

$$f(s_t|Y_t) = \int f(s_t, s_{t-1}|Y_t) ds_{t-1}. \quad (10.35)$$

The goal of the EIS filter is to approximate  $f(s_t, s_{t-1}|Y_t)$  by constructing an importance sampler in  $(s_{t-1}, s_t)$  tailored to  $\varphi_t(s_{t-1}, s_t)$ . In the

present context, recall from section 4 of chapter 9 that tailoring entails the construction of a sampler  $g(s_{t-1}, s_t; \hat{a}_t)$ , with kernel  $k(s_{t-1}, s_t; \hat{a}_t)$  and integrating constant  $\chi(\hat{a}_t)$ , where  $\hat{a}_t$  is chosen to best match the location and shape of  $\varphi_t(s_{t-1}, s_t)$ . Specifically,  $\hat{a}_t$  obtains as the solution to the fixed-point problem

$$(\hat{a}_t^{l+1}, \hat{c}_t^{l+1}) = \arg \min_{(a_t, c_t)} \sum_{i=1}^R [\ln \varphi_t(s_{t-1}^i, s_t^i) - c_t - \ln k(s_{t-1}^i, s_t^i; a_t)]^2, \quad (10.36)$$

where  $c_t$  is an intercept term that accounts for factors in  $\varphi_t()$  and  $k()$  that do not depend upon  $(s_{t-1}, s_t)$ , and  $(s_{t-1}^i, s_t^i)$  represent draws from  $g(s_{t-1}, s_t; \hat{a}_t^l)$ . The inclusion of the subscript  $t$  in characterizing  $\hat{a}_t$  reflects its dependence on  $\gamma_t$ .

Upon convergence, the likelihood function is approximated as

$$\hat{f}(y_t | \gamma_{t-1}) = \frac{1}{N} \sum_{i=1}^N \omega_t(s_{t-1}^i, s_t^i; \hat{a}_t), \quad (10.37)$$

$$\omega_t(s_{t-1}, s_t; \hat{a}_t) = \frac{\varphi_t(s_{t-1}, s_t)}{g_t(s_{t-1}, s_t; \hat{a}_t)}. \quad (10.38)$$

In turn, the filtering density is approximated as the marginal distribution of  $g(s_{t-1}, s_t; \hat{a}_t)$  in  $s_t$ :

$$\hat{f}(s_t | \gamma_t) = \int g(s_{t-1}, s_t; \hat{a}_t) ds_{t-1}. \quad (10.39)$$

Two notes bear emphasis at this point. First, particularly given the presence of outliers and/or tightly distributed measurement densities, the selection of a good initial sampler  $g_t(s_{t-1}, s_t; \hat{a}_t^0)$  is necessary for achieving rapid and robust convergence. Below we characterize in detail a systematic means of constructing such a sampler. The sampler is Normally distributed, and is parameterized using Taylor Series approximations of potential nonlinear relationships featured in underlying measurement and state-transition equations. This construction follows that proposed by Durbin and Koopman (1997) for producing final importance-sampling distributions in filtering applications; under the EIS filter, this instead serves to initialize construction of the optimized sampler  $g(s_{t-1}, s_t; \hat{a}_t)$ . As DeJong et al. (2010) demonstrate, this approach to initializing  $g_t(s_{t-1}, s_t; \hat{a}_t^0)$  is highly effective in applications to DSGE models.

Second, while the EIS filter seeks to attain unconditional optimality, it cannot do so exactly absent the analytical tractability of  $f(s_t, s_{t-1} | \gamma_t)$ .

#### 10.4 Unconditional Optimality: The EIS Filter

That is,  $g_t(s_{t-1}, s_t; \hat{a}_t^0)$  merely serves as an approximation of  $f(s_t, s_{t-1} | \gamma_t)$ . Moreover, while the approximation is designed to achieve optimal numerical efficiency, optimization is conditional upon the family of distributions from which  $g_t()$  is chosen. Currently, the EIS filter is fully operational only for auxiliary kernels that belong to the exponential family of distributions, since the auxiliary least-squares problem (10.36) is then linear in  $a$ . Nevertheless, as DeJong et al. (2010) demonstrate, specifications of  $g_t()$  chosen from the exponential family have proven to yield highly efficient importance samplers in applications to DSGE models.

Given a distributional family chosen for  $g_t()$ , a systematic means of constructing  $g_t(s_{t-1}, s_t; \hat{a}_t^0)$ , and with  $\hat{f}(s_0 | \gamma_0)$  initialized using  $f(s_0)$ , a summary of the period- $t$  algorithm of the EIS filter is as follows.

**Propagation:** Inheriting  $\hat{f}(s_{t-1} | \gamma_{t-1})$  from period  $(t-1)$ , obtain the integrand

$$\varphi_t(s_{t-1}, s_t) = f(y_t | s_t, \gamma_{t-1}) \cdot f(s_t | s_{t-1}, \gamma_{t-1}) \cdot \hat{f}(s_{t-1} | \gamma_{t-1})$$

appearing in the likelihood integral (10.33).

**EIS optimization:** Construct an initialized sampler  $g_t(s_{t-1}, s_t; \hat{a}_t^0)$ , and obtain the optimized parameterization  $\hat{a}_t$  as the solution to (10.36). This yields  $g_t(s_{t-1}, s_t; \hat{a}_t)$ .

**Likelihood integral:** Obtain draws  $\{s_{t-1}^i, s_t^i\}_{i=1}^N$  from  $g_t(s_{t-1}, s_t; \hat{a}_t)$ , and approximate  $\hat{f}(y_t | \gamma_{t-1})$  as in (10.37).

**Filtering:** Approximate  $\hat{f}(s_t | \gamma_t)$  as in (10.39).

**Continuation:** Pass  $\hat{f}(s_t | \gamma_t)$  to the period- $(t+1)$  propagation step and proceed through period  $T$ .

Before we detail the construction of  $g_t(s_{t-1}, s_t; \hat{a}_t^0)$ , it is necessary to describe a special case involving the presence of identities among the system of state-transition equations included in the model. In this context, identities are termed degenerate transitions.

##### 10.4.1 Degenerate Transitions

State transitions often include identities that effectively reduce the dimension of integration in  $(s_t, s_{t-1})$ . For example, in the RBC model presented in section 1 of chapter 3, the state-transition equations for physical capital  $k_t$  and total factor productivity  $z_t$  are given by

$$k_{t+1} = i_t + (1 - \delta)k_t, \\ \ln z_t = (1 - \rho) \ln (\bar{z}) + \rho \ln z_{t-1} + \varepsilon_t,$$

with  $\varepsilon_t$  distributed as  $N(0, \sigma^2)$ . The former is an identity, implying a degenerate distribution for  $k_{t+1}$  given  $(k_t, z_t)$  (since the behavior of  $i_t$  is

determined by a policy function of the form  $i(k_t, z_t)$ . In working with continuous representations of  $\widehat{f}(s_{t-1}|\Upsilon_{t-1})$ , identities require special care.

Let  $s_t$  partition into  $s_t = (p_t, q_t)$  such that the transition consists of a set of identities

$$q_t \equiv \phi(p_t, s_{t-1}), \quad (10.40)$$

in combination with a proper transition density  $f(p_t|s_{t-1}, \Upsilon_{t-1})$ . With reference to (10.33), the product  $f(p_t|s_{t-1}, \Upsilon_{t-1}) \cdot f(s_{t-1}|\Upsilon_{t-1})$  defines a joint density for  $(p_t, s_{t-1}) = (p_t, p_{t-1}, q_{t-1})$  given  $\Upsilon_{t-1}$ , which turns out to be difficult to work with in achieving filtering. However, if we reinterpret (10.40) as a transformation of variables from  $(p_t, p_{t-1}, q_{t-1})$  into  $(p_t, q_t, p_{t-1}) = (s_t, p_{t-1})$ , filtering becomes straightforward. We proceed under the assumption that this transformation is one-to-one. Denote its inverse and associated Jacobian respectively as

$$q_t = \psi(s_t, p_{t-1}), \quad (10.41)$$

$$J(s_t, p_{t-1}) = \left\| \frac{\partial}{\partial q_t} \psi(s_t, p_{t-1}) \right\|, \quad (10.42)$$

where  $\|\bullet\|$  denotes the absolute value of a determinant. Then via a standard change-of-variables argument, the likelihood integral in (10.33) can be rewritten as

$$\begin{aligned} f(y_t|\Upsilon_{t-1}) &= \int \int f(y_t|s_t, \Upsilon_{t-1}) \cdot J(s_t, p_{t-1}) [f(p_t|s_{t-1}, \Upsilon_{t-1}) \\ &\quad \cdot \widehat{f}(s_{t-1}|\Upsilon_{t-1})] |_{q_{t-1}=\psi(s_t, p_{t-1})} ds_t dp_{t-1}. \end{aligned} \quad (10.43)$$

Application of the EIS filter to (10.43) rather than (10.33) then proceeds precisely as described above, producing an EIS sampler  $g_t(s_t, p_{t-1}|\widehat{a}_t)$  for which an approximation of the filtering density  $f(s_t|\Upsilon_t)$  obtains directly by marginalization with respect to  $p_{t-1}$ .

#### 10.4.2 Initializing the Importance Sampler

We now describe an automated procedure for producing effective initialized samplers  $g_t(s_{t-1}, s_t|\widehat{a}_t^0)$ . The idea behind the procedure is to approximate the individual components of the targeted integrands in either (10.33) or (10.43) as Normal distributions parameterized using local Taylor Series approximations of nonlinear relationships featured in their underlying equations. Here we shall focus on the case in which (10.33) is the targeted integral. Although specialization to (10.43) is straightforward, we highlight this special case explicitly in section 4, in the context of applications to DSGE models.

For ease of notation, here we ignore the dependence of  $f(y_t|s_t, \Upsilon_{t-1})$ ,  $f(s_t|s_{t-1}, \Upsilon_{t-1})$ , and  $\widehat{f}(s_{t-1}|\Upsilon_{t-1})$  on  $\Upsilon_{t-1}$ . Further, we ignore the dependence of the parameterization of  $\widehat{f}(s_{t-1}|\Upsilon_{t-1})$  on time ( $t-1$ ); we shall do the same for the approximations  $(\widehat{f}(y_t|s_t), \widehat{f}(s_t|s_{t-1}))$  of  $(f(y_t|s_t, \Upsilon_{t-1}), f(s_t|s_{t-1}, \Upsilon_{t-1}))$ . We take  $y_t$  as  $m$ -dimensional, and  $s_t$  as  $n$ -dimensional. Finally, we use the following functional notation to represent the Normal density of the  $r$ -dimensional vector  $\chi$ :

$$f_N^r(\chi|\mu, \Omega) = (2\pi)^{-\frac{r}{2}} |\Omega|^{-\frac{1}{2}} \exp -\frac{1}{2} (\chi - \mu)' \Omega^{-1} (\chi - \mu). \quad (10.44)$$

The initial sampler  $g_t(s_{t-1}, s_t|\widehat{a}_t^0)$  is built sequentially, beginning with a representation of the period- $(t-1)$  filtering density given by

$$\widehat{f}(s_{t-1}) = f_N^n(s_{t-1}|\mu, \Omega). \quad (10.45)$$

To repeat, here we have ignored the dependence of  $(\mu, \Omega)$  on time ( $t-1$ ). Next, we construct a first-order approximation of the state-transition equations  $s_t = \gamma(s_{t-1}, \Upsilon_{t-1}, v_t)$  around  $(s_{t-1} = \mu, v_t = 0)$ , obtaining

$$s_t = A + B s_{t-1},$$

where  $A$  is  $n \times 1$ , and  $B$  is  $n \times n$ . Using this approximation, the state-transition density is approximated as

$$\widehat{f}(s_t|s_{t-1}) = f_N^n(s_t|A + B s_{t-1}, \Sigma), \quad (10.46)$$

where  $\Sigma$  is specified in accordance with  $f(s_t|s_{t-1}, \Upsilon_{t-1})$ . Finally, we construct a first-order approximation of the measurement equations  $y_t = \delta(s_t, \Upsilon_{t-1}, u_t)$  around  $(s_t = A + B\mu, u_t = 0)$ , obtaining

$$y_t = C + D s_t,$$

where  $C$  is  $m \times 1$ , and  $D$  is  $m \times n$ . Using this approximation, the measurement density is approximated as

$$\widehat{f}(y_t|s_t) = f_N^m(y_t|C + D s_t, \Lambda), \quad (10.47)$$

where  $\Lambda$  is specified in accordance with  $f(y_t|s_t, \Upsilon_{t-1})$ .

Given (10.45)–(10.47),  $g_t(s_{t-1}, s_t|\widehat{a}_t^0)$  is constructed via the straightforward manipulation of Normal densities (for textbook references, see, e.g., Zellner, 1971; or Bauwens, Lubrano, and Richard, 1999). To begin, (10.45) and (10.46) combine to form

$$\begin{aligned}\hat{f}\left(\begin{array}{c} s_t \\ s_{t-1} \end{array}\right) &= f_N^{2n}\left(\begin{array}{c} s_t \\ s_{t-1} \end{array} \mid \begin{bmatrix} A + B\mu \\ \mu \end{bmatrix}, \begin{bmatrix} \Sigma + B\Omega B' & B\Omega \\ \Omega B' & \Omega \end{bmatrix}\right) \\ &\equiv f_N^{2n}\left(\begin{array}{c} s_t \\ s_{t-1} \end{array} \mid \tilde{\mu}, \tilde{\Omega}\right),\end{aligned}\quad (10.48)$$

where

$$\tilde{\mu} = \tilde{A} + \tilde{B}\mu, \quad \underbrace{\tilde{A}}_{2n \times 1} = \begin{bmatrix} A \\ 0 \\ n \times 1 \end{bmatrix}, \quad \underbrace{\tilde{B}}_{2n \times n} = \begin{bmatrix} B \\ I \\ n \times n \end{bmatrix}, \quad (10.49)$$

$$\tilde{\Omega} = \begin{bmatrix} \Sigma & 0 \\ 0 & 0 \end{bmatrix} + \tilde{B}\Omega\tilde{B}', \quad (10.50)$$

Next, rewrite (10.47) as

$$\begin{aligned}\hat{f}\left(y_t \mid \begin{array}{c} s_t \\ s_{t-1} \end{array}\right) &= f_N^m(y_t \mid C + \tilde{D}\begin{bmatrix} s_t \\ s_{t-1} \end{bmatrix}, \Lambda), \\ \underbrace{\tilde{D}}_{m \times 2n} &= \begin{bmatrix} D & 0 \\ m \times n & \end{bmatrix},\end{aligned}$$

and combine with (10.48) to obtain

$$\begin{aligned}\hat{f}\left(\begin{array}{c} y_t \\ s_t \\ s_{t-1} \end{array}\right) &= f_N^{m+2n}\left(\begin{array}{c} y_t \\ s_t \\ s_{t-1} \end{array} \mid \begin{bmatrix} C + \tilde{D}\tilde{\mu} \\ \tilde{\mu} \end{bmatrix}, \begin{bmatrix} \Lambda + \tilde{D}\tilde{\Omega}\tilde{D}' & \tilde{D}\tilde{\Omega} \\ \tilde{\Omega}\tilde{D}' & \tilde{\Omega} \end{bmatrix}\right) \\ &= f_N^{m+2n}\left(\begin{array}{c} y_t \\ s_t \\ s_{t-1} \end{array} \mid m, \Gamma\right),\end{aligned}\quad (10.51)$$

where

$$m = \tilde{C} + \tilde{D}\tilde{\mu}, \quad \underbrace{\tilde{C}}_{m+2n \times 1} = \begin{bmatrix} C \\ 0 \\ 2n \times 1 \end{bmatrix}, \quad \underbrace{\tilde{D}}_{m+2n \times 2n} = \begin{bmatrix} D \\ I \\ 2n \times 2n \end{bmatrix}, \quad (10.52)$$

$$\Gamma = \begin{bmatrix} \Lambda & 0 \\ 0 & 0 \end{bmatrix} + \tilde{D}\tilde{\Omega}\tilde{D}', \quad (10.53)$$

Finally, divide  $[y_t \ s_t \ s_{t-1}]'$  into  $(y_t, [s_t \ s_{t-1}]')$ , and partition  $m$  and  $\Gamma$  conformably with this division; that is,

$$m = \begin{bmatrix} \underbrace{m_1}_{m \times 1} \\ \underbrace{m_2}_{2n \times 1} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \underbrace{\Gamma_{11}}_{m \times m} & \underbrace{\Gamma_{12}}_{m \times 2n} \\ \underbrace{\Gamma_{21}}_{2n \times m} & \underbrace{\Gamma_{22}}_{2n \times 2n} \end{bmatrix}.$$

Then the initialization we seek is the conditional distribution of  $(s_t, s_{t-1})$  given  $y_t$ , which is constructed from the joint distribution of  $[y_t \ s_t \ s_{t-1}]'$  in (10.51):

$$\begin{aligned}\hat{f}\left(\begin{array}{c} s_t \\ s_{t-1} \end{array} \mid Y_t\right) &= f_N^{2n}\left(\begin{array}{c} s_t \\ s_{t-1} \end{array} \mid m_{2,1} + \Delta_{21} \cdot y_t, \Gamma_{22,1}\right) \\ &\equiv g_t(s_{t-1}, s_t | \hat{a}_t^0), \\ \Delta_{21} &= \Gamma_{21}\Gamma_{11}^{-1}, \quad \Gamma_{22,1} = \Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}, \\ m_{2,1} &= m_2 - \Delta_{21}m_1.\end{aligned}\quad (10.54)$$

Given the optimized sampler

$$g(s_{t-1}, s_t; \hat{a}_t) = f_N^{2n}\left(\begin{array}{c} s_t \\ s_{t-1} \end{array} \mid \mu_{\hat{a}_t}, \Omega_{\hat{a}_t}\right),$$

likelihood approximation is achieved as in (10.37). Then in light of its specification as a Normal density, the period- $t$  filtering density is given by

$$\hat{f}(s_t) = f_N^n(s_t | \mu, \Omega),$$

where  $\mu$  is defined as the first  $n$  elements of  $\mu_{\hat{a}_t}$ , and  $\Omega$  is defined as the first  $n \times n$  elements of  $\Omega_{\hat{a}_t}$ .

### 10.4.3 Example

We conclude this section by demonstrating the application of the EIS filter to the example state-space representation analyzed in section 2. Under this model, inputs into the construction of  $g_t(s_{t-1}, s_t | \hat{a}_t^0)$  are as follows:

$$\begin{aligned}A &= \alpha + \beta \frac{2\mu^3}{(1+\mu^2)^2}, \quad B = \beta \frac{1}{(1+\mu^2)} \left(1 - \frac{2\mu^2}{(1+\mu^2)}\right), \quad \Sigma = \sigma_v^2, \\ C &= 0, \quad D = 1, \quad \Lambda = \gamma^2 \frac{v}{v-2}.\end{aligned}$$

As with the bootstrap particle filter, we again conduct a series of Monte Carlo experiments involving the repeated approximation of the sample log-likelihood  $\ln L(Y_T | \mu)$  for a given data set  $Y_T$ . Each sample log-likelihood

is approximated using a distinct set of common random numbers (CRNs); the standard deviation in the approximations obtained across sets of CRNs provides an estimate of the numerical standard error (NSE) associated with the approximation of  $\ln L(Y_T|\mu)$ .

Using the same eight data sets used in constructing table 10.1, we once again obtain 100 approximations of  $\ln L(Y_T|\mu)$  for each data set, here setting  $R = 30$ ,  $N = 500$  for the EIS filter. The choice of  $N$  reflects calibration of the NSE of the EIS filter to roughly match that of the bootstrap particle filter set at  $N = 10,000$  given the model specification  $\sigma_v = 1$ ,  $\gamma = 1$ .

Mean likelihood values obtained using the EIS filter closely mimic those reported in table 10.1, and thus are not reported again here. Instead, table 10.2 reports NSEs, along with the NSE obtained using the bootstrap particle filter measured as a percentage of the NSE obtained using the EIS filter.

Note that NSEs once again increase as the measurement density narrows, but not nearly as dramatically as with the bootstrap particle filter: by a factor of 5.6 given  $\sigma_v = 0.5$ , and 4.4 given  $\sigma_v = 1$ . As a result, the relative performance of the bootstrap particle filter deteriorates as the measurement density narrows. With the numerical accuracy of the two filters roughly on par given  $\sigma_v = 1$ ,  $\gamma = 1$ , the NSE of the bootstrap particle filter is 2.59 times that of the EIS filter given  $\sigma_v = 0.5$ ,  $\gamma = 0.01$ , and 3.77 that given  $\sigma_v = 1$ ,  $\gamma = 0.01$ . Given that reductions in NSEs are achieved at the square-root rate of increase in  $N$ , these figures indicate that for  $\sigma_v = 1$ ,  $\gamma = 0.01$ , the bootstrap particle filter can match the numerical accuracy of the EIS filter by setting  $N$  as approximately 142,000. Finally, it bears noting that on the desktop computer used to produce these results, 30 seconds are required to obtain one evaluation of the likelihood function given  $N = 10,000$ ; in turn, 30 seconds is sufficient to yield 200 likelihood evaluations using the EIS filter given  $R = 30$ ,  $N = 500$ .

**TABLE 10.2**  
NSE Comparisons, Bootstrap Particle Filter  
and EIS Filter

$\sigma_v$		0.5	1
0.01		0.0994, 259	0.0999, 377
0.05		0.0935, 118	0.0959, 165
0.5		0.0213, 104	0.0599, 68
1		0.0178, 63	0.0225, 99

*Note:* Each cell reports the NSE of the EIS filter, along with the NSE of the bootstrap particle filter reported as a percentage of the EIS filter.

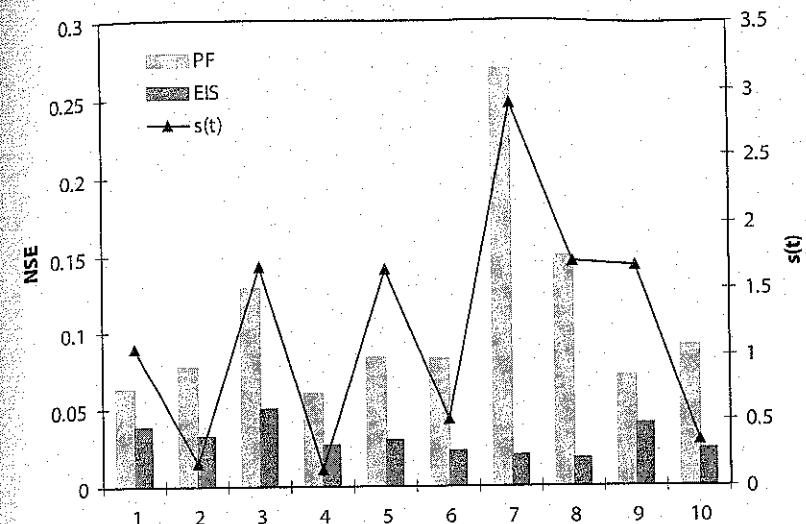


Figure 10.1 NSEs and Corresponding  $s_t$  Given  $\sigma_v = 1$ ,  $\gamma = 0.01$ .

Above we observed that the combination of narrow measurement densities and outlier observations is a particularly challenging scenario for the bootstrap particle filter. Figure 10.1 provides an illustration of this phenomenon by plotting period-by-period NSEs along with corresponding realizations of  $s_t$  obtained under the parameterization  $\sigma_v = 1$ ,  $\gamma = 0.01$ .

Note that while NSEs associated with the EIS filter are relatively uniform over time, those associated with the bootstrap particle filter are volatile, with the large spike in period 7 corresponding with the large increase in  $s_t$  particularly distinct.

### Exercise 10.3

The GAUSS code used to generate the results reported for the EIS filter are available for downloading from the textbook website. Use this code to experiment with alternative parameterizations of the model.

## 10.5 Application to DSGE Models

Here we present implementation of the EIS filter to DSGE models, closely following DeJong et al. (2010). Implementation in this case represents a specialization of the strategy presented in section 4; our approach here is

to map the implementation of DSGE models into the generalized strategy presented above.

The presentation is based on the following assumptions and specialized notation. First, approximations of filtering densities  $\hat{f}(s_t | \mathcal{Y}_t)$  are based on Normal distributions:

$$\hat{f}(s_t | \mathcal{Y}_t) = f_N^n(s_t | \mu_t, \Omega_t). \quad (10.55)$$

Second, measurement densities are also Normally distributed, and are given by

$$f(y_t | s_t, \mathcal{Y}_{t-1}) = f_N^m(y_t | \mu(s_t), \Lambda), \quad (10.56)$$

where  $\mu(s_t)$  denotes an  $m$ -dimensional nonlinear vector function of  $s_t$ . Third, the transition density consists of a nondegenerate density for  $p_t$  (which is  $n_p \times 1$ ) and a nonlinear degenerate transition for  $q_t$  (which is  $n_q \times 1$ , with  $n_p + n_q = n$ ). The former is based on a linear mapping from  $s_{t-1}$  to  $p_t$ ; the latter depends only on  $s_{t-1}$ :

$$f(p_t | s_{t-1}, \mathcal{Y}_{t-1}) = f_N^{n_p}(p_t | B \cdot s_{t-1}, \Sigma), \quad (10.57)$$

$$q_t = \phi(p_{t-1}, q_{t-1}). \quad (10.58)$$

As in section 4.1,  $\phi$  is assumed to be invertible; in this case, its inverse and Jacobian are denoted by

$$q_{t-1} = \psi(q_t, p_{t-1}), \quad J(q_t, p_{t-1}) = \left\| \frac{\partial}{\partial q_t} \psi(q_t, p_{t-1}) \right\|. \quad (10.59)$$

We emphasize that implementation of the EIS filter is not constrained to models that fit these assumptions. Instead, deviations merely entail an alternative approach to implementation than that outlined here. That said, these assumptions encompass a wide range of DSGE models.

Recall the nested likelihood integral pertaining to the case in which transition densities are partially degenerate, repeated here for convenience:

$$f(y_t | \mathcal{Y}_{t-1}) = \int \int f(y_t | s_t, \mathcal{Y}_{t-1}) \cdot J(s_t, p_{t-1}) [f(p_t | s_{t-1}, \mathcal{Y}_{t-1}) \\ \cdot \hat{f}(s_{t-1} | \mathcal{Y}_{t-1})] \Big|_{q_{t-1}=\psi(s_t, p_{t-1})} ds_t dp_{t-1}. \quad (10.60)$$

Under the foregoing assumptions, the likelihood integrand specializes to

$$\varphi_t(p_{t-1}, s_t) = f_N^m(y_t | \mu(s_t), V_y) \cdot J(q_t, p_{t-1}) \cdot [f_N^{n_p}(p_t | B s_{t-1}, \Sigma) \\ \cdot f_N^n(s_{t-1} | \mu_{t-1}, \Omega_{t-1})] \Big|_{q_{t-1}=\psi(s_t, p_{t-1})}. \quad (10.61)$$

Its associated EIS sampler is given by  $g(p_{t-1}, s_t | \hat{\alpha}_t)$ , which is initialized as  $g(p_{t-1}, s_t | \hat{\alpha}_t^0)$ .

### 10.5.1 Initializing the Importance Sampler

Initialization is based on the local linear approximations

$$\hat{\mu}(s_t) = C + D s_t, \quad (10.62)$$

$$\hat{\psi}(q_t, p_{t-1}) = \psi_0 + \psi_q q_t + \psi_p p_{t-1}, \quad (10.63)$$

which obtain by first-order Taylor Series expansions around the base points

$$\bar{s}_{t-1} = \mu_{t-1}, \quad \bar{p}_t = B \mu_{t-1} \quad \text{and} \quad \bar{q}_t = \phi(\bar{s}_{t-1}). \quad (10.64)$$

As in section 4.2, the dependence of  $C$ ,  $D$ , and so forth on time  $t$  is suppressed for ease of notation.

The construction of  $g(p_{t-1}, s_t | \hat{\alpha}_t^0)$  is initiated once again by combining the filtering and state-transition densities, yielding

$$\begin{aligned} \hat{f}\left(\begin{array}{c} p_t \\ s_{t-1} \end{array}\right) &= f_N^{n_p+n}\left(\begin{array}{c} p_t \\ s_{t-1} \end{array} \middle| \begin{bmatrix} B\mu \\ \mu \end{bmatrix}, \begin{bmatrix} \Sigma + B\Omega B' & B\Omega \\ \Omega B' & \Omega \end{bmatrix}\right) \quad (10.65) \\ &\equiv f_N^{n_p+n}\left(\begin{array}{c} p_t \\ s_{t-1} \end{array} \middle| \tilde{\mu}^*, \tilde{\Omega}^*\right), \end{aligned}$$

where  $(\tilde{\mu}^*, \tilde{\Omega}^*)$  are defined following the definitions of  $(\tilde{\mu}, \tilde{\Omega})$  in (10.49) and (10.50) (note that in this case  $A = 0$ ).

Relative to the steps outlined in section 4.2, the next step in this case is novel: we must transform (10.65) into a distribution for  $(s_t, p_{t-1})$ . This is accomplished by application of the linear transformation (10.63). Specifically, (10.63) implies

$$\begin{aligned} \left(\begin{array}{c} p_t \\ s_{t-1} \end{array}\right) &\equiv \left(\begin{array}{c} p_t \\ p_{t-1} \\ q_{t-1} \end{array}\right) = \left(\begin{array}{c} 0 \\ 0 \\ \psi_0 \end{array}\right) + \left(\begin{array}{ccc} I_{n_p} & 0 & 0 \\ 0 & 0 & I_{n_q} \\ 0 & \psi_q & \psi_p \end{array}\right) \cdot \left(\begin{array}{c} p_t \\ q_t \\ p_{t-1} \end{array}\right) \\ &\equiv \gamma + \Theta^{-1} \cdot \left(\begin{array}{c} s_t \\ p_{t-1} \end{array}\right), \\ \gamma &= \left(\begin{array}{c} 0 \\ 0 \\ \psi_0 \end{array}\right), \quad \Theta^{-1} = \left(\begin{array}{ccc} I_{n_p} & 0 & 0 \\ 0 & 0 & I_{n_q} \\ 0 & \psi_q & \psi_p \end{array}\right), \quad \|\Theta^{-1}\| = \psi_q. \end{aligned}$$

Application of this transformation to (10.65) yields

$$\begin{aligned} \hat{f}\left(\begin{array}{c} s_t \\ p_{t-1} \end{array}\right) &= \|\Theta^{-1}\| f_N^{n_p+n}\left(\begin{array}{c} s_t \\ p_{t-1} \end{array} \middle| \tilde{\mu}, \tilde{\Omega}\right), \quad (10.66) \\ \tilde{\mu} &= \Theta(\tilde{\mu}^* - \gamma), \quad \tilde{\Omega} = \Theta \tilde{\Omega}^* \Theta'. \end{aligned}$$

This approximation of the state-transition density is the analogue to that of (10.48) in section 4.2.

The remaining steps proceed exactly as described above. We next construct

$$\widehat{f}\left(y_t \mid \begin{matrix} s_t \\ p_{t-1} \end{matrix}\right) = f_N^m(y_t \mid C + \tilde{D} \begin{bmatrix} s_t \\ p_{t-1} \end{bmatrix}, \Lambda),$$

$$\tilde{D} = \begin{bmatrix} D & \underbrace{0}_{m \times n_p} \end{bmatrix},$$

where  $(C, D)$  are now defined in (10.62), and combine this density with (10.66) to obtain

$$\begin{aligned} \widehat{f}\left(\begin{matrix} y_t \\ s_t \\ p_{t-1} \end{matrix}\right) &= \|\Theta^{-1}\| f_N^{m+n+n_p} \left( \begin{matrix} y_t \\ s_t \\ p_{t-1} \end{matrix} \middle| \begin{bmatrix} C + \tilde{D}\tilde{\mu} \\ \tilde{\mu} \\ \Lambda + \tilde{D}\tilde{\Omega}\tilde{D}' \end{bmatrix}, \begin{bmatrix} \tilde{\Omega} & \tilde{D}\tilde{\Omega}' \\ \tilde{\Omega}\tilde{D}' & \tilde{\Omega} \end{bmatrix} \right) \\ &= \|\Theta^{-1}\| f_N^{m+n+n_p} \left( \begin{matrix} y_t \\ s_t \\ p_{t-1} \end{matrix} \middle| m, \Gamma \right), \end{aligned} \quad (10.67)$$

where  $(m, \Gamma)$  are as defined in (10.52) and (10.53).

Finally, divide  $[y_t \ s_t \ p_{t-1}]'$  into  $(y_t \ s_t \ p_{t-1})'$ , and partition  $m$  and  $\Gamma$  conformably with this division; that is,

$$m = \begin{bmatrix} \underbrace{m_1}_{m \times 1} \\ \underbrace{m_2}_{n+n_p \times 1} \end{bmatrix}, \quad \Gamma = \begin{bmatrix} \underbrace{\Gamma_{11}}_{m \times m} & \underbrace{\Gamma_{12}}_{m \times n+n_p} \\ \underbrace{\Gamma_{21}}_{n+n_p \times m} & \underbrace{\Gamma_{22}}_{n+n_p \times n+n_p} \end{bmatrix}.$$

Then the initialization we seek is the conditional distribution of  $(s_t, p_{t-1})$  given  $y_t$ , which is constructed from the joint distribution of  $[y_t \ s_t \ p_{t-1}]'$  in (10.67):

$$\begin{aligned} \widehat{f}\left(\begin{matrix} s_t \\ p_{t-1} \end{matrix} \mid Y_t\right) &= \|\Theta^{-1}\| f_N^{n+n_p} \left( \begin{matrix} s_t \\ p_{t-1} \end{matrix} \middle| m_{2,1} + \Delta_{21} \cdot y_t, \Gamma_{22,1} \right) \\ &\equiv g_t(p_{t-1}, s_t | \widehat{a}_t^0), \end{aligned} \quad (10.68)$$

$$\begin{aligned} \Delta_{21} &= \Gamma_{21}\Gamma_{11}^{-1}, \quad \Gamma_{22,1} = \Gamma_{22} - \Gamma_{21}\Gamma_{11}^{-1}\Gamma_{12}, \\ m_{2,1} &= m_2 - \Delta_{21}m_1. \end{aligned}$$

Given the optimized sampler

$$g(p_{t-1}, s_t; \widehat{a}_t) = f_N^{n+n_p} \left( \begin{matrix} s_t \\ p_{t-1} \end{matrix} \middle| \mu_{\widehat{a}_t}, \Omega_{\widehat{a}_t} \right),$$

## 10.5 Application to DSGE Models

likelihood approximation is achieved as follows:

$$\widehat{f}(y_t | Y_{t-1}) = \frac{1}{N} \sum_{i=1}^N \omega_t(p_{t-1}^i, s_t^i; \widehat{a}_t), \quad (10.69)$$

$$\omega_t(p_{t-1}, s_t; \widehat{a}_t) = \frac{\varphi_t(p_{t-1}, s_t)}{g_t(p_{t-1}, s_t | \widehat{a}_t)}, \quad (10.70)$$

where  $\varphi_t(p_{t-1}, s_t)$  is defined in (10.61). Then in light of its specification as a Normal density, the period- $t$  filtering density is given by

$$\widehat{f}(s_t) = f_N^n(s_t | \mu, \Omega),$$

where  $\mu$  is defined as the first  $n$  elements of  $\mu_{\widehat{a}_t}$ , and  $\Omega$  is defined as the first  $n \times n$  elements of  $\Omega_{\widehat{a}_t}$ .

### 10.5.2 Initializing the Filtering Density

There are two alternatives for constructing the approximation  $\widehat{f}(s_0)$ , which is used to initialize likelihood construction by representing the period-0 filtering density. If  $s_t$  represents state variables measured as deviations or logged deviations from steady states, it is convenient to work with a linear or log-linear approximation of the associated DSGE model. Recall that such approximations are of the form

$$\begin{aligned} x_t &= Fx_{t-1} + Gv_t \\ &= Fx_{t-1} + e_t. \end{aligned}$$

With  $Ee_t e_t' = Q$ , the variance-covariance matrix of  $x_t$ , denoted as  $\Gamma(0)$ , is given by

$$\text{vec}[\Gamma(0)] = [I - F \otimes F]^{-1} \text{vec}[Q],$$

where  $\otimes$  denotes the Kronecker product. Then with  $\Omega_0$  denoting the submatrix of  $\Gamma(0)$  appropriate given the location of  $s_t$  in  $x_t$ , the approximation we seek is simply

$$\widehat{f}(s_0) = f_N^n(s_0 | 0, \Omega_0).$$

If instead  $s_t$  represents state variables measured in terms of levels, it may be more convenient to work with a nonlinear model representation of the form

$$s_t = f(s_{t-1}, v_t). \quad (10.71)$$

As discussed in section 2.1 of chapter 9, the unconditional mean and covariance matrix of  $s_t$  can be approximated using sample statistics calculated

from  $\{s_t\}_{t=1}^T$ , generated from (10.71) via direct simulation. Simulation is initialized by an arbitrary specification  $s_0$ , whose influence is eliminated using a burn-in stage.

### 10.5.3 Application to the RBC Model

We demonstrate implementation of the EIS filter using the RBC model introduced in section 1 of chapter 3. This is precisely the model used by Fernandez-Villaverde and Rubio-Ramirez (2005) to demonstrate implementation of the particle filter. Recall that the model is given by

$$\left(\frac{1-\varphi}{\varphi}\right) \frac{c_t}{l_t} = (1-\alpha)z_t \left(\frac{k_t}{n_t}\right)^\alpha \quad (10.72)$$

$$c_t^\kappa l_t^\lambda = \beta E_t \left\{ c_{t+1}^\kappa l_{t+1}^\lambda \left[ \alpha z_{t+1} \left(\frac{n_{t+1}}{k_{t+1}}\right)^{1-\alpha} + (1-\delta) \right] \right\} \quad (10.73)$$

$$\zeta_t = z_t k_t^\alpha n_t^{1-\alpha} \quad (10.74)$$

$$\xi_t = c_t + i_t \quad (10.75)$$

$$k_{t+1} = i_t + (1-\delta)k_t \quad (10.76)$$

$$l = n_t + l_t \quad (10.77)$$

$$\ln z_t = (1-\rho) \ln (\bar{z}) + \rho \ln z_{t-1} + \varepsilon_t, \quad (10.78)$$

where

$$\kappa = \varphi(1-\phi) - 1$$

and

$$\lambda = (1-\varphi)(1-\phi).$$

We proceed by defining  $s_t = (\ln k_t \ \ln z_t)'$ ,  $v_t = \varepsilon_t$ ,  $p_t = \ln z_t$ ,  $q_t = \ln k_t$ . Given policy functions for the control variables, the model is represented in state-space form using the state-transition equations

$$\ln z_t = (1-\rho) \ln (z_0) + \rho \ln z_{t-1} + \varepsilon_t, \quad (10.79)$$

$$e^{\ln k_{t+1}} = i_t (\ln k_t, \ln z_t) + (1-\delta)e^{\ln k_{t+1}}. \quad (10.80)$$

Then working with  $y_t = (\zeta_t \ i_t \ n_t)'$  as the observables, the observation equations are given by

$$\zeta_t = e^{\ln z_t} e^{\alpha \ln k_t} n_t (\ln k_t, \ln z_t)^{1-\alpha} + u_{\zeta t} \quad (10.81)$$

$$i_t = i_t (\ln k_t, \ln z_t) + u_{it} \quad (10.82)$$

$$n_t = n_t (\ln k_t, \ln z_t) + u_{nt}, \quad (10.83)$$

### 10.5 Application to DSGE Models

with  $u_t \sim N(0, \Lambda)$ . Note that (10.79)–(10.80) are the specializations of (10.57)–(10.58) implied by this model, and the linear approximation of (10.80) yields (10.63). Likewise, (10.81)–(10.83) are the specializations of (10.56), and the linear approximation of these equations yields (10.62). Finally, returning to (10.80), its inversion yields

$$\ln k_{t-1} = \psi(\ln k_t, \ln z_{t-1}), \quad J(\ln k_t, \ln z_{t-1}) = \left\| \frac{\partial}{\partial \ln k_t} \psi(\ln k_t, \ln z_{t-1}) \right\|, \quad (10.84)$$

which represents the specialization of (10.59) implied by the model.

To demonstrate the performance of the EIS filter in this setting, we conduct a Monte Carlo experiment using an artificial data set consisting of 100 realizations of  $\{\zeta_t, i_t, n_t\}$  generated from the model. The data set was constructed by Fernandez-Villaverde and Rubio-Ramirez (2005) under the model parameterization presented in table 10.3. Note from the table that the standard deviations of the measurement errors ( $\sigma_\zeta, \sigma_i, \sigma_n$ ) are small relative to  $\sigma_\varepsilon$ , which as we have seen can lead to problems associated with sample impoverishment. The data set is available for downloading at <http://qed.econ.queensu.ca/jae/2005-v20.7/fernandez-rubio/>.

For experiments involving additional data sets and DSGE models, see DeJong et al. (2010).

The experiment we conduct follows those associated with tables 10.1 and 10.2. Using 100 alternative sets of CRNs, we obtain 100 log-likelihood approximations; the standard deviation of these approximations calculated across CRNs is the NSE associated with the filter. We conduct the experiment for both the bootstrap particle filter and the EIS filter; the former is implemented using  $R = 50$ ,  $N = 100$ ; the latter is implemented using  $N = 60,000$ , following Fernandez-Villaverde and Rubio-Ramirez. Sample means of log-likelihoods, along with RNEs, are reported in table 10.4.

The difference in NSEs in this example is large: its elimination implies the need to specify particle swarm sizes on the order of  $8.4 \times 10^{15}$  under the bootstrap particle filter. Note also the nontrivial discrepancy in sample means obtained in this application. This discrepancy does not reflect bias

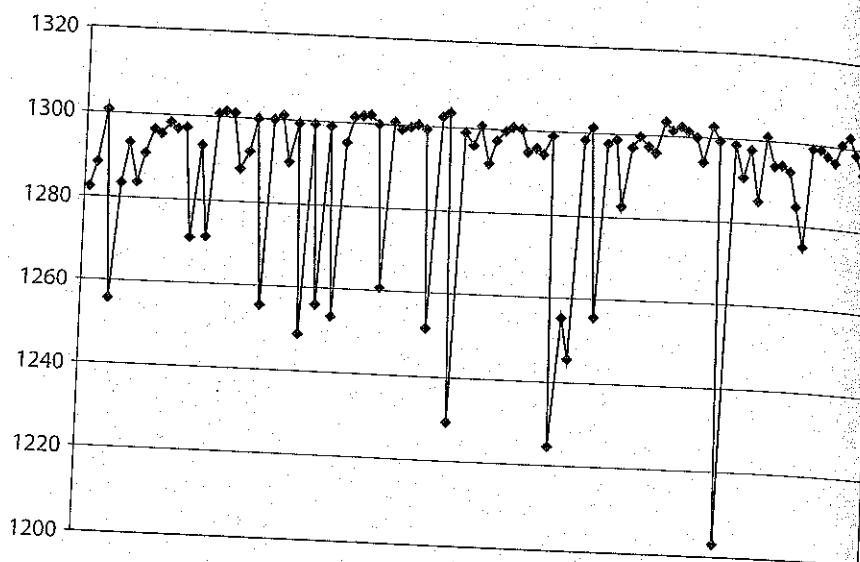
TABLE 10.3  
Parameter Values, RBC Model

$\alpha$	$\beta$	$\phi$	$\varphi$	$\delta$	$\rho$	$\sigma_\varepsilon$	$\sigma_\zeta$	$\sigma_i$	$\sigma_n$
0.4	0.99	2	0.357	0.02	0.95	0.007	0.000158	0.000866	0.0011

TABLE 10.4

Sample Means and NSE Comparisons, Bootstrap Particle Filter and EIS Filter

	Sample mean	NSE
Bootstrap	1,289.452	19.023
EIS	1,300.067	0.000051

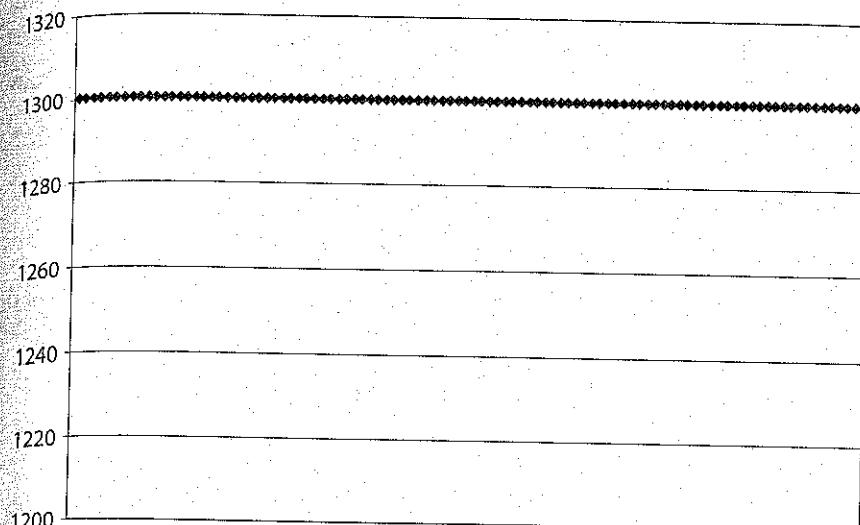


10.2a Bootstrap Particle Filter

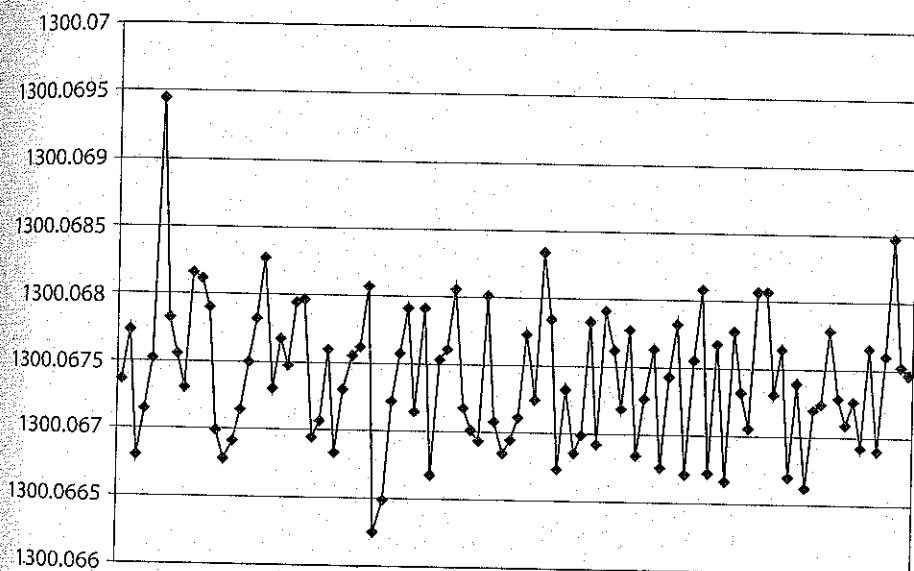
Figure 10.2 Log-Likelihood Approximations

in either filter. Rather, it reflects a tendency of the bootstrap particle filter to provide very low log-likelihood estimates in certain cases, a problem that can be remedied by increasing  $N$ . This tendency is illustrated in figure 10.2, which plots log-likelihood estimates obtained across CRNs using both filters.

We conclude this application by illustrating a second advantage of implementing the EIS filter: given the use of CRNs, the likelihood approximations it produces are continuous functions of model parameters. To demonstrate this, figure 10.3 illustrates likelihood surfaces produced by allowing each model parameter to vary individually above and below its actual value, holding all additional parameters fixed at their true values.



10.2b EIS Filter



10.2c EIS Filter, Rescaled

Figure 10.2 Continued

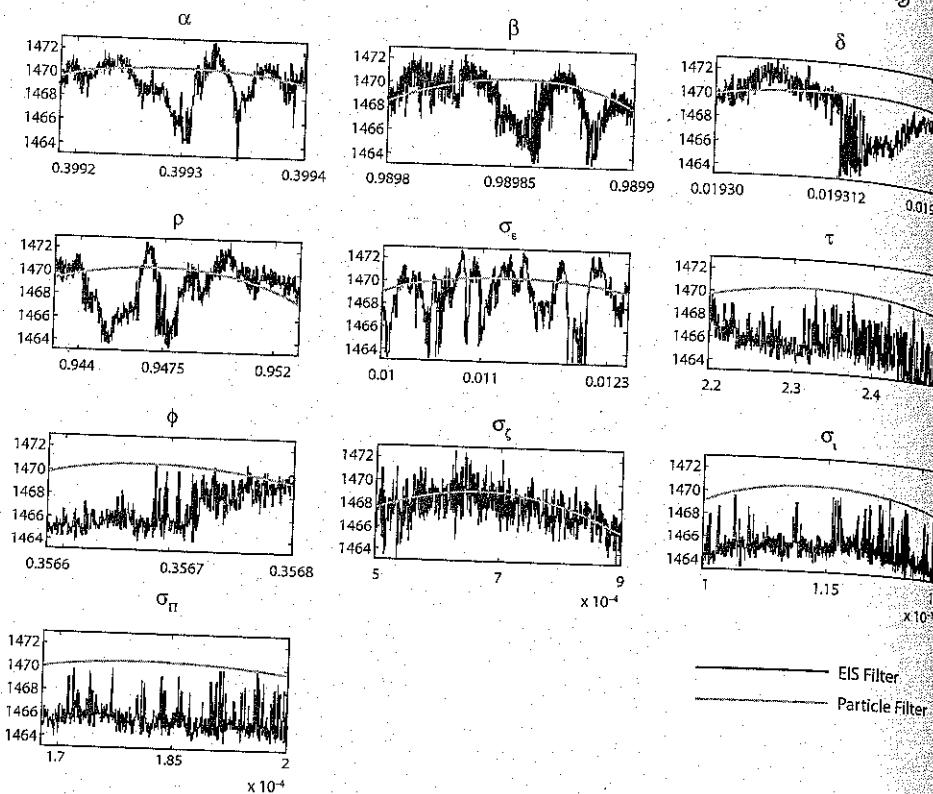


Figure 10.3 Likelihood Contours

This is done using both filters, each implemented using CRNs. Regarding the labeling of parameters,  $\tau$  is the notation Fernandez-Villaverde and Rubio-Ramirez used for the CRRA parameter we denote as  $\phi$ , and  $\phi$  is the notation for the share parameter we denote as  $\varphi$ .

We consider the following exercise to be most important among all those included in this text. We hope that the foregoing results have provided sufficient inspiration for attempting its execution, and that the details regarding implementation presented here are sufficient for ensuring a successful attempt.

#### Exercise 10.4

Replicate the results reported for the EIS filter in table 10.4.

## Part V

### Empirical Methods

# *Chapter 11*

---

## *Calibration*

Models are to be used, not believed.

—Henri Theil, *Principles of Econometrics*

### **11.1 Historical Origins and Philosophy**

With their seminal analysis of business cycles, Kydland and Prescott (1982) capped a paradigm shift in the conduct of empirical work in macroeconomics. They did so using a methodology that enabled them to cast the DSGE model they analyzed as the centerpiece of their empirical analysis. The analysis contributed towards the Nobel Prize in Economics they received in 2004, and the methodology has come to be known as a calibration exercise.<sup>1</sup> Calibration not only remains a popular tool for analyzing DSGEs, but has also served as the building block for subsequent methodologies developed towards this end. Thus it provides a natural point of departure for our presentation of these methodologies.

Although undoubtedly an empirical methodology, calibration is distinct from the branch of econometrics under which theoretical models are represented as complete probability models that can be estimated, tested, and used to generate predictions using formal statistical procedures. Haavelmo (1944) provided an early and forceful articulation of this latter approach to econometrics, and in 1989 received the Nobel Prize in Economics “for his clarification of the probability theory foundations of econometrics and his analyses of simultaneous economic structures” (Bank of Sweden, 1989). Henceforth, we will refer to this as the probability approach to econometrics.

Regarding Haavelmo’s “analyses of simultaneous economic structures,” otherwise known as systems-of-equations models, at the time of his work this was the most sophisticated class of structural models that could be subjected to formal empirical analysis. As characterized by Koopmans (1949), such models include equations falling into one of four classes: identities

<sup>1</sup>Specifically, Kydland and Prescott received the Nobel Prize “for their contributions to dynamic macroeconomics: the time consistency of economic policy and the driving forces behind business cycles” (Bank of Sweden, 2004).

(e.g., the national income accounting identity), institutional rules (e.g., tax rates), technology constraints (e.g., a production function), and behavioral equations (e.g., a consumption function relating consumption to disposable income). For a textbook presentation of systems-of-equations models, see Sargent (1987a).

This latter class of equations distinguishes systems-of-equations models from DSGE models. Behavioral equations cast endogenous variables as *ad hoc* functions of additional variables included in the model. An important objective in designing these specifications is to capture relationships between variables observed from historical data. Indeed, econometric implementations of such models proceed under the assumption that the parameters of the behavioral equations are fixed, and thus may be estimated using historical data; the estimated models are then used to address quantitative questions. Such analyses represented state-of-the-art practice in econometrics into the 1970s.

In place of behavioral equations, DSGE models feature equations that reflect the pursuit of explicit objectives (e.g., the maximization of lifetime utility) by purposeful decision makers (e.g., representative households). For example, the RBC model presented in section 1 of chapter 3 features two such equations: the intratemporal optimality condition that determines the labor-leisure trade-off, and the intertemporal optimality condition that determines the consumption-investment trade-off. The parameters in these equations reflect either the preferences of the decision maker (e.g., discount factors, intertemporal elasticities of substitution, etc.) or features of their environment (e.g., capital's share of labor in the production technology).

Two important developments ultimately ended the predominance of the systems-of-equations approach. The first was empirical: systems-of-equations models suffered "spectacular predictive failure" in the policy guidance they provided during the episode of stagflation experienced during the early 1970s (Kydland and Prescott, 1991a, 166). Quoting Lucas and Sargent (1979):

In the present decade, the U.S. economy has undergone its first major depression since the 1930's, to the accompaniment of inflation rates in excess of 10 per cent per annum. . . . These events . . . were accompanied by massive government budget deficits and high rates of monetary expansion, policies which, although bearing an admitted risk of inflation, promised according to modern Keynesian doctrine rapid real growth and low rates of unemployment. That these predictions were wildly incorrect and that the doctrine on which they were based is fundamentally flawed are now simple matters of fact, involving no novelties in economic theory. (1)

The second development was theoretical: the underlying assumption that the parameters of behavioral equations in such models are fixed was

### 11.1 Historical Origins and Philosophy

recognized as being inconsistent with optimizing behavior on the part of purposeful decision makers. This is the thrust of Lucas' (1976) critique of policy evaluation based on systems-of-equations models:

Given that the structure of an econometric model consists of optimal decision rules of economic agents, and that optimal decision rules vary systematically with changes in the structure of series relevant to the decision maker, it follows that any change in policy will systematically alter the structure of econometric models. (41)

Or as summarized by Lucas and Sargent (1979):

The casual treatment of expectations is not a peripheral problem in these models, for the role of expectations is pervasive in them and exerts a massive influence on their dynamic properties. . . . The failure of existing models to derive restrictions on expectations from any first principles grounded in economic theory is a symptom of a deeper and more general failure to derive behavioral relationships from any consistently posed dynamic optimization problem. . . . There are, therefore, . . . theoretical reasons for believing that the parameters identified as structural by current macroeconomic methods are not in fact structural. That is, we see no reason to believe that these models have isolated structures which will remain invariant across the class of interventions that figure in contemporary discussions of economic policy. (5-6)

In consequence, Lucas (1976) concludes that "simulations using these models can, in principle, provide no useful information as to the actual consequences of alternative economic policies" (20). In turn, again referring to the predictive failure of these models during the stagflation episode of the 1970s, Lucas and Sargent (1979) conclude that "the difficulties are fatal: that modern macroeconomic models are of *no* value in guiding policy and that this condition will not be remedied by modifications along any line which is currently being pursued" (2).

Two leading reactions to these developments ensued. First, remaining in the tradition of the probability approach to econometrics, the methodological contributions of Sims (1972) and Hansen and Sargent (1980) made possible the imposition of theoretical discipline on reduced-form models of macroeconomic activity. DSGE models featuring rational decision makers provided the source of this discipline, and the form of the discipline most commonly took the form of "cross-equation restrictions" imposed on vector autoregressive (VAR) models. This development represents an intermediate step towards the implementation of DSGE models in empirical applications, because reduced-form models serve as the focal point of such analyses. Moreover, early empirical applications spawned by this development proved to be disappointing, because rejections of particular parametric implementations of the restrictions were commonplace

(e.g., see Hansen and Sargent, 1981; Hansen and Singleton 1982, 1983, and Eichenbaum, 1983). The second reaction was the development of the modern calibration exercise.

In place of estimation and testing, the goal in a calibration exercise is to use a parameterized structural model to address a specific quantitative question. The model is constructed and parameterized subject to the constraint that it mimic features of the actual economy that have been identified a priori. Questions fall under two general headings. They may involve the ability of the model to account for an additional set of features of the actual economy; that is, they may involve assessments of fit. Alternatively, they may involve assessments of the theoretical implications of changes in economic policy. This characterization stems from Kydland and Prescott (1991a, 1996), who traced the historical roots of the use of calibration exercises as an empirical methodology, and outlined their view of what such exercises entail.

Kydland and Prescott (1991a) identify calibration as embodying the approach to econometrics articulated and practiced by Frisch (1933a, 1933b). Regarding articulation, this is provided by Frisch's (1933a) editorial opening the inaugural issue of the flagship journal of the Econometric Society: *Econometrica*. As stated in its constitution, the main objective of the Econometric Society is to

promote studies that aim at a unification of the theoretical-quantitative and empirical-quantitative approach to economic problems and that are penetrated by constructive and rigorous thinking.... Any activity which promises ultimately to further such unification of theoretical and factual studies in economics shall be within the sphere of interest of the Society. (1)

Such studies personified Frisch's vision of econometrics: "This mutual penetration of quantitative economic theory and statistical observation is the essence of econometrics" (2).

Of course, this vision was also shared by the developers of the probability approach to econometrics. To quote Haavelmo (1944): "The method of econometric research aims, essentially, at a conjunction of economic theory and actual measurements, using the theory and technique of statistical inference as a bridge pier" (iii). However, in practice Frisch pursued this vision without strict adherence to the probability approach: for example, his (1933b) analysis of the propagation of business-cycle shocks was based on a production technology with parameters calibrated on the basis of micro data. This work contributed towards the inaugural Nobel Prize in Economics he shared with Jan Tinbergen in 1969:

Let me take, as an example, Professor Frisch's pioneer work in the early thirties involving a dynamic formulation of the theory of cycles. He demonstrated

how a dynamic system with difference and differential equations for investments and consumption expenditure, with certain monetary restrictions, produced a damped wave movement with wavelengths of 4 and 8 years. By exposing the system to random disruptions, he could demonstrate also how these wave movements became permanent and uneven in a rather realistic manner. Frisch was before his time in the building of mathematical models, and he has many successors. The same is true of his contribution to methods for the statistical testing of hypotheses. (Lundberg, 1969)

In their own analysis of business cycles, Kydland and Prescott (1982) eschewed the probability approach in favor of a calibration experiment that enabled them to cast the DSGE model they analyzed as the focal point of their empirical analysis. (Tools for working with general-equilibrium models in static and nonstochastic settings had been developed earlier by Shoven and Whalley, 1972; and Scarf and Hansen, 1973.) It is tempting to view this as a decision made due to practical considerations, because formal statistical tools for implementing DSGE models empirically had yet to be developed. However, an important component of Kydland and Prescott's advocacy of calibration is based on a criticism of the probability approach. For example, writing with specific reference to calibration exercises involving real business cycle models, Prescott (1986) makes the case as follows:

The models constructed within this theoretical framework are necessarily highly abstract. Consequently, they are necessarily false, and statistical hypothesis testing will reject them. This does not imply, however, that nothing can be learned from such quantitative theoretical exercises. (10)

A similar sentiment was expressed earlier by Lucas (1980): "Any model that is well enough articulated to give clear answers to the questions we put to it will necessarily be artificial, abstract, patently 'unreal'" (696). As another example, in discussing Frisch's (1970) characterization of the state of econometrics, Kydland and Prescott (1991a) offer the following observation:

In this review (Frisch) discusses what he considers to be "econometric analysis of the genuine kind" (p. 163), and gives four examples of such analysis. None of these examples involve the estimation and statistical testing of some model. None involve an attempt to discover some true relationship. All use a model, which is an abstraction of a complex reality, to address some clear-cut question or issue. (162)

In sum, the use of calibration exercises as a means for facilitating the empirical implementation of DSGE models arose in the aftermath of the

demise of systems-of-equations analyses. Estimation and testing are purposefully de-emphasized under this approach, yet calibration exercises are decidedly an empirical tool, in that they are designed to provide concrete answers to quantitative questions. We now describe their implementation.

## 11.2 Implementation

Enunciations of the specific methodology advocated by Kydland and Prescott for implementing calibration exercises in applications to DSGE models are available in a variety of sources (e.g., Kydland and Prescott 1991a, 1996; Prescott 1986, 2006; and Cooley and Prescott 1995). Here we begin by outlining the five-step procedure presented by Kydland and Prescott (1996). We then discuss its implementation in the context of the notation established in Part I of this book.

The first step is to pose a question, which will fall under one of two general headings: questions may involve assessments of the theoretical implications of changes in policy (e.g., the potential welfare gains associated with a given tax reform), or they may involve assessments of the ability of a model to mimic features of the actual economy. Kydland and Prescott characterize the latter class of questions as follows:

Other questions are concerned with the testing and development of theory. These questions typically ask about the quantitative implications of theory for some phenomena. If the answer to these questions is that the predictions of theory match the observations, theory has passed that particular test. If the answer is that there is a discrepancy, a deviation from theory has been documented. (70–71)

The second step is to use “well-tested theory” to address the question: “With a particular question in mind, a researcher needs some strong theory to carry out a computational experiment: that is, a researcher needs a theory that has been tested through use and found to provide reliable answers to a class of questions” (72). This step comes with a caveat: “We recognize, of course, that although the economist should choose a well-tested theory, every theory has some issues and questions that it does not address well” (72).

Of course, this caveat does not apply exclusively to calibration exercises: as simplifications of reality, all models suffer empirical shortcomings along certain dimensions, and any procedure that enables their empirical implementation must be applied with this problem in mind. The point here is that the chosen theory must be suitably developed along the dimensions of relevance to the question at hand. To take the example offered by

Kydland and Prescott: “In the case of neoclassical growth theory . . . it fails spectacularly when used to address economic development issues. . . . This does not preclude its usefulness in evaluating tax policies and in business cycle research.” (72).

The third step involves the construction of the model economy: “With a particular theoretical framework in mind, the third step in conducting a computational experiment is to construct a model economy. Here, key issues are the amount of detail and the feasibility of computing the equilibrium process” (72). Regarding this last point, the more detailed and complex is a given model, the harder it is to analyze; thus in the words of Solow (1956): “The art of successful theorizing is to make the inevitable simplifying assumptions in such a way that the final results are not very sensitive” (65). So in close relation with step 2, the specific model chosen for analysis is ideally constructed to be sufficiently rich for addressing the question at hand without being unnecessarily complex. For example, Rowenhorst (1991) studied versions of an RBC model with and without the “time-to-build” feature of the production technology included in the model analyzed by Kydland and Prescott (1982). (Under “time to build,” current investments yield productive capital only in future dates.) His analysis demonstrated that the time-to-build feature was relatively unimportant in contributing to the propagation of technology shocks; today, time-to-build rarely serves as a central feature of RBC models.

Beyond ease of analysis, model simplicity has an additional virtue: it is valuable for helping to disentangle the importance of various features of a given specification for generating a particular result. Consider the simultaneous inclusion of a set of additional features to a baseline model with known properties. Given the outcome of an interesting modification of the model’s properties, the attribution of importance to the individual features in generating this result is at minimum a significant challenge. In contrast, analysis of the impact of the individual features in isolation or in smaller subsets is a far more effective means of achieving attribution. It may turn out that each additional feature is necessary for achieving the result; alternatively, certain features may turn out to be unimportant and usefully discarded.

These first three steps apply quite broadly to empirical applications; the fourth step, the calibration of model parameters, does not:

Generally, some economic questions have known answers, and the model should give an approximately correct answer to them if we are to have any confidence in the answer given to the question with unknown answer. Thus, data are used to calibrate the model economy so that it mimics the world as closely as possible along a limited, but clearly specified, number of dimensions. (Kydland and Prescott, 1996, 74)

Upon offering this definition, calibration is then distinguished from estimation: "Note that calibration is not an attempt at assessing the size of something; it is not estimation" (74). Moreover:

It is important to emphasize that the parameter values selected are not the ones that provide the best fit in some statistical sense. In some cases, the presence of a particular discrepancy between the data and the model economy is a test of the theory being used. In these cases, absence of that discrepancy is grounds to reject the use of the theory to address the question. (74)

Although in general this definition implies no restrictions on the specific dimensions of the world used to pin down parameter values, certain dimensions have come to be applied rather extensively in a wide range of applications. Long-run averages such as the share of output paid to labor, and the fraction of available hours worked per household, both of which have been remarkably stable over time, serve as primary examples. In addition, empirical results obtained in micro studies conducted at the individual or household level are often used as a means of pinning down certain parameters. For example, the time-allocation study of Ghez and Becker (1975), conducted using panel data on individual allocations of time to market activities, was used by Cooley and Prescott (1995) to determine the relative weights assigned to consumption and leisure in the instantaneous utility function featured in the RBC model they analyzed.

The final step is to run the experiment. Just how this is done depends upon the question at hand, but at a minimum this involves the solution of the model, for example, as outlined in chapter 4. The model solution yields a state-space representation of the form

$$x_{t+1} = F(\mu)x_t + G(\mu)v_{t+1} \quad (11.1)$$

$$X_t = H(\mu)'x_t \quad (11.2)$$

$$E(e_t e_t') = G(\mu)E(v_t v_t')G(\mu)' \equiv Q(\mu). \quad (11.3)$$

Recall that  $x_t$  represents the full set of variables included in the model, represented as deviations from steady state values;  $X_t$  represents the collection of associated observable variables, represented analogously; and  $\mu$  contains the structural parameters of the model.

At this point it is useful to distinguish between two versions of  $X_t$ . We will denote model versions as  $X_t^M$ , and the actual data as  $X_t$ . In the context of this notation, the calibration step involves the specification of the individual elements of  $\mu$ . Representing the real-world criteria used to specify  $\mu$  as  $\Omega(\{X_t\}_{t=1}^T)$ , the calibration step involves the choice of  $\mu$  such that

$$\Omega(\{X_t^M\}_{t=1}^T) = \Omega(\{X_t\}_{t=1}^T). \quad (11.4)$$

### 11.3 The Welfare Cost of Business Cycles

If the question posed in the calibration exercise is to compare model predictions with an additional collection of features of the real world, then denoting these additional features as  $\Phi(\{X_t\}_{t=1}^T)$ , the question is addressed via comparison of  $\Phi(\{X_t\}_{t=1}^T)$  and  $\Phi(\{X_t^M\}_{t=1}^T)$ . Depending upon the specification of  $\Phi(\cdot)$ ,  $\Phi(\{X_t^M\}_{t=1}^T)$  may either be calculated analytically or via simulation.

As noted above, Kydland and Prescott characterize exercises of this sort as a means of facilitating the "testing and development of theory" (70). As will be apparent in chapter 12, wherein we present generalized and simulated method-of-moment procedures, such exercises are closely related to moment-matching exercises, which provide a powerful means of estimating models and evaluating their ability to capture real-world phenomena. The only substantive difference lies in the level of statistical formality upon which comparisons are based.

If the question posed in the exercise involves an assessment of the theoretical implications of a change in policy, then once again the experiment begins with the choice of  $\mu$  such that (11.4) is satisfied. In this case, given that  $\mu$  contains as a subset of elements parameters characterizing the nature of the policy under investigation, the policy change will be reflected by a new specification  $\mu'$ , and thus  $X_t^{M'}$ . The question can then be cast in the form of a comparison between  $\Phi(\{X_t^M\}_{t=1}^T)$  and  $\Phi(\{X_t^{M'}\}_{t=1}^T)$ . Examples of both sorts of exercises follow.

### 11.3 The Welfare Cost of Business Cycles

We begin with an example of the first sort, based on an exercise conducted originally by Lucas (1987), and updated by Lucas (2003). The question involves a calculation of the potential gains in welfare that could be achieved through improvements in the management of business cycle fluctuations. In particular, consider the availability of a policy capable of eliminating all variability in consumption, beyond long-term growth. Given risk aversion on the part of consumers, the implementation of such a policy will lead to an improvement in welfare. The question is: to what extent?

A quantitative answer to this question is available from a comparison of the utility derived from consumption streams  $\{c_t^A\}$  and  $\{c_t^B\}$  associated with the implementation of alternative policies  $A$  and  $B$ . Suppose the latter is preferred to the former, so that

$$U(\{c_t^A\}) < U(\{c_t^B\}).$$

To quantify the potential welfare gains to be had in moving from  $A$  to  $B$ , or alternatively, the welfare cost associated with adherence to policy  $A$ , Lucas proposed the calculation of  $\lambda$  such that

$$U((1 + \lambda)c_t^A) = U(c_t^B). \quad (11.5)$$

In this way, welfare costs are measured in units of a percentage of the level of consumption realized under policy  $A$ .

Lucas' implementation of this question entailed a comparison of the expected discounted lifetime utility derived by a representative consumer from two alternative consumption streams: one that mimics the behavior of actual postwar consumption, and one that mimics the deterministic growth pattern followed by postwar consumption. Under the first, consumption is stochastic, and obeys

$$c_t = A e^{\mu t} e^{-\frac{1}{2}\sigma^2} \varepsilon_t, \quad (11.6)$$

where  $\log \varepsilon_t$  is distributed as  $N(0, \sigma^2)$ . Under this distributional assumption for  $\varepsilon_t$ ,

$$E(\varepsilon_t^m) = e^{\frac{1}{2}\sigma^2 m^2}, \quad (11.7)$$

and thus

$$E(c_t) = A e^{\mu t}, \quad (11.8)$$

the growth rate of which is  $\mu$ . Under the second, consumption is deterministic, and at time  $t$  is given by  $A e^{\mu t}$ , as in (11.8).

Modeling the lifetime utility generated by a given consumption stream  $\{c_t\}$  as

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma}}{1-\gamma},$$

where  $\beta$  is the consumer's discount factor and  $\gamma$  measures the consumer's degree of relative risk aversion, Lucas' welfare comparison involves the calculation of  $\lambda$  such that

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{[(1 + \lambda)c_t]^1 - \gamma}{1 - \gamma} = \sum_{t=0}^{\infty} \beta^t \frac{(A e^{\mu t})^{1-\gamma}}{1-\gamma}, \quad (11.9)$$

with the behavior of  $c_t$  given by (11.6). Given (11.7),

$$E_0(\varepsilon_t^{1-\gamma}) = e^{\frac{1}{2}(1-\gamma)^2 \sigma^2},$$

and thus (11.9) simplifies to

$$(1 + \lambda)^{1-\gamma} e^{-\frac{1}{2}\gamma(1-\gamma)\sigma^2} = 1.$$

Finally, taking logs and using the approximation  $\log(1 + \lambda) \approx \lambda$ , the comparison reduces to

$$\lambda = \frac{1}{2}\sigma^2\gamma.$$

Thus the calculation of  $\lambda$  boils down to this simple relationship involving two parameters:  $\sigma^2$  and  $\gamma$ . The former can be estimated directly as the residual variance in a regression of  $\log c_t$  on a constant and trend. Using annual data on real per capita consumption spanning 1947–2001, Lucas (2003) estimated  $\sigma^2$  as  $(0.032)^2$ . Extending these data through 2009 yields an estimate of  $(0.02996)^2$ ; the data used for this purpose are contained in the file `welddat.txt`, available at the textbook website.

Regarding the specification of  $\gamma$ , Lucas (2003) appeals to an intertemporal optimality condition for consumption growth  $\mu$  that is a standard feature of theoretical growth models featuring consumer optimization (e.g., see Barro and Sala-i-Martin, 2004):

$$\mu = \frac{1}{\gamma}(r - \rho),$$

where  $r$  is the after-tax rate of return generated by physical capital and  $\rho$  is the consumer's subjective discount rate. In this context,  $\frac{1}{\gamma}$  represents the consumer's intertemporal elasticity of substitution. With  $r$  averaging approximately 0.05 in postwar data,  $\mu$  estimated as 0.023 in the regression of  $\log c_t$  on a constant and trend, and  $\rho$  restricted to be positive, an upper bound for  $\gamma$  is approximately 2.2, and a value of 1 is often chosen as a benchmark specification.

Table 11.1 reports values of  $\lambda$  calculated for alternative specifications of  $\gamma$ , based on the estimate  $\sigma^2 = (0.02996)^2$ . It also reports the associated (chain-weighted 2005) dollar value of per capita consumption in 2009.

As these figures indicate, potential welfare gains offered by the complete elimination of cyclical fluctuations in consumption are strikingly low: as high as only \$35.70 per person given  $\gamma = 2.5$ , and only \$14.28 given  $\gamma = 1$ . They led Lucas (2003) to two conclusions. First, the figures serve as a tribute to the success of the stabilization policies that have been implemented over the postwar period: the policies have not left much room for improvement. Second, efforts designed to deliver further gains in stabilization have little to offer in generating further improvements in welfare.

**TABLE 11.1**  
Welfare Costs, CRRA Preferences

$\gamma$	$\lambda$	\$/person, 2009
0.5	0.00022	\$7.14
1	0.00045	\$14.28
1.5	0.00067	\$21.42
2	0.00090	\$28.56
2.5	0.00112	\$35.70

Lucas' (1987) analysis prompted an extensive literature that grew out of efforts to analyze the robustness of these figures; a summary of this literature is provided by Lucas (2003). For example, one prominent strand of this literature involves investigations of robustness to departures from the representative-agent framework (e.g., as pursued by Krusell and Smith, 1999). Such departures enable the calculation of differential welfare effects for, for example, low-income households, who may suffer disproportionately from cyclical uncertainty. Another prominent strand involves departures from the CRRA specification chosen for instantaneous utility. Lucas' (2003) reading of this literature led him to conclude that his original calculations have proven to be remarkably robust. Here we demonstrate an extension that underscores this conclusion, adopted from Otrok (2001).

Otrok's extension involves the replacement of the CRRA specification for instantaneous utility with the habit/durability specification introduced in chapter 3. Otrok's analysis was conducted in the context of a fully specified RBC model featuring a labor/leisure trade-off, along with a consumption/investment trade-off. The parameters of the model were estimated using Bayesian methods discussed in chapter 14. Here, we adopt Otrok's estimates of the habit/durability parameters in pursuit of calculations of  $\lambda$  akin to those made by Lucas (1987, 2003) for the CRRA case.

Recall that under habit/durability preferences, instantaneous utility is given by

$$u(h_t) = \frac{h_t^{1-\gamma}}{1-\gamma}, \quad (11.10)$$

with

$$h_t = h_t^d - \alpha h_t^b, \quad (11.11)$$

### 11.3 The Welfare Cost of Business Cycles

where  $\alpha \in (0, 1)$ ,  $h_t^d$  is the household's durability stock, and  $h_t^b$  its habit stock. The stocks are defined by

$$h_t^d = \sum_{j=0}^{\infty} \delta^j c_{t-j} \quad (11.12)$$

$$\begin{aligned} h_t^b &= (1-\theta) \sum_{j=0}^{\infty} \theta^j h_{t-1-j}^d \\ &= (1-\theta) \sum_{j=0}^{\infty} \theta^j \sum_{i=0}^{\infty} \delta^i c_{t-1-i} \end{aligned} \quad (11.13)$$

where  $\delta \in (0, 1)$  and  $\theta \in (0, 1)$ . Substituting for  $h_t^d$  and  $h_t^b$  in (11.11) thus yields

$$\begin{aligned} h_t &= c_t + \sum_{j=1}^{\infty} \left[ \delta^j - \alpha(1-\theta) \sum_{i=0}^{j-1} \delta^i \theta^{j-i-1} \right] c_{t-j} \\ &\equiv \sum_{j=0}^{\infty} \Phi_j c_{t-j}, \end{aligned} \quad (11.14)$$

where  $\Phi_0 \equiv 1$ .

Combining (11.10) and (11.14) with (11.9), Lucas' welfare comparison in this case involves the calculation of  $\lambda$  such that

$$\begin{aligned} E_0 \sum_{t=0}^{\infty} \beta^t &\frac{\left[ (1+\lambda) \left\{ \sum_{j=0}^{\infty} \Phi_j c_{t-j} \right\} \right]^{1-\gamma}}{1-\gamma} \\ &= \sum_{t=0}^{\infty} \beta^t \frac{\left( \sum_{j=0}^{\infty} \Phi_j A e^{\mu t - j} \right)^{1-\gamma}}{1-\gamma}, \end{aligned} \quad (11.15)$$

where once again the behavior of  $c_t$  is given by (11.6). In this case, a convenient analytical expression for  $\lambda$  is unavailable, due to the complicated nature of the integral needed to calculate the expectation on the left-hand side of (11.15). Hence we proceed in this case by approximating  $\lambda$  numerically.

The algorithm we use for this purpose takes the collection of parameters that appear in (11.15) and (11.6), including  $\lambda$ , as given. The parameters  $(A, \mu, \sigma^2)$ , dictating the behavior of consumption, were obtained

once again from an OLS regression of  $\log c_t$  on a constant and trend; the parameters  $(\beta, \gamma, \alpha, \delta, \theta)$  were obtained from Otronk (2001), and a grid of values was specified for  $\lambda$ . Because Otronk's estimates were obtained using postwar quarterly data on per capita consumption, defined as the consumption of nondurables and services, we reestimated  $(A, \mu, \sigma^2)$  using this alternative measure of consumption. The series is contained in `ycih.txt`, available at the textbook website. It is slightly more volatile than the annual measure of consumption considered by Lucas: the estimate of  $\sigma^2$  it yields is  $(0.0394)^2$ .

Given a specification of parameters, the right-hand side of (11.15) may be calculated directly, subject to two approximations. First, for each value of  $t$ , the infinite-order expression  $\sum_{j=0}^{\infty} \Phi_j A e^{\mu t - j}$  must be calculated using a finite-ordered approximation. It turns out that under the range of values for  $(\beta, \gamma, \alpha, \delta, \theta)$  estimated by Otronk, the corresponding sequence  $\{\Phi_j\}$  decays rapidly as  $j$  increases, reaching approximately zero for  $j = 6$ ; we set an upper bound of  $J = 10$  to be conservative. Second, the infinite time horizon over which the discounted value of instantaneous utility is aggregated must also be truncated. It turns out that summation over approximately a 1,500-period time span provides an accurate approximation of the infinite time horizon; we worked with an upper bound of  $T = 3,000$  to be conservative.

Using the same upper limits  $J$  and  $T$ , an additional approximation is required to calculate the expectation that appears on the left-hand side of (11.15). This is accomplished via numerical integration. In the present application, the process begins by obtaining a simulated realization of  $\{u_t\}_{t=1}^T$  from an  $N(0, \sigma^2)$  distribution, calculating  $\{\varepsilon_t\}_{t=1}^T$  using  $\varepsilon_t = e^{u_t}$ , and then inserting  $\{\varepsilon_t\}_{t=1}^T$  into (11.6) to obtain a simulated drawing of  $\{c_t\}_{t=1}^T$ . Using this drawing, the corresponding realization of the discounted value of instantaneous utility is calculated. These steps deliver the value of a single realization of lifetime utility. Computing the average value of realized lifetime utilities calculated over many replications of this process yields an approximation of the expected value of lifetime utility. The results reported below were obtained using 1,000 replications of this process. (A discussion of the accuracy with which this calculation approximates the actual integral we seek is provided in chapter 9.)

Finally, to determine the value of  $\lambda$  that satisfies (11.15), the left-hand side of this equation was approximated over a sequence of values specified for  $\lambda$ ; the specific sequence we used began at 0 and increased in increments of 0.000025. Given the risk aversion implied by the habit/durability specification, the left-hand side of (11.15) is guaranteed to be less than the right-hand side given  $\lambda = 0$ ; and as  $\lambda$  increases, the left-hand side increases, ultimately reaching the value of the right-hand side. The value of  $\lambda$  we seek

TABLE 11.2  
Welfare Costs, Habit/Durability Preferences

Parameterization	$\lambda$	\$/person, 2009
Baseline	0.000425	\$13.52
$\gamma = 0.5363$	0.000325	\$10.34
$\gamma = 0.9471$	0.000550	\$17.50
$\theta = 0.0178$	0.000425	\$13.52
$\theta = 0.3039$	0.000425	\$13.52
$\alpha = 0.2618$	0.000425	\$13.52
$\alpha = 0.6133$	0.000425	\$13.52
$\delta = 0.0223$	0.000550	\$17.50
$\delta = 0.3428$	0.000275	\$8.75
$\gamma = 1$	0.000570	\$18.25

in the approximation is the value that generates equality between the two sides.

Results of this exercise are reported in table 11.2 for nine specifications of  $(\beta, \gamma, \alpha, \delta, \theta)$ . The first is referenced as a baseline: these correspond with median values of the marginal posterior distributions Otronk calculated for each parameter: 0.9878, 0.7228, 0.446, 0.1533, 0.1299. Next, sensitivity to the specification of  $\gamma$  is demonstrated by respecifying  $\gamma$  first at the 5% quantile value of its marginal posterior distribution, then at its 95% quantile value. The remaining parameters were held fixed at their median values given the respecification of  $\gamma$ . Sensitivity to the specifications of  $\alpha$ ,  $\delta$ , and  $\theta$  is demonstrated analogously. Finally, a modification of the baseline under which  $\gamma = 1$  is reported, to facilitate a closer comparison of Lucas' estimate of  $\lambda$  obtained under CRRA preferences given  $\gamma = 1$ .

Under Otronk's baseline parameterization,  $\lambda$  is estimated as 0.00425. In terms of the annual version of consumption considered by Lucas, this corresponds with a consumption cost of \$13.52/person in 2009. For the modification of the baseline under which  $\gamma = 1$ , the cost rises to \$18.25/person, which slightly exceeds the cost of \$14.28/person calculated using CRRA preferences given  $\gamma = 1$ . The estimated value of  $\lambda$  is most sensitive to changes in  $\delta$ , which determines the persistence of the flow of services from past consumption in contributing to the durability stock. However, even in this case the range of estimates is modest: \$8.75/person given  $\delta = 0.3428$ ; \$17.50/person given  $\delta = 0.0223$ . These results serve to demonstrate the general insensitivity of Lucas' results to this particular modification of the specification of instantaneous utility; as characterized by Otronk: "It is a result that casts doubt that empirically plausible modifications to preferences alone could lead to large costs of consumption volatility" (88).

**Exercise 11.1**

Recalculate the welfare cost of business cycles for the CRRA case given a relaxation of the log-Normality assumption for  $\{\varepsilon_t\}$ . Do so using the following steps.

1. Using the consumption data contained in `welddat.txt`, regress the log of consumption on a constant and trend. Use the resulting parameter estimates to specify  $(A, \mu, \sigma^2)$  in (11.6), and save the resulting residuals in the vector  $u$ .
2. Construct a simulated realization of  $\{c_t\}_{t=0}^T$ ,  $T = 3,000$ , by obtaining random drawings (with replacement) of the individual elements of  $u$ . For each of the  $t$  drawings  $u_t$  you obtain, calculate  $\varepsilon_t = e^{u_t}$ ; then insert the resulting drawing  $\{\varepsilon_t\}_{t=0}^T$  into (11.6) to obtain  $\{c_t\}_{t=0}^T$ .
3. Using the drawing  $\{c_t\}_{t=0}^T$ , calculate  $\sum_{t=0}^T \beta^t \frac{[(1+\lambda)\varepsilon_t]^{1-\gamma}}{1-\gamma}$  using  $\beta = 0.96$ ,  $\gamma = 0.5, 1, 1.5, 2$  (recalling that for  $\gamma = 1$ ,  $\frac{x^{1-\gamma}}{1-\gamma} = \ln(x)$ ), and  $\lambda = 0, 0.000025, \dots, 0.0001$ .
4. Repeat steps 2 and 3 1,000 times, and record the average values of  $\sum_{t=0}^T \beta^t \frac{[(1+\lambda)\varepsilon_t]^{1-\gamma}}{1-\gamma}$  you obtain as approximations of  $E_0 \sum_{t=0}^T \beta^t \frac{[(1+\lambda)\varepsilon_t]^{1-\gamma}}{1-\gamma}$ .
5. Calculate the right-hand side of (11.9) using the estimates of  $A$  and  $\mu$  obtained in step 1. Do so for each value of  $\gamma$  considered in step 3.
6. For each value of  $\gamma$ , find the value of  $\lambda$  that most closely satisfies (11.9). Compare the values you obtain with those reported in table 11.1. Are Lucas' original calculations robust to departures from the log-Normality assumption adopted for  $\{\varepsilon_t\}$ ?

**Exercise 11.2**

Using the baseline estimates of the habit/durability specification reported in table 11.2, evaluate the robustness of the estimates of  $\lambda$  reported in table 11.2 to departures from the log-Normality assumption adopted for  $\{\varepsilon_t\}$ . Do so using simulations of  $\{c_t\}_{t=0}^T$  generated as described in step 2 of Exercise 11.1. Once again, are the results reported in table 11.2 robust to departures from the log-Normality assumption adopted for  $\{\varepsilon_t\}$ ?

**11.4 Productivity Shocks and Business Cycle Fluctuations**

As an example of an experiment involving a question of fit, we calibrate the RBC model presented in chapter 3, section 1, and examine

the extent to which it is capable of accounting for aspects of the behavior of output, consumption, investment, and hours introduced in chapters 6 and 7 (the data are contained in `ycih.txt`, available at the textbook website). Reverting to the notation used in the specification of the model, the parameters to be calibrated are as follows: capital's share of output  $\alpha$ ; the subjective discount factor  $\beta = \frac{1}{1+\varrho}$ , where  $\varrho$  is the subjective discount rate; the degree of relative risk aversion  $\phi$ ; consumption's share (relative to leisure) of instantaneous utility  $\varphi$ ; the depreciation rate of physical capital  $\delta$ ; the AR parameter specified for the productivity shock  $\rho$ ; and the standard deviation of innovations to the productivity shock  $\sigma$ .

Standard specifications of  $\beta$  result from the association of the subjective discount rate  $\varrho$  with average real interest rates: roughly 4%–5% on an annual basis, or 1%–1.25% on a quarterly basis. As a baseline, we select a rate of 1%, implying  $\beta = \frac{1}{1.01} = 0.99$ . The parameterization of the CRRA parameter was discussed in the previous section; as a baseline, we set  $\phi = 1.5$ , and consider [0.5, 2.5] as a plausible range of alternative values.

We use the long-run relationship observed between output and investment to identify plausible parameterizations of  $\alpha$  and  $\delta$ . Recall that steady state values of  $\frac{\bar{y}}{n}$  and  $\frac{\bar{i}}{n}$  are given by

$$\frac{\bar{y}}{n} = \eta,$$

$$\frac{\bar{i}}{n} = \delta\theta,$$

where

$$\theta = \left( \frac{\alpha}{\varrho + \delta} \right)^{\frac{1}{1-\alpha}},$$

$$\eta = \theta^\alpha.$$

Combining these expressions yields the relationship

$$\alpha = \left( \frac{\delta + \varrho}{\delta} \right) \frac{\bar{i}}{\bar{y}}. \quad (11.16)$$

Using the sample average of 0.175 as a measure of  $\frac{\bar{i}}{\bar{y}}$ , and given  $\varrho = 0.01$ , a simple relationship is established between  $\alpha$  and  $\delta$ .

Before illustrating this relationship explicitly, we use a similar step to identify a plausible range of values for  $\varphi$ . Here we use the steady state

expression for  $\bar{n}$ , the fraction of discretionary time spent on job-market activities:

$$\bar{n} = \frac{1}{1 + \left(\frac{1}{1-\alpha}\right) \left(\frac{1-\varphi}{\varphi}\right) [1 - \delta \theta^{1-\alpha}]} \quad (11.17)$$

According to the time-allocation study of Ghez and Becker (1975), which is based on household survey data,  $\bar{n}$  is approximately  $1/3$ . Using this figure in (11.17), and exploiting the fact that  $\bar{\epsilon} = [1 - \delta \theta^{1-\alpha}]$  (the sample average of which is 0.825), we obtain the following relationship between  $\alpha$  and  $\varphi$ :

$$\varphi = \frac{1}{1 + \frac{2(1-\alpha)}{\bar{\epsilon}/y}} \quad (11.18)$$

In sum, using sample averages to pin down  $\bar{\epsilon}$ ,  $\bar{\eta}$ , and  $\bar{n}$ , we obtain the relationships between  $\alpha$ ,  $\delta$ , and  $\varphi$  characterized by (11.16) and (11.18). Table 11.3 presents combinations of these parameters for a grid of values specified for  $\delta$  over the range [0.01, 0.04], implying a range of annual depreciation rates of [4%, 16%]. As a baseline, we specify  $\delta = 0.025$  (10% annual depreciation), yielding specifications of  $\alpha = 0.24$  and  $\varphi = 0.35$  that match the empirical values of  $\bar{\epsilon}$ ,  $\bar{\eta}$ , and  $\bar{n}$ . Regarding the specification of  $\alpha$ , this is somewhat lower than the residual value attributed to capital's share given the measure of labor's share calculated from National Income and Product Accounts (NIPA) data. In the NIPA data, labor's share is approximately  $2/3$ , implying the specification of  $\alpha = 1/3$ . The reason for the difference in this application is the particular measure of investment we use (real gross private domestic investment). Using the same measure of

TABLE 11.3  
Trade-offs between  $\delta$ ,  $\alpha$ , and  $\varphi$

$\delta$	$\alpha$	$\varphi$
0.010	0.35	0.39
0.015	0.29	0.37
0.020	0.26	0.36
0.025	0.24	0.35
0.030	0.23	0.35
0.035	0.22	0.35
0.040	0.22	0.35

Note: The moments used to establish these trade-offs are  $\bar{\epsilon} = 0.175$ ,  $\bar{\eta} = 0.825$ , and  $\bar{n} = \frac{1}{3}$ .

investment over the period 1948:I–1995:IV, DeJong and Ingram (2001) obtain an estimate of  $\alpha = 0.23$  (with a corresponding estimate of  $\delta = 0.02$ ) using Bayesian methods described in chapter 14.

The final parameters to be established are those associated with the behavior of the productivity shock  $z_t$ :  $\rho$  and  $\sigma$ . We obtain these by first measuring the behavior of  $z_t$  implied by the specification of the production function, coupled with the observed behavior of output, hours, and physical capital. Given this measure, we then estimate the AR parameters directly via OLS.

Regarding physical capital, the narrow definition of investment we use must be taken into account in measuring the corresponding behavior of capital. We do so using a tool known as the perpetual inventory method. This involves the input of  $\{i_t\}_{t=1}^T$  and  $k_0$  into the law of motion of capital

$$k_{t+1} = i_t + (1 - \delta) k_t \quad (11.19)$$

to obtain a corresponding sequence  $\{k_t\}_{t=1}^T$ . A measure of  $k_0$  may be obtained in four steps: divide both sides of (11.19) by  $y_t$ ; use beginning-of-sample averages to measure the resulting ratios  $\bar{x}_y$ ,  $x = i, k$ ; solve for

$$\bar{k} = \frac{1}{\delta} \bar{i}; \quad (11.20)$$

and finally, multiply  $\frac{1}{\delta} \bar{i}$  by  $y_0$ . The results reported here are based on the specification of  $\bar{i}$  obtained using an eight-period, or two-year, sample average.

Given this measure of  $\{k_t\}_{t=1}^T$ ,  $\log z_t$  is derived following Solow (1956) as the unexplained component of  $\log y_t$  given the input of  $\log k_t$  and  $\log n_t$  in the production function:

$$\log z_t = \log y_t - \alpha \log k_t - (1 - \alpha) \log n_t. \quad (11.21)$$

For this reason,  $z_t$  is often referred to as a Solow residual. Finally, we apply the Hodrick-Prescott (H-P) filter to each variable, and estimate  $\rho$  and  $\sigma$  using the HP-filtered version of  $z_t$ . The resulting estimates are 0.78 and 0.0067.

Having parameterized the model, we characterize its implications regarding the collection of moments reported for the HP-filtered data in table 7.1. Moments calculated using both the model and data are reported in table 11.4.

As these results indicate, the model performs well in characterizing the patterns of serial correlation observed in the data, and also replicates the patterns of volatility observed for consumption and investment relative to output: the former is quite smooth relative to the latter. However,

TABLE 11.4  
Moment Comparison

H-P filtered data					
$j$	$\sigma_j$	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\varphi_{j,y}(1)$
$y$	0.0178	1.00	0.87	1.00	0.87
$c$	0.0078	0.44	0.84	0.83	0.76
$i$	0.0761	4.26	0.80	0.96	0.81
$n$	0.0191	1.07	0.91	0.83	0.62

RBC model					
$j$	$\sigma_j$	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\varphi_{j,y}(1)$
$y$	0.0184	1.00	0.79	1.00	0.79
$c$	0.0085	0.46	0.96	0.77	0.62
$i$	0.0799	4.34	0.74	0.95	0.75
$n$	0.0087	0.47	0.75	0.91	0.72

Note:  $\varphi(1)$  denotes first-order serial correlation;  $\varphi_{j,y}(l)$  denotes  $l^{\text{th}}$  order correlation between variables  $j$  and  $y$ . Model moments based on the parameterization

$$\mu = [\alpha \beta \phi \delta \rho \sigma]^T = [0.24 \ 0.99 \ 1.5 \ 0.35 \ 0.025 \ 0.78 \ 0.0067]^T.$$

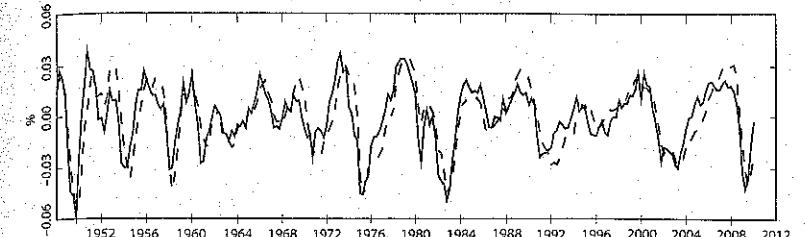
it performs poorly in characterizing the relative volatility of hours, which are roughly equally as volatile as output in the data, but less than  $\frac{1}{2}$  as volatile in the model. Figure 11.1 illustrates this shortcoming by depicting the time-series paths followed by output and hours in the actual data, along with paths followed by counterparts simulated from the model. The simulated hours series is far smoother than the corresponding output series.

The standard RBC model's characterization of the volatility of hours is a well-known shortcoming, and has prompted many extensions that improve upon its performance along this dimension. Examples include specifications featuring indivisible labor hours (Hansen, 1985; Rogerson, 1988; Kydland and Prescott, 1991b); home production (Benhabib, Rogerson, and Wright, 1991; Greenwood and Hercowitz, 1991); labor hoarding (Burnside, Eichenbaum, and Rebelo, 1993; Burnside and Eichenbaum, 1997); fiscal disturbances (McGrattan, 1994); and skill-acquisition activities (Einarsson and Marquis, 1997; Perli and Sakellaris, 1998; and DeJong and Ingram, 2001). Thus this example serves as a prime case study for the use of calibration experiments as a means of facilitating the “testing and development of theory,” as advocated by Kydland and Prescott (1996, 70).

### Exercise 11.3

Reconstruct table 11.4 by using the alternative combinations of values specified for  $\delta$ ,  $\alpha$ , and  $\varphi$  listed in table 11.3. Also, consider 0.5 and 2.5 as

Output and Hours, Actual Data



Output and Hours, Simulated Data

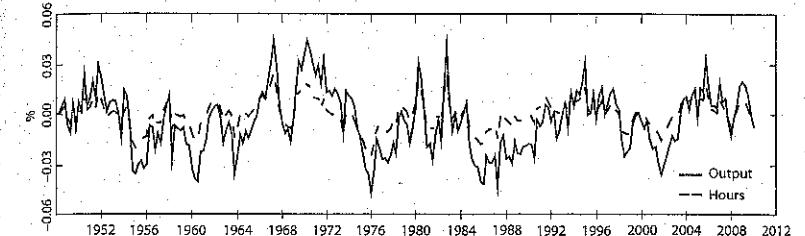


Figure 11.1 Comparisons of Output and Hours

alternative specifications of  $\phi$ . Are the results of table 11.4 materially altered by any of the alternative choices you considered?

### Exercise 11.4

Reconduct the calibration exercise of this section using the extension of the RBC model outlined in chapter 3, section 1 that explicitly incorporates long-term growth. Pay careful attention to the fact that equations (11.16) and (11.18), used to restrict the parameterizations of  $\delta$ ,  $\alpha$ , and  $\varphi$ , will be altered given this extension. So too will be (11.20), which was used to specify  $k_0$ . Once again, are the results of table 11.4 materially altered given this extension of the model?

## 11.5 The Equity Premium Puzzle

The calibration experiment of Mehra and Prescott (1985) also serves as a prime case study involving the testing and development of theory. The test they conducted sought to determine whether the asset-pricing model presented in chapter 3, section 3.2, is capable of characterizing patterns of returns generated by relatively risky assets (equity) and riskless assets

(Treasury bills). The data they considered for this purpose consist of the annual real return on the Standard and Poor's 500 composite index, the annualized real return on Treasury bills, and the growth rate of real per capita consumption on nondurables and services. The time span of their data is 1889–1978; here we demonstrate a brief replication of their study using data updated through 2009. The original data set is contained in `mpepp.txt`, and the updated data set is contained in `mpeppext.txt`; both are available at the textbook website.

As characterized in chapter 3, section 3.2, Mehra and Prescott's statement of the equity premium puzzle amounts to a presentation of the empirical incompatibility of the following equations, derived in a two-asset environment:

$$\beta E_t \left\{ \frac{u'(c_{t+1})}{u'(c_t)} r_{t+1}^f - 1 \right\} = 0 \quad (11.22)$$

$$E_t \left\{ \frac{u'(c_{t+1})}{u'(c_t)} [r_{t+1}^e - r_{t+1}^f] \right\} = 0, \quad (11.23)$$

where  $r_{t+1}^f$  and  $r_{t+1}^e$  denote the gross returns associated with a risk-free and risky asset. The difference  $r_{t+1}^e - r_{t+1}^f$  is referred to as the equity premium. Sample averages (standard deviations) of  $r_{t+1}^f$  and  $r_{t+1}^e$  are 0.77% (5.92%) and 6.98% (16.53%) in the data ending in 1978, and 1.12% (5.28%) and 7.28% (16.57%) in the data ending through 2009. Thus equity returns are large and volatile relative to risk-free returns. The question is: can this pattern of returns be reconciled with the attitudes towards risk embodied by the CRRA preference specification?

Given CRRA preferences, with  $\gamma$  representing the risk-aversion parameter, the equations are given by

$$\beta E_t \left\{ \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} r_{t+1}^f - 1 \right\} = 0 \quad (11.24)$$

$$E_t \left\{ \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} [r_{t+1}^e - r_{t+1}^f] \right\} = 0. \quad (11.25)$$

Here, we use sample averages to estimate the expectations represented in (11.24) and (11.25), and seek to determine whether it is possible to specify  $\gamma$  in order to jointly account for each equation, given  $\beta = 0.99$ . The results of this exercise are presented in table 11.5.

Note that for a specification of  $\gamma$  between 0 and 0.5, it is possible to account for the first equation. The reason for the requirement of such a low value is that in addition to characterizing attitudes towards risk,  $\gamma$  plays

TABLE 11.5  
The Equity Premium and Risk-Free Rate Puzzles

$\gamma$	$\beta \left( \frac{c_{t+1}}{c_t}^{-\gamma} r_{t+1}^f \right) - 1$	$\left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} [r_{t+1}^e - r_{t+1}^f]$
0	0.001	0.062
0.5	-0.007	0.060
1	-0.015	0.059
1.5	-0.023	0.057
2	-0.030	0.056
2.5	-0.037	0.054
3	-0.044	0.053

Note:  $\bar{x}$  denotes the sample mean of  $x$ . Results obtained using  $\beta = 0.99$ .

the role of determining the household's intertemporal elasticity of substitution, given by  $1/\gamma$ . The presence in the data of a relatively large average consumption growth rate (1.83% in the data ending in 1978, and 1.75% in the data ending in 2009), coupled with the low average returns generated by the risk-free rate, implies a highly elastic intertemporal substitution parameter, and thus a low value of  $\gamma$ . Of course, the specification of a smaller value for  $\beta$  requires an even lower value for  $\gamma$ ; thus  $\beta = 0.99$  represents nearly a best-case scenario along this dimension.

In turn, a large specification of  $\gamma$  is required to account for the second equation. In fact, the sample average first becomes negative given  $\gamma = 20$ . The reason is that the relatively large premium associated with returns generated by the risky asset implies substantial risk aversion on the part of households. This in turn implies a highly inelastic intertemporal elasticity of substitution given CRRA preferences, which is inconsistent with the first equation. Thus the puzzle.

Mehra and Prescott's (1985) statement of the puzzle spawned an enormous literature seeking to determine whether alterations of preferences or additional features of the environment are capable of accounting for this behavior. Surveys of this literature are given by Kocherlakota (1996) and Mehra and Prescott (2003). Certainly these efforts have served to make headway towards a resolution; however, the conclusion of both surveys is that the equity premium remains a puzzle.

### Exercise 11.5

Replicate the calculations presented in table 11.5 using the baseline parameterization of the habit/durability specification presented in section 3. Does this modification represent progress in resolving the puzzle? Ponder the intuition behind your findings.

## 11.6 Critiques and Extensions

### 11.6.1 Critiques

As noted, Kydland and Prescott's (1982) use of a calibration exercise to implement the DSGE model they studied represents a pathbreaking advance in the conduct of empirical work in macroeconomic applications. Moreover, as illustrated in the applications above, calibration exercises can certainly serve as an effective means of making headway in empirical applications involving general-equilibrium models. At a minimum, they are well-suited for conveying a quick initial impression regarding the empirical strengths and weaknesses of a given model along specific dimensions chosen by the researcher. This latter attribute is particularly important, in that it provides an effective means of discovering dimensions along which extensions to the model are most likely to bear fruit.

This being said, the lack of statistical formality associated with calibration exercises imposes distinct limitations upon what can be learned and communicated via their use. Moreover, the particular approach advocated by Kydland and Prescott (as presented in section 2) for addressing empirical questions in the absence of a formal statistical framework has been criticized on a variety of fronts. We conclude this chapter by summarizing certain aspects of this criticism, and discussing some closely related extensions to calibration that retain the general spirit of the exercise.

In the preface to his (1944) articulation of the probability approach to econometrics, Haavelmo opened with a criticism of the approach that prevailed at the time. The criticism is striking in its applicability to calibration exercises devoted to questions regarding fit (the included quotation marks and italics are Haavelmo's):

So far, the common procedure has been, first to construct an economic theory involving *exact* functional relationships, then to compare this theory with some actual measurements, and, finally, "to judge" whether the correspondence is "good" or "bad." Tools of statistical inference have been introduced, in some degree, to support such judgements, e.g., the calculation of a few standard errors and multiple-correlation coefficients. The application of such simple "statistics" has been considered legitimate, while, at the same time, the adoption of definite probability models has been deemed a crime in economic research, a violation of the very nature of economic data. That is to say, it has been considered legitimate to use some of the *tools* developed in statistical theory *without accepting the very foundation* upon which statistical theory is built. For *no tool developed in the theory of statistics has any meaning—except, perhaps, for descriptive purposes—without being referred to some stochastic scheme.* (iii)

## 11.6 Critiques and Extensions

Our interpretation of the thrust of this criticism is that in the absence of statistical formality, communication regarding the results of an experiment is problematic. Judgments of "good" or "bad," or as Kydland and Prescott (1996, 71) put it, judgments of whether "the predictions of theory match the observations," are necessarily subjective. In applications such as Mehra and Prescott's (1985) identification of the equity premium puzzle, empirical shortcomings are admittedly fairly self-evident. But in evaluating marginal gains in empirical performance generated by various modifications of a baseline model, the availability of coherent and objective reporting mechanisms is invaluable. No such mechanism is available in conducting calibration exercises.

Reacting to Kydland and Prescott (1996), Sims (1996) makes a similar point:

Economists can do very little experimentation to produce crucial data. This is particularly true of macroeconomics. Important policy questions demand opinions from economic experts from month to month, regardless of whether professional consensus has emerged on the questions. As a result, economists normally find themselves considering many theories and models with legitimate claims to matching the data and predicting the effects of policy. We have to deliver recommendations or accurate descriptions of the nature of the uncertainty about the consequences of alternative policies, despite the lack of a single accepted theory. (107)

Moreover:

Axiomatic arguments can produce the conclusion that anyone making decisions under uncertainty must act as if that agent has a probability distribution over the uncertainty, updating the probability distribution by Bayes' rule as new evidence accumulates. People making decisions whose results depend on which of a set of scientific theories is correct should therefore be interested in probabilistic characterizations of the state of the evidence. (108)

These observations lead him to conclude:

formal statistical inference is not necessary when there is no need to choose among competing theories among which the data do not distinguish decisively. But if the data do not make the choice of theory obvious, and if decisions depend on the choice, experts can report and discuss their conclusions reasonably only using notions of probability. (110)

Regarding the problem of choosing among given theories, there is no doubt that from a classical hypothesis testing perspective, under which one model is posed as the null hypothesis, this is complicated by the fact that the models in question are "necessarily false." But this problem is not unique to the analysis of DSGE models, as the epigraph to this chapter by Theil

implies. Moreover, this problem is not an issue given the adoption of a Bayesian perspective, under which model comparisons involve calculations of the relative probability assigned to alternative models, conditional on the observed data. Under this perspective, there is no need for the declaration of a null model; rather, all models are treated symmetrically, and none are assumed a priori to be "true." Details regarding this approach are provided in chapter 14.

Also in reaction to Kydland and Prescott (1996), Hansen and Heckman (1996) offer additional criticisms. For one, they challenge the view that calibration experiments involving questions of fit are to be considered as distinct from estimation: "The distinction drawn between calibrating and estimating the parameters of a model is artificial at best" (91). Their reasoning fits nicely with the description provided in section 1 of the means by which the specification of  $\mu$  is achieved via use of (11.4); and the means by which judgments of fit are achieved via comparisons of  $\Phi(\{X_t\}_{t=1}^T)$  with  $\Phi(\{X_t^M\}_{t=1}^T)$ :

Econometricians refer to the first stage as *estimation* and the second stage as *testing*. . . . From this perspective, the Kydland-Prescott objection to mainstream econometrics is simply a complaint about the use of certain loss functions for describing the fit of a model to the data or for producing parameter estimates. (92)

In addition, Hansen and Heckman call into question the practice of importing parameter estimates obtained from micro studies into macroeconomic models. They do so on two grounds. First, such parameter estimates are associated with uncertainty. In part this uncertainty can be conveyed by reporting corresponding standard errors, but in addition, model uncertainty plays a substantial role (a point emphasized by Sims). Second, "It is only under very special circumstances that a micro parameter . . . can be 'plugged into' a representative consumer model to produce an empirically concordant aggregate model" (88). As an example of such a pitfall, they cite Houthakker (1956), who demonstrated that the aggregation of Leontief micro production technologies yields an aggregate Cobb-Douglas production function.

These shortcomings lead Hansen and Heckman and Sims to similar conclusions. To quote Hansen and Heckman:

Calibration should only be the starting point of an empirical analysis of general-equilibrium models. In the absence of firmly established estimates of key parameters, sensitivity analyses should be routine in real business cycle simulations. Properly used and qualified simulation methods can be an important source of information and an important stimulus to high-quality empirical economic research. (101)

And to quote Sims:

A focus on solving and calibrating models, rather than carefully fitting them to data, is reasonable at a stage where solving the models is by itself a major research task. When plausible theories have been advanced, though, and when decisions depend on evaluating them, more systematic collection and comparison of evidence cannot be avoided. (109)

It is precisely in the spirit of these sentiments that we present the additional material contained in Part V of this book.

### 11.6.2 Extensions

We conclude this chapter by presenting two extensions to the basic calibration exercise. In contrast to the extensions presented in the chapters that follow, the extensions presented here are distinct in that they do not entail an estimation stage. Instead, both are designed to provide measures of fit for calibrated models.

The first extension is due to Watson (1993), who proposed a measure of fit based on the size of the stochastic error necessary for reconciling discrepancies observed between the second moments corresponding to the model under investigation and those corresponding with the actual data. Specifically, letting  $X_t$  denote the  $m \times 1$  vector of observable variables corresponding with the model, and  $Y_t$  their empirical counterparts, Watson's measure is based on the question: how much error  $U_t$  would have to be added to  $X_t$  so that the autocovariances of  $X_t + U_t$  are equal to the autocovariances of  $Y_t$ ?

To quantify this question, recall from chapter 7 that the  $s^{\text{th}}$ -order autocovariance of a mean-zero covariance-stationary stochastic process  $z_t$  is given by the  $m \times m$  matrix

$$E(z_t z_{t-s}') \equiv \Gamma_z(s).$$

Therefore the  $s^{\text{th}}$ -order autocovariance of  $U_t = Y_t - X_t$  is given by

$$\Gamma_U(s) = \Gamma_Y(s) + \Gamma_X(s) - \Gamma_{XY}(s) - \Gamma_{YX}(s), \quad (11.26)$$

where

$$\Gamma_{XY}(s) = E(X_t Y_{t-s}')$$

and

$$\Gamma_{YX}(s) = \Gamma_{YX}(-s)'.$$

With the autocovariance-generating function (ACGF) of  $z_t$  given by

$$A_z(e^{-i\omega}) = \sum_{s=-\infty}^{\infty} \Gamma_z(s)e^{-i\omega s},$$

where  $i$  is complex and  $\omega \in [0, 2\pi]$  represents a particular frequency, the ACGF of  $U_t$  implied by (11.26) is given by

$$\begin{aligned} A_U(e^{-i\omega}) &= A_Y(e^{-i\omega}) + A_X(e^{-i\omega}) \\ &\quad - A_{XY}(e^{-i\omega}) - A_{YX}(e^{i\omega}), \end{aligned} \quad (11.27)$$

where

$$A_{XY}(e^{i\omega})' = A_{YX}(e^{-i\omega}).$$

As discussed in chapter 7, it is straightforward to construct  $A_Y(e^{-i\omega})$  and  $A_X(e^{-i\omega})$  given observations on  $Y_t$  and a model specified for  $X_t$ . However, absent theoretical restrictions imposed upon  $A_{XY}(e^{-i\omega})$ , and absent joint observations on  $(Y_t, X_t)$ , the specification of  $A_{XY}(e^{-i\omega})$  is arbitrary. Watson overcame this problem by proposing a restriction on  $A_{XY}(e^{-i\omega})$  that produces a lower bound for the variance of  $U_t$ , or in other words, a best-case scenario for the model's ability to account for the second moments of the data.

To characterize the restriction, it is useful to begin with the implausible but illustrative case in which  $X_t$  and  $Y_t$  are serially uncorrelated. In this case, the restriction involves minimizing the size of the covariance matrix of  $U_t$ , given by

$$\Sigma_U = \Sigma_Y + \Sigma_X - \Sigma_{XY} - \Sigma_{YX}. \quad (11.28)$$

Because there is no unique measure of the size of  $\Sigma_U$ , Watson proposes the minimization of the trace of  $W\Sigma_U$ , where  $W$  is an  $m \times m$  weighting matrix that enables the researcher to assign alternative importance to linear combinations of the variables under investigation. If  $W$  is specified as the identity matrix, then each individual variable is treated as equally important. If alternatively the researcher is interested in  $GY_t$  and  $GX_t$ , then  $W$  can be chosen as  $G'G$ , since

$$\text{tr}(G\Sigma_U G') = \text{tr}(G'G\Sigma_U).$$

Given  $W$ , Watson shows that

$$\Sigma_{XY} = C_X V U' C_Y' \quad (11.29)$$

is the unique specification of  $\Sigma_{XY}$  that minimizes  $\text{tr}(W\Sigma_U)$ . In (11.29),  $C_X$  and  $C_Y$  are arbitrary  $m \times m$  matrix square roots of  $\Sigma_X$  and  $\Sigma_Y$  (e.g.,

## 11.6 Critiques and Extensions

$\Sigma_X = C_X C_X'$ , an example of which is the Cholesky decomposition), and the matrices  $U$  and  $V$  are obtained by computing the singular value decomposition of  $C_Y' W C_X$ . Specifically, the singular value decomposition of  $C_Y' W C_X$  is given by

$$C_Y' W C_X = USV', \quad (11.30)$$

where  $U$  is an  $m \times k$  orthonormal matrix (i.e.,  $U'U$  is the  $k \times k$  identity matrix),  $S$  is a  $k \times k$  matrix, and  $V$  is a  $k \times k$  orthonormal matrix.<sup>2</sup>

For the general case in which  $X_t$  and  $Y_t$  are serially correlated, Watson's minimization objective translates directly from  $\Sigma_{XY}$  to  $A_{XY}(e^{-i\omega})$ . Recall from chapter 7 the relationship between ACGFs and spectra; in the multivariate case of interest here, the relationship is given by

$$s_z(\omega) = \left( \frac{1}{2\pi} \right) \sum_{s=-\infty}^{\infty} \Gamma_z(s)e^{-i\omega s}, \quad \omega \in [-\pi, \pi].$$

With  $C_X(\omega)$  now denoting the matrix square root of  $s_X(\omega)$  calculated for a given specification of  $\omega$ , and so on, the specification of  $A_{XY}(e^{-i\omega})$  that minimizes  $\text{tr}(W\Sigma_U)$  is given by

$$A_{XY}(e^{-i\omega}) = C_X(\omega)V(\omega)U'(\omega)C_Y'(\omega),$$

where  $U(\omega)$  and  $V(\omega)$  are as indicated in (11.30).

In this way, the behavior of the minimum-variance stochastic process  $U_t$  can be analyzed on a frequency-by-frequency basis. As a summary of the overall performance of the model, Watson proposes a relative mean square approximation error (RMSAE) statistic, analogous to a lower bound on  $1 - R^2$  statistics in regression analyses (lower is better):

$$r_j(\omega) = \frac{A_U(e^{-i\omega})_{jj}}{A_Y(e^{-i\omega})_{jj}}, \quad (11.31)$$

where  $A_U(e^{-i\omega})_{jj}$  denotes the  $j^{\text{th}}$  diagonal element of  $A_U(e^{-i\omega})$ .

As an illustration, figure 11.2 closely follows Watson by demonstrating the application of his procedure to the RBC model presented in section 4, parameterized as indicated in table 11.4. Following Watson, the figure illustrates spectra corresponding to first differences of both the model variables and their empirical counterparts. This is done to better accentuate behavior over business-cycle frequencies (i.e., frequencies between 1/40

<sup>2</sup> In GAUSS, Cholesky decompositions can be obtained using the command chol, and singular value decompositions can be obtained using the command svd1.

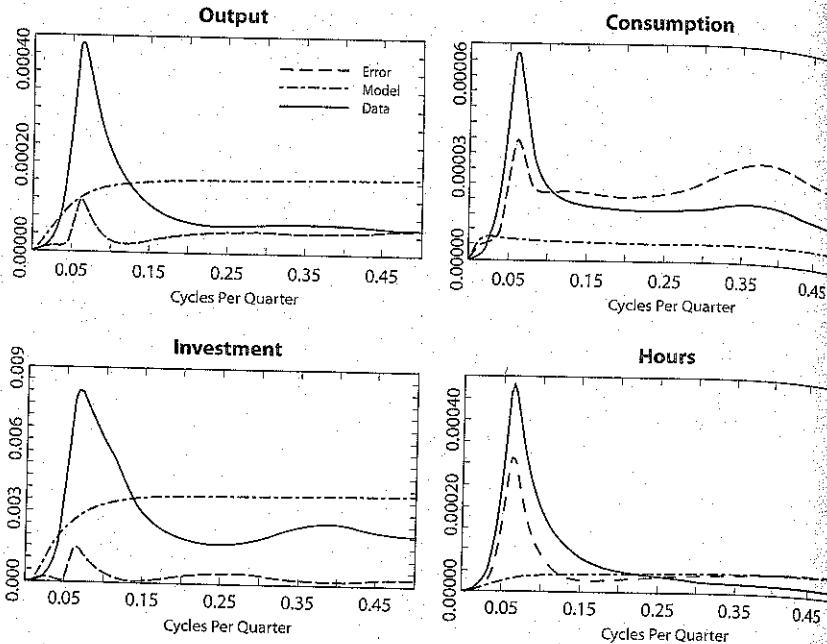


Figure 11.2. Decomposition of Spectra

and  $1/6$  cycles per quarter). The data were also HP-filtered, and  $W$  was specified as the  $4 \times 4$  identity matrix.<sup>3</sup>

The striking aspect of this figure is that the model fails to capture the spectral peaks observed in the data over business cycle frequencies. Given this failure, associated RMSAE statistics are decidedly anticlimactic. Nevertheless, RMSAEs calculated for output, consumption, investment, and hours over the  $[1/40, 1/6]$  frequency range are given by  $0.17, 0.78, 0.08$ , and  $0.53$ ; thus the model's characterization of consumption and hours over this range is seen to be particularly poor.

Watson's demonstration of the failure of the standard RBC model to produce spectral peaks prompted several efforts devoted to determining whether plausible modifications of the model are capable of improving empirical performance along this dimension. For example, Wen (1998) obtained an affirmative answer to this question by introducing two modifications: an employment externality, under which the level of aggregate

<sup>3</sup> Three GAUSS procedures were used to produce this figure: `modspec.prc`, `dataspec.prc`, and `errspectrum.prc`. All are available at the textbook website.

employment has an external effect on sectoral output; and the specification of habit formation in the enjoyment of leisure activities. So too did Otron (2001), for the model (described in section 3) he used to analyze the welfare cost of business cycles. Thus the results of figure 11.2 do not provide a general characterization of the empirical performance of RBC models along this dimension.

The second extension is due to related work by Canova (1995) and DeJong, Ingram, and Whiteman (1996). Each study proposed the replacement of a single set of values specified over  $\mu$  with a prior distribution  $\pi(\mu)$ . The distribution induced by  $\pi(\mu)$  over a collection of empirical targets chosen by the researcher is then constructed and compared with a corresponding distribution calculated from the actual data. For example, the collection of moments featured in table 11.4 used to evaluate the RBC model presented in section 4 represents a common choice of empirical targets. Reverting to the notation used in section 2, hereafter we represent the selected targets as  $\Omega$ .

The difference between these studies lies in the empirical distributions over  $\Omega$  they use. DeJong et al. worked with posterior distributions over  $\Omega$  obtained from the specification of a vector autoregressive (VAR) model for the actual data; Canova worked with sampling distributions. Methodologies available for calculating posterior distributions are not presented in this book until chapter 14, so details regarding the implementation of the extension proposed by DeJong et al. are not provided here. Instead, we present an algorithm for calculating sampling distributions over  $\Omega$  induced by a VAR specification for the data, and thus for implementing Canova's measure of fit. This represents a specialization of the Monte Carlo method presented in chapter 7, section 2, as a means of approximating standard errors numerically.

Using the notation of chapter 7, let the VAR specified for  $X_t$  be given by

$$X_t = \gamma_1 X_{t-1} + \gamma_2 X_{t-2} + \cdots + \gamma_p X_{t-p} + \varepsilon_t, \quad E(\varepsilon_t \varepsilon_t') = \Sigma. \quad (11.32)$$

The algorithm takes as inputs OLS estimates  $\hat{\gamma}_j, j = 1, \dots, p$ ,  $\hat{\Sigma}$ , estimated residuals  $\{\hat{\varepsilon}_t\}$ , and the first  $p$  observations of  $X_t$ , which serve as starting values. Given these inputs, a simulated drawing  $\tilde{X}_{p+1}$  can be constructed by obtaining an artificial drawing  $\tilde{\varepsilon}_p$  from a given distribution, and evaluating the right-hand side of (11.32) using  $\{X_1, X_2, \dots, X_{p-1}, X_p\}$  and  $\tilde{\varepsilon}_p$ . Next,  $\tilde{X}_{p+2}$  is constructed by obtaining a second drawing  $\tilde{\varepsilon}_{p+1}$ , and evaluating the right-hand side of (11.32) using  $\{X_2, X_3, \dots, X_p, \tilde{X}_{p+1}\}$  and  $\tilde{\varepsilon}_p$ . Performing  $T$  replications of this process yields an artificial drawing  $\{\tilde{X}_t\}$ . (The influence of the initial observations  $\{X_1, X_2, \dots, X_{p-1}, X_p\}$  can be reduced by obtaining  $T + T'$  drawings of  $\tilde{X}_t$ , and discarding the first  $T'$

drawings in constructing  $\{\tilde{X}_t\}$ .) For each of  $J$  replications of this process, a collection of  $J$  estimates of  $\Omega$  is calculated; the resulting distribution of  $\Omega$  approximates the sampling distribution we seek.

There are two common general methods for obtaining artificial drawings  $\tilde{\varepsilon}_t$ . One method involves a distributional assumption for  $\{\varepsilon_t\}$ , parameterized by  $\hat{\Sigma}$ . For example, let  $\hat{S}$  denote the Cholesky decomposition of  $\hat{\Sigma}$ . Then under the assumption of Normality for  $\{\varepsilon_t\}$ , drawings  $\tilde{\varepsilon}_t$  may be obtained using  $\tilde{\varepsilon}_t = \hat{S}\tilde{u}$ , where  $\tilde{u}$  represents an  $m \times 1$  vector of independent  $N(0, 1)$  random variables. Alternatively,  $\tilde{\varepsilon}_t$  may be obtained as a drawing (with replacement) from the collection of residuals  $\{\hat{\varepsilon}_t\}$ .

Let the distribution obtained for the  $i^{\text{th}}$  element of  $\Omega$  be given by  $S(\Omega_i)$ , and let  $[a, b]_i$  denote the range of values for  $\Omega_i$  derived by subtracting and adding one standard deviation of  $S(\Omega_i)$  to the mean value of  $S(\Omega_i)$ . Also, let  $\pi(\Omega_i)$  denote the prior distribution over  $\Omega_i$  induced by  $\pi(\mu)$ . Then Canova's measure of fit along this dimension of the target space is the proportion  $\pi(\Omega_i)$  that lies within  $[a, b]_i$ :

$$f_i = \int_{a_i}^{b_i} \pi(\Omega_i) d\Omega_i. \quad (11.33)$$

Thus  $f_i \in [0, 1]$ , and the greater is the proportion of  $\pi(\Omega_i)$  contained in the range  $[a, b]_i$ , the closer  $f_i$  will be to 1. A measure of fit related to that proposed by DeJong et al. is obtained by replacing the range  $[a, b]_i$  with an analogous coverage interval corresponding to a posterior distribution specified over  $\Omega_i$ .

### Exercise 11.6

Using the RBC model presented in section 4, calculate  $f_i$  for the collection of moments analyzed in table 11.4. Use a six-lag VAR specified for the data to construct  $[a, b]_i$  for each moment. Do so using the version of the Monte Carlo algorithm described above under which artificial drawings  $\tilde{\varepsilon}_t$  are obtained as drawings (with replacement) from the collection of residuals  $\{\hat{\varepsilon}_t\}$ . Finally, use the following prior distribution specified over the elements of  $\mu$ :  $\beta \sim U[0.988, 0.9925]$ ;  $\delta \sim U[0.01, 0.04]$ ;  $\rho \sim U[0.75, 0.81]$ , and  $U$  denoting the uniform distribution. (Uniform drawings of, e.g.,  $\beta$  over  $[\underline{\beta}, \bar{\beta}]$  may be obtained using the transformation  $\tilde{\beta} = \beta + (\bar{\beta} - \underline{\beta})\tilde{u}$ , where  $\tilde{u}$  is a drawing from a  $U[0, 1]$  distribution.) Given a drawing of  $\delta$  and  $Q = 1/\beta - 1$ , calculate the corresponding value of  $\alpha$  using (11.16), and the corresponding value of  $\varphi$  using (11.18). Throughout, hold  $\sigma$  fixed at 0.0067.

## Chapter 12

### Matching Moments

Know the right moment.

—“The Seven Sages,” From Diogenes Laertius,  
*Lives of Eminent Philosophers*

### 12.1 Overview

In the previous chapter, we characterized calibration as an exercise under which a set of empirical targets is used to pin down the parameters of the model under investigation, and a second set of targets is used to judge the model's empirical performance. Here we present a collection of procedures that establish a statistical foundation upon which structural models can be parameterized and evaluated. Under these procedures, parameterization is accomplished via estimation, and empirical performance is assessed via hypothesis testing.

As with calibration, the focus of these procedures remains on a set of empirical targets chosen by the researcher. Thus we broadly characterize their implementation as involving the matching of moments (although our use of the term moments extends rather liberally to include empirical targets such as spectra and impulse response functions). But in contrast with calibration, these procedures involve the adoption of a classical statistical perspective under which the model under investigation is interpreted as a potential data generation process from which the actual data, and thus the moments used to characterize the data, were realized. As discussed in chapter 7, statistical uncertainty associated with the realization of these moments is represented by their corresponding standard errors.

In the estimation stage, moment-matching procedures seek to determine the parameterization  $\mu$  that best enables the underlying structural model to match a collection of preselected moments. The parameterization  $\hat{\mu}$  that accomplishes this objective is interpreted as an estimate of the actual value of these parameters. Because the estimate  $\hat{\mu}$  is a function of the data, it too is associated with statistical uncertainty, represented in the form of standard errors. In the testing stage, the goal is to determine whether the collection of moments selected as targets can plausibly be

interpreted as a random drawing from the underlying structural model. If the probability associated with this outcome falls below a preselected threshold, or critical value, the model is rejected as a potential data-generating process; otherwise it is accepted. Typically, moment-matching procedures accomplish estimation and testing simultaneously, through a process known as overidentification.

Because the procedures described in this chapter accomplish estimation and testing on the basis of a preselected collection of moments, rather than on the full range of statistical implications associated with the model under investigation, they are often referred to as limited-information procedures. This characteristic is at once a strength and a weakness. The full range of statistical implications associated with a given model is conveyed by its corresponding likelihood function. By working instead with a collection of moments, these procedures entail a loss of efficiency that can sometimes be problematic in working with small sample sizes. Also, resulting inferences can potentially be sensitive to the particular collection of moments chosen for the analysis. On the other hand, the establishment of a likelihood function involves an assumption regarding the distribution from which the structural shocks  $v_t$  have been realized. Unless they require a simulation step, moment-matching procedures require no such assumption, and are therefore not subject to potential concerns regarding misspecification along this dimension.

Moment-matching procedures date at least to Pearson (1894), and a wide range of specific alternatives remain in use today. Here we focus on three: the Generalized Method of Moments (GMM), the Simulated Method of Moments (SMM), and indirect inference (II). General textbook discussions are available, for example, in Hamilton (1994) and Greene (2003) (for GMM), and Gourieroux and Monfort (1996) (for SMM and II). Applications of these procedures to DSGE models are discussed in Adda and Cooper (2003).

This chapter is organized as follows. In order to help establish intuition, section 2 presents each procedure in relatively general terms. Section 3 then discusses their implementation in the specific context of analyses involving DSGE models. Finally, section 4 presents an example application to the RBC model introduced in section 1 of chapter 3.

## 12.2 Implementation

### 12.2.1 The Generalized Method of Moments

Although moment-based estimation procedures have enjoyed a long tradition, their continued prominence stems from the work of Hansen

## 12.2 Implementation

(1982), who generalized their use to applications involving serially correlated stochastic processes, and established their asymptotic properties. Our presentation here closely follows his original development; as noted, in section 3 we then specialize the discussion of implementation in applications involving DSGE models.<sup>1</sup>

Let  $z_t$  denote an  $a \times 1$  vector of variables observed at time  $t$ , and let  $f(z_t, \theta)$  denote an  $r \times 1$  vector-valued function. The  $q \times 1$  vector of parameters  $\theta$  have a true but unknown value  $\theta_0$  that is to be estimated. When evaluated at  $\theta_0$ , the unconditional expectation of  $f(z_t, \theta)$  satisfies

$$E[f(z_t, \theta_0)] = 0, \quad (12.1)$$

which are known as orthogonality conditions. The idea behind GMM estimation is to choose parameter estimates  $\hat{\theta}$  such that the sample analogs of (12.1) are as close as possible to zero. Denote these analogues as

$$g(Z, \theta) = \frac{1}{T} \sum_{t=1}^T f(z_t, \theta), \quad (12.2)$$

where

$$Z = [z'_T, z'_{T-1}, \dots, z'_1]'$$

is a  $(T \times a) \times 1$  vector containing the observed data.

As stated, the specification of  $\hat{\theta}$  involves the minimization of the  $r$ -valued function  $g(Z, \theta)$  with respect to the  $q \times 1$  vector  $\theta$ . If  $q > r$ , then the possibility arises that multiple specifications of  $\hat{\theta}$  will yield  $g(Z, \hat{\theta}) = 0$ , and we are said to be faced with an identification problem. This problem is circumvented by augmenting  $f(z_t, \theta)$  so that  $r \geq q$ . For the case in which  $r = q$ , it may be possible to choose  $\hat{\theta}$  such that  $g(Z, \hat{\theta}) = 0$ . This is known as the just-identified case. Otherwise, when  $r > q$ , it will not be possible in general to achieve  $g(Z, \hat{\theta}) = 0$ . In this case, the objective is to choose  $\hat{\theta}$  to make  $g(Z, \hat{\theta})$  as close as possible to zero, in a manner made precise below. This is known as the overidentified case. The case of overidentification is of particular interest, because it enables a formal test of the null hypothesis that the orthogonality conditions are in fact consistent with the data generation process from which  $Z$  was realized.

<sup>1</sup> GAUSS code for performing the GMM estimation and testing procedures described here is available at [www.american.edu/academic.depts/cas/econ/gaussres/GMM/GMM.HTM](http://www.american.edu/academic.depts/cas/econ/gaussres/GMM/GMM.HTM). The code was prepared by Lars Hansen, John Heaton, and Masao Ogaki, and was designed specifically to estimate the asset-pricing model analyzed by Hansen and Singleton (1982). Details for adopting the code in pursuit of alternative applications is provided in supporting documentation available at this website.

Given  $r \geq q$ , the GMM estimate  $\hat{\theta}$  is obtained as the solution to a quadratic minimization problem of the form

$$\min_{\theta} \Gamma(\theta) = g(Z, \theta)' \times \Omega \times g(Z, \theta), \quad (12.3)$$

where  $\Omega$  is a positive-definite weighting matrix. (Techniques available for minimizing objective functions such as (12.3) are discussed in detail in chapter 13.) This weighting matrix determines the relative importance of the individual components of  $g(Z, \theta)$  in determining  $\hat{\theta}$ . Following Hansen (1982), the optimal weighting matrix  $\Omega^*$  is that which minimizes the asymptotic variance of  $\hat{\theta}$ . This optimal weighting matrix turns out to be given by the inverse of the variance-covariance matrix  $\Sigma$  associated with the sample mean  $g(Z, \theta)$ :  $\Omega^* = \Sigma^{-1}$ . Intuitively, the optimal weighting matrix serves to decrease the influence on  $\hat{\theta}$  of elements of  $g(Z, \theta)$  that provide relatively imprecise estimates of their population analogues  $Ef(z_t, \theta)$ . We now characterize a sequential procedure for constructing an estimate  $\hat{\Omega}^*$  of the optimal weighting matrix.

The asymptotic variance-covariance matrix of the sample mean  $g(Z, \theta)$  is given by

$$\Sigma = \lim_{T \rightarrow \infty} T E[g(Z, \theta_0)g(Z, \theta_0)'].$$

Notice that  $\Sigma$  is a function of the unknown value of  $\theta_0$ , and thus must be estimated. For the case in which  $f(z_t, \theta_0)$  is serially uncorrelated, a consistent estimator for  $\Sigma$  is given by

$$\hat{\Sigma} = \frac{1}{T} \sum_{t=1}^T f(z_t, \hat{\theta})f(z_t, \hat{\theta})', \quad (12.4)$$

where  $\hat{\theta}$  is any consistent estimate of  $\theta_0$ . For the case in which  $f(z_t, \theta_0)$  is serially correlated, a consistent estimator for  $\Sigma$  is given by

$$\hat{\Sigma} = \hat{\Gamma}_0 + \sum_{v=1}^s \left(1 - \frac{v}{s+1}\right) (\hat{\Gamma}_v + \hat{\Gamma}'_v), \quad (12.5)$$

where estimates of the  $v^{\text{th}}$ -order covariance matrices,  $v = 0, \dots, s$  are given by

$$\hat{\Gamma}_v = \frac{1}{T} \sum_{t=v+1}^T f(z_t, \hat{\theta})f(z_{t-v}, \hat{\theta})'.$$

This latter estimator for  $\Sigma$  is due to Newey and West (1987a).

## 12.2 Implementation

Notice that the estimates  $\hat{\Sigma}$  given in (12.4) and (12.5) are both functions of  $\hat{\theta}$ , which in turn is obtained using a weighting matrix ideally constructed as  $\Omega^* = \hat{\Sigma}^{-1}$ . Due to this circularity, an iterative procedure is typically used in constructing  $\hat{\theta}$  and  $\hat{\Sigma}$ . This begins by obtaining an initial estimate  $\hat{\theta}^0$  using an identity matrix for  $\Omega$  in (12.3), and then obtaining an estimate  $\hat{\Sigma}^0$  by inserting  $\hat{\theta}^0$  into either (12.4) or (12.5). Next,  $\hat{\theta}^1$  is derived by substituting  $[\hat{\Sigma}^0]^{-1}$  for  $\Omega$  in (12.3), and  $\hat{\Sigma}^1$  is derived by inserting  $\hat{\theta}^1$  into either (12.4) or (12.5). Repeating this process until the estimates  $\hat{\theta}$  and  $\hat{\Sigma}$  converge yields optimal GMM estimates of  $\theta_0$  and  $\Sigma$ .

If the orthogonality conditions (12.1) are in fact consistent with the data generation process from which  $Z$  was realized, then by the law of large numbers the limiting distribution of  $g(Z, \theta_0)$  is given by

$$\sqrt{T}g(Z, \theta_0) \rightarrow N(0, \Sigma). \quad (12.6)$$

In turn, under mild regularity conditions established by Hansen (1982), the limiting distribution of  $\hat{\theta}$  is given by

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow N(0, V), \quad (12.7)$$

where

$$V = (D\Sigma^{-1}D')^{-1},$$

and  $D'$  is a  $q \times r$  gradient matrix:

$$D' = \frac{\partial g(Z, \theta)}{\partial \theta'}, \quad (12.8)$$

with the  $(i, j)^{\text{th}}$  element containing the derivative of the  $i^{\text{th}}$  component of  $g(Z, \theta)$  with respect to the  $j^{\text{th}}$  component of  $\theta$ . Letting  $\hat{D}$  denote the evaluation of  $D$  obtained using  $\hat{\theta}$ ,  $V$  may be estimated using<sup>2</sup>

$$\hat{V} = (\hat{D}\hat{\Sigma}^{-1}\hat{D}')^{-1}. \quad (12.9)$$

Given  $\hat{\theta}$  and  $\hat{V}$ , tests of various sorts of parameter restrictions are straightforward to conduct. Consider the following three classes of restrictions. First, the statistical significance of the difference between an individual element  $\theta^j$  and a hypothesized value  $\bar{\theta}^j$  may be evaluated using the  $t$ -statistic

$$t^j = \frac{\hat{\theta}^j - \bar{\theta}^j}{\hat{\sigma}_{\theta^j}}, \quad (12.10)$$

where  $\hat{\sigma}_{\theta^j}$  is the standard error of  $\hat{\theta}^j$ , or the square root of the  $j^{\text{th}}$  diagonal element of  $\hat{V}$ . Under the null hypothesis that  $\hat{\theta}^j = \bar{\theta}^j$ ,  $t^j$  is distributed as

<sup>2</sup>The GAUSS command `gradp` is available for use in constructing  $D$ .

Student's  $t$  with  $T - 1$  degrees of freedom; large values of  $t^j$  are assigned low probabilities, or  $p$ -values, under this distribution. When  $p$ -values lie below a prespecified critical value, the null hypothesis is rejected; otherwise it is accepted.

Second, following Newey and West (1987b), tests of joint sets of restrictions imposed upon  $\theta$  may be conducted using GMM counterparts to likelihood-based tests. Consider a collection of restrictions imposed upon  $\theta$  of the form

$$\gamma(\theta) = 0, \quad (12.11)$$

where  $\gamma(\theta)$  is an  $l$ -valued differentiable function, with  $l < q$ . Then letting  $\widehat{\theta}_R$  denote the value of  $\theta$  that minimizes  $\Gamma(\theta)$  subject to (12.11), the null hypothesis that the restrictions in fact hold can be tested using the GMM analogue to the likelihood ratio statistic:

$$LR = \Gamma(\widehat{\theta}_R) - \Gamma(\widehat{\theta}). \quad (12.12)$$

Under the null,  $LR$  is distributed as  $\chi^2(l)$ . As the gap between the restricted and unrestricted objective functions widens,  $LR$  increases, and its associated  $p$ -value falls. As with the  $t$ -test, if the  $p$ -value falls below a prespecified critical value the null is rejected; otherwise it is accepted. Regarding the specification chosen for  $\Omega$  in estimating  $\widehat{\theta}_R$ , this should be the optimal weighting matrix  $\widehat{\Sigma}^{-1}$  used to estimate  $\widehat{\theta}$ , because  $\widehat{\theta}$  is a consistent estimator of  $\theta_0$  regardless of whether the restrictions (12.11) in fact hold. Finally, regarding implementation, consider the case in which (12.11) can be rewritten as

$$\theta = r(\omega), \quad (12.13)$$

where  $\omega$  is an  $\ell$ -vector such that  $\ell = q - l$ . Then  $\Gamma(\widehat{\theta}_R)$  can be calculated by optimizing over  $\omega$ , and setting  $\widehat{\theta}_R = r(\widehat{\omega})$ .

Third, consider a test designed to assess the stability of  $\theta$  over the sample period. Following Andrews and Fair (1988), this can be conducted as follows. Let  $T_0$  denote a potential date at which  $\theta$  undergoes a structural change. Also, let  $(\widehat{\theta}_1, \widehat{V}_1)$  and  $(\widehat{\theta}_2, \widehat{V}_2)$  denote estimates of  $(\theta, V)$  obtained over the subsamples  $[1, T_0]$  and  $[T_0 + 1, T]$ . Finally, let  $\pi = T_0/T$  denote the fraction of observations contained in the first subsample. Then the null hypothesis of structural stability (i.e.,  $\theta_1 = \theta_2$ ) can be evaluated using a Wald test:

$$\lambda = T(\widehat{\theta}_1 - \widehat{\theta}_2)'[\pi^{-1}\widehat{V}_1 + (1 - \pi^{-1})\widehat{V}_2]^{-1}(\widehat{\theta}_1 - \widehat{\theta}_2). \quad (12.14)$$

Under the null,  $\lambda$  is distributed as  $\chi^2(q)$ ; once again, the null is rejected if the  $p$ -value associated with  $\lambda$  falls below a prespecified critical value.

## 12.2 Implementation

However, if a systematic search is conducted over a range of potential break dates, then the test statistic associated with any one potential break date (e.g., the date at which the largest value of  $\lambda$  is obtained) will no longer be appropriately interpreted as being distributed as  $\chi^2(q)$ . Andrews (1993) reports relevant distributions for  $\lambda$  in this case.

For the case in which  $r > q$ , so that the model is overidentified, Hansen (1982) derives a test of the null hypothesis that the orthogonality conditions (12.1) are consistent with the data generation process from which  $Z$  was obtained. Recall that under the null,  $\sqrt{T}g(Z, \theta_0)$  is asymptotically distributed as  $N(0, \Sigma)$ . Thus replacing  $(\theta_0, \Sigma)$  with their estimated values  $(\widehat{\theta}, \widehat{\Sigma})$ , under the null the test statistic

$$J = [\sqrt{T}g(Z, \widehat{\theta})]' \widehat{\Sigma}^{-1} [\sqrt{T}g(Z, \widehat{\theta})] \quad (12.15)$$

is distributed as  $\chi^2(r - q)$ , where note that the degrees of freedom are given by the difference between the number of moment conditions and the number of parameters determined by these conditions.

To further enhance the intuition behind the GMM approach to estimation, consider the following three examples. First, following Hamilton (1994), consider the implementation of GMM in the context of an ordinary least squares (OLS) regression model. Specifically, consider a regression model of the form

$$y_t = w_t' \beta + \varepsilon_t, \quad (12.16)$$

where  $y_t$  is a single time series and  $w_t$  is a  $d \times 1$  matrix of explanatory variables. Moment conditions for this model are derived under the assumption that the residuals  $\{\varepsilon_t\}$  are orthogonal to the explanatory variables:  $E[w_t \varepsilon_t] = 0$ . Under this assumption, the true value  $\beta_0$  satisfies the  $d \times 1$  system of equations

$$E[w_t(y_t - w_t' \beta_0)] = 0. \quad (12.17)$$

Thus the analogue for  $f(z_t, \theta_0)$  in (12.1) in the linear regression context is  $w_t(y_t - w_t' \beta_0)$ . Notice that  $\beta$  is a  $d \times 1$  vector, so that we are working in the context of the just-identified case. Therefore, the GMM estimate  $\widehat{\beta}$  is obtained as the vector of parameters that sets the sample average of (12.17) equal to zero:

$$g(Z, \beta) = \frac{1}{T} \sum_{t=1}^T w_t(y_t - w_t' \widehat{\beta}) = 0, \quad (12.18)$$

where

$$Z = [y_T \sim w_T', y_{T-1} \sim w_{T-1}', \dots, y_1 \sim w_1']'$$

Rearranging (12.18) and solving for  $\hat{\beta}$  yields the familiar least-squares expression

$$\hat{\beta} = \left[ \sum_{t=1}^T w_t w_t' \right]^{-1} \left[ \sum_{t=1}^T w_t y_t \right]. \quad (12.19)$$

Proceeding under the assumption that  $\{\varepsilon_t\}$  is serially uncorrelated, the estimate  $\hat{\Sigma}$  is given by

$$\begin{aligned} \hat{\Sigma} &= \frac{1}{T} \sum_{t=1}^T [w_t(y_t - w_t' \hat{\beta})][w_t(y_t - w_t' \hat{\beta})]' \\ &= \hat{\sigma}^2 \frac{1}{T} \sum_{t=1}^T w_t w_t', \end{aligned} \quad (12.20)$$

where

$$\hat{\sigma}^2 = \frac{1}{T} \sum_{t=1}^T (y_t - w_t' \hat{\beta})(y_t - w_t' \hat{\beta})'.$$

Finally,

$$\begin{aligned} \hat{D}' &= \frac{\partial}{\partial \hat{\beta}'} \frac{1}{T} \sum_{t=1}^T w_t(y_t - w_t' \hat{\beta}) \\ &= \frac{1}{T} \sum_{t=1}^T w_t w_t', \end{aligned} \quad (12.21)$$

and thus

$$\begin{aligned} \hat{V} &= \left[ \left( \frac{1}{T} \sum_{t=1}^T w_t w_t' \right) \left( \hat{\sigma}^2 \frac{1}{T} \sum_{t=1}^T w_t w_t' \right)^{-1} \left( \frac{1}{T} \sum_{t=1}^T w_t w_t' \right) \right]^{-1} \\ &= \hat{\sigma}^2 \left( \frac{1}{T} \sum_{t=1}^T w_t w_t' \right)^{-1}. \end{aligned}$$

### Exercise 12.1

As a second example, in the context of the regression model (12.16), suppose that the orthogonality assumption  $E(w_t \varepsilon_t) = 0$  is inappropriate, due to the presence of endogenous variables included among the elements of  $w_t$ . In this

case, assume the availability of a  $d \times 1$  collection of instrumental variables  $u_t$  that are correlated with  $w_t$  but orthogonal to  $\varepsilon_t$ . Using the collection of  $d$  orthogonality conditions  $E[u_t(y_t - w_t' \hat{\beta}_0)] = 0$ , obtain the corresponding GMM estimates  $(\hat{\beta}, \hat{V})$ , once again under the assumption that  $\{\varepsilon_t\}$  is serially uncorrelated. The estimates you obtain are known as instrumental variable (IV) estimates.

As a third example, consider the case in which (12.1) specializes to

$$E[f(z_t, \theta_0)] = E \left\{ \begin{array}{l} [f_1(w_t, \theta_0)] u_t \\ [f_2(w_t, \theta_0)] u_t \\ \vdots \\ [f_m(w_t, \theta_0)] u_t \end{array} \right\} = 0, \quad (12.22)$$

where  $u_t$  is an  $\ell \times 1$  vector of instruments. In this case,

$$z_t = [w_t' u_t']'$$

and the number of orthogonality conditions is  $r = m\ell$ . Although a specialization of (12.1), (12.22) is nevertheless quite general, and in fact encompasses the full range of applications to analyses involving DSGE models discussed in section 3.

We conclude this subsection with three comments, made in the context of GMM applications cast in the form of (12.22). First, cases in which the number of orthogonality conditions  $r$  exceeds  $q$  (the dimensionality of the parameter vector  $\theta$ ) are in no way guaranteed to be free from identification problems. Symptoms of such problems include the attainment of large standard errors associated with particular elements of  $\theta$ . The presence of large standard errors indicates that the objective function is relatively insensitive to alternative specifications of  $\theta$  along certain dimensions, and thus that a wide range of alternative parameterizations represent roughly equally plausible estimates of  $\theta_0$ . Potential remedies for this problem involve adjustments to  $\{f_i(w_t, \theta_0)\}_{i=1}^m$  and/or  $u_t$ , the goal being to produce orthogonality conditions that are more informative regarding the parameterization of  $\theta$ . We caution, though, that remedies simply involving expansions of the instruments included in  $u_t$  can cause additional problems, as Tauchen (1986b) and Kocherlakota (1990) have illustrated. In particular, applying GMM estimation and testing procedures to artificial data simulated from asset-pricing models, they show that the small-sample performance of estimators and tests based on parsimonious specifications of  $u_t$  is superior to tests based on profligate specifications, in terms of both the accuracy of parameter estimates and the rejection rates obtained in applications of the  $J$  test in (12.15). Further discussion of identification problems, cast in the context of maximum-likelihood estimation, is provided in chapter 13.

Second, depending upon the specific context of the empirical exercise, the researcher will often have considerable leeway regarding the specification of (12.22), in terms of both  $\{f_i(w_t, \theta_0)\}_{i=1}^m$  and  $u_t$ . The results of the exercise may depend sensitively upon the particular specification that has been chosen. For example, in the context of the specification of the regression model (12.16), coefficient estimates associated with particular elements of  $w_t$  may depend sensitively upon the inclusion or exclusion of additional variables in  $w_t$ . As we shall see in section 3, it is often the case in working with DSGE models that the specification of  $\{f_i(w_t, \theta_0)\}_{i=1}^m$  is determined rather directly by the particular question the model has been designed to address, or by the specification of the model itself. In addition, theoretical considerations can also be helpful in specifying  $u_t$ . However, guidance from theory rarely obviates the need for sensitivity analysis.

Third, we have noted that because GMM estimates are not based upon distributional assumptions regarding random disturbances included in the model, they are not subject to misspecification along this dimension. However, they are of course sensitive to the validity and effectiveness of  $u_t$  as instruments. Simultaneity between  $f_i(w_t, \theta_0)$  and individual elements of  $u_t$  has long been known to produce bias in the resulting estimates of  $\theta$ , but bias can also result from the use of instruments that are weakly correlated with the elements of  $w_t$ , a point made dramatically by Nelson and Startz (1990a, 1990b) (for a survey of problems associated with weak instruments, see Stock, Wright, and Yogo, 2002). Once again, guidance from theory is often helpful in identifying appropriate specifications of  $u_t$ , particularly in applications involving DSGE models; but to repeat: this guidance rarely obviates the need for sensitivity analysis.

### 12.2.2 The Simulated Method of Moments

In certain applications, it may be the case that orthogonality conditions associated with the model under investigation cannot be assessed analytically. In such cases, the work of McFadden (1989), Pakes and Pollard (1989), Lee and Ingram (1991) and Duffie and Singleton (1993) has demonstrated that moment-matching estimation and testing procedures based on model simulations retain asymptotic characteristics akin to those based on analytical calculations. Here we present a straightforward means of implementing so-called SMM procedures. For a general textbook presentation, see Gourieroux and Monfort (1996).

Consider a collection of empirical targets contained in the  $r \times 1$  vector  $b(z_t)$ , where as in section 2.1  $z_t$  denotes an  $a \times 1$  vector of variables observed at time  $t$ . Let  $y_t$  denote an  $a \times 1$  vector of model variables that correspond directly to their empirical counterparts  $z_t$ , and  $b(y_t, \theta)$  denote the theoretical counterparts to  $b(z_t)$ . Under GMM,  $E[b(y_t, \theta)]$  is

### 12.2 Implementation

calculated analytically; under SMM this need not be the case. Instead, all that is required is the ability to simulate  $\{y_t\}$  using a parameterization of the structural model under investigation (model simulation is discussed in section 3). Proceeding under the usual assumption that the model has been specified for stationary versions of the data, the sample average

$$b(Y, \theta) = \frac{1}{N} \sum_{i=1}^N b(y_t, \theta), \quad (12.23)$$

where

$$Y = [y'_N, y'_{N-1}, \dots, y'_1],$$

is an asymptotically consistent estimator of  $E[b(y_t, \theta)]$ . The idea behind SMM estimators is to use sample averages in place of population averages in evaluating implications of the structural model regarding  $b(y_t, \theta)$ .

SMM estimation proceeds under the hypothesis that there exists a parameterization of the structural model  $\theta_0$  under which

$$Eb(z_t) = Eb(y_t, \theta_0); \quad (12.24)$$

the objective is to choose parameter estimates  $\hat{\theta}$  that come as close as possible to satisfying this equality. In terms of the notation used in discussing the GMM estimator,  $f(z_t, \theta_0)$  in (12.1) is given by

$$f(z_t, \theta_0) = b(z_t) - b(y_t, \theta_0), \quad (12.25)$$

and the sample analogue  $g(Z, \theta)$  in (12.2) is given by

$$g(Z, \theta) = \left( \frac{1}{T} \sum_{t=1}^T b(z_t) \right) - \left( \frac{1}{N} \sum_{i=1}^N b(y_t, \theta) \right). \quad (12.26)$$

Having made these adjustments to  $f(z_t, \theta_0)$  and  $g(Z, \theta)$ , estimation and hypothesis testing proceed as described in section 2.1, with a minor modification needed for the weighting matrix  $\Omega$  in (12.3). The optimal weighting matrix  $\Omega^*$  is once again given by the inverse of the variance-covariance matrix  $\Sigma$  associated with the sample mean of  $g(Z, \theta)$ . Under the null hypothesis (12.24), the sample means of  $b(z_t)$  and  $b(y_t, \theta)$  in (12.26) are identical. Moreover, the empirical and artificial samples  $\{z_t\}$  and  $\{y_t\}$  are independent. Thus because  $g(Z, \theta)$  is a linear combination of independent sample means, its variance-covariance matrix is given by

$$\Sigma(1 + T/N).$$

Note that as the artificial sample size  $N$  increases relative to  $T$ , the covariance matrix of  $g(Z, \theta)$  converges to its GMM counterpart. This is true

because as  $N$  increases, the influence of the numerical error associated with the use of a sample average to approximate  $E[b(y_t, \theta)]$  decreases. See chapter 9 for a discussion of numerical errors. For the reasons outlined there, the specification of  $N = 10,000$  is standard, because it is associated with a numerical standard error of 1% of the standard error associated with the simulated moment.

With the adjustment

$$\Omega^* = \Sigma(1 + T/N),$$

estimation and testing via GMM and SMM proceed identically. To briefly summarize, the objective function used to estimate  $\hat{\theta}$  is given by

$$\min_{\theta} \Gamma(\theta) = g(Z, \theta)' \times [\Sigma(1 + T/N)]^{-1} \times g(Z, \theta),$$

where  $\Sigma$  is given by either

$$\Sigma = \frac{1}{T} \sum_{t=1}^T f(z_t, \theta) f(z_t, \theta)'$$

or

$$\Sigma = \Gamma_0 + \sum_{v=1}^s \left(1 - \frac{v}{s+1}\right) (\Gamma_v + \Gamma_v'),$$

$$\Gamma_v = \frac{1}{T} \sum_{t=v+1}^T f(z_t, \theta) f(z_{t-v}, \theta)'.$$

The circular relationship between  $\theta$  and  $\Sigma$  once again necessitates an iterative estimation procedure under which an initial estimate  $\hat{\theta}^0$  is obtained using an identity matrix for  $\Sigma$ ,  $\hat{\theta}^0$  is used to construct  $\hat{\Sigma}^0$ , and the process is repeated until the estimates  $\hat{\theta}$  and  $\hat{\Sigma}$  converge.

Under mild regularity conditions, the limiting distribution of  $\hat{\theta}$  is once again given by

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow N(0, V),$$

where

$$V = (D\Sigma^{-1}D')^{-1},$$

and  $D'$  is a  $q \times r$  gradient matrix:

$$D' = \frac{\partial g(Z, \theta)}{\partial \theta'}.$$

## 12.2 Implementation

With  $\hat{D}$  denoting the evaluation of  $D$  obtained using  $\hat{\theta}$ ,  $V$  is estimated using

$$\hat{V} = (\hat{D}\hat{\Sigma}^{-1}\hat{D}')^{-1}.$$

Finally, the null hypothesis (12.24) can be evaluated using a modification of Hansen's (1982)  $J$  statistic:

$$J = [\sqrt{T}g(Z, \hat{\theta})]'[\hat{\Sigma}(1 + T/N)]^{-1}[\sqrt{T}g(Z, \hat{\theta})],$$

which is distributed asymptotically as  $\chi^2(r - q)$ .

### 12.2.3 Indirect Inference

Although it is becoming increasingly common to address specific empirical questions using structural models, reduced-form models remain a workhorse for empirical practitioners. The use of structural specifications to account for and help interpret findings obtained using reduced-form models is the objective of the indirect inference methodology, which is the subject of this sub-section. The goal of indirect inference is to determine whether data simulated from a given structural model can be used to replicate estimates obtained using actual data in a given reduced-form analysis. If so, then a structural interpretation for the reduced-form phenomenon can be offered. Note that indirect inference amounts to a moment-matching exercise, in which estimates obtained in the reduced-form analysis serve as target moments. An attractive feature of this approach to structural estimation is that clear guidance regarding relevant moments is available a priori: this is provided by the reduced-form exercise for which a structural interpretation is sought. The method of indirect inference was developed by Gourieroux, Monfort, and Renault (1993) and Smith (1993); a general textbook presentation is available in Gourieroux and Monfort (1996).

Once again, let  $z_t$  and  $y_t$  denote  $a \times 1$  vectors of actual and model variables. In this case, the empirical target is an  $r \times 1$  vector of reduced-form parameters  $\delta$ . To match the notation used in characterizing SMM, let  $\delta(z_t)$  denote the true value of  $\delta$ , and  $\delta(y_t, \theta)$  its theoretical counterpart (Gourieroux, Monfort, and Renault (1993) refer to  $\delta(y_t, \theta)$  as the binding function). Indirect inference proceeds under the hypothesis that there exists a parameterization of the structural model  $\theta_0$  under which

$$\delta(z_t) = \delta(y_t, \theta_0). \quad (12.27)$$

The sample analogue of  $\delta(z_t)$ , denoted as  $\delta(Z)$ , is obtained by estimating  $\delta$  using the actual data  $Z = [z'_T, z'_{T-1}, \dots, z'_1]'$ :

$$\delta(Z) = \arg \max_{\delta} \Delta(Z, \delta). \quad (12.28)$$

In (12.28),  $\Delta(Z, \delta)$  represents the objective function used to estimate  $\delta$ . For example,  $\Delta(Z, \delta)$  could represent the likelihood function associated with the reduced-form model under investigation, in which case  $\delta(Z)$  would represent the maximum-likelihood estimate of  $\delta$ . Alternatively,  $\Delta(Z, \delta)$  could represent the negative of quadratic objective functions associated with OLS, IV, GMM, or SMM estimators. In turn, let  $\delta(Y, \theta)$  denote an estimate of  $\delta$  obtained using a simulated realization of artificial data  $Y = [y'_T, y'_{T-1}, \dots, y'_1]'$ :

$$\delta(Y, \theta) = \arg \max_{\delta} \Delta(Y, \delta). \quad (12.29)$$

Note that the sample sizes of  $Y$  and  $Z$  are identical. Letting  $Y^j$  denote the  $j^{\text{th}}$  of  $S$  realizations of  $Y$ , the sample analogue of  $\delta(y_t, \theta)$ , denoted as  $\delta_S(Y, \theta)$ , is obtained by averaging over  $\delta(Y^j, \theta)$ :

$$\delta_S(Y, \theta) = \frac{1}{S} \sum_{j=1}^S \delta(Y^j, \theta). \quad (12.30)$$

Given these sample analogues,  $g(Z, \theta)$  is in this case given by

$$g(Z, \theta) = \delta(Z) - \delta_S(Y, \theta). \quad (12.31)$$

In turn, the objective function used to obtain the estimate  $\hat{\theta}$  is given by

$$\begin{aligned} \min_{\theta} \Gamma(\theta) &= g(Z, \theta)' \times \Omega^* \times g(Z, \theta) \\ &= [\delta(Z) - \delta_S(Y, \theta)]' \times \Omega^* \times [\delta(Z) - \delta_S(Y, \theta)]. \end{aligned} \quad (12.32)$$

Once again,  $\Omega^*$  denotes the optimal weighting matrix given by the inverse of the variance-covariance matrix  $\Sigma$  associated with  $g(Z, \theta)$ . We now turn to a characterization of this matrix.

Under certain regularity conditions Gourieroux, Monfort, and Renault (1993) demonstrate that the indirect inference estimator is consistent and asymptotically Normal (for fixed  $S$ , as  $T \rightarrow \infty$ ):

$$\sqrt{T}(\hat{\theta} - \theta_0) \rightarrow N(0, W(S, \Omega)).$$

The asymptotic variance-covariance matrix  $W(S, \Omega)$  is given by

$$\begin{aligned} &\left(1 + \frac{1}{S}\right) \left( \frac{\partial \delta(y_t, \theta)}{\partial \theta} \Omega \frac{\partial \delta(y_t, \theta)}{\partial \theta'} \right)^{-1} \frac{\partial \delta(y_t, \theta)}{\partial \theta} \Omega J_0^{-1} (I_0 - K_0) J_0^{-1} \Omega \\ &\times \frac{\partial \delta(y_t, \theta)}{\partial \theta'} \left( \frac{\partial \delta(y_t, \theta)}{\partial \theta} \Omega \frac{\partial \delta(y_t, \theta)}{\partial \theta'} \right)^{-1}, \end{aligned} \quad (12.33)$$

## 12.2 Implementation

where

$$J_0 = p \lim - \frac{\partial^2 \Delta(Y, \delta)}{\partial \delta \partial \delta'} \quad (12.34)$$

$$I_0 = \lim \text{Var} \left[ \sqrt{T} \frac{\partial \Delta(Y, \delta)}{\partial \delta} \right] \quad (12.35)$$

$$K_0 = \lim \text{Var} \left[ E \left( \sqrt{T} \frac{\partial \Delta(Z, \delta)}{\partial \delta} \right) \right]. \quad (12.36)$$

This expression simplifies considerably when evaluated using the optimal weighting matrix  $\Omega^*$ , which once again is defined as that which minimizes  $W$ . This is given by

$$\Omega^* = J_0 (I_0 - K_0)^{-1} J_0. \quad (12.37)$$

Substituting  $\Omega^*$  into (12.33), the asymptotic variance-covariance matrix  $W(S, \Omega)$  becomes

$$W(S, \Omega) = \left(1 + \frac{1}{S}\right) \left[ \left( \frac{\partial^2 \Delta(Y, \delta)}{\partial \theta \partial \theta'} \right) (I_0 - K_0)^{-1} \left( \frac{\partial^2 \Delta(Y, \delta)}{\partial \delta \partial \delta'} \right) \right]^{-1}. \quad (12.38)$$

Finally,  $J_0$  may be estimated using

$$\hat{J}_0 = - \frac{\partial^2 \Delta(Y, \delta_S(Y, \hat{\theta}))}{\partial \delta \partial \delta'}, \quad (12.39)$$

$\frac{\partial^2 \Delta(Y, \delta)}{\partial \theta \partial \theta'}$  may be estimated using

$$\frac{\partial^2 \Delta(Y, \delta_S(Y, \hat{\theta}))}{\partial \theta \partial \theta'}, \quad (12.40)$$

and  $I_0 - K_0$  may be estimated using

$$\begin{aligned} \hat{\Xi} &= \frac{T}{S} \sum_{s=1}^S \left( \frac{\partial \Delta(Y, \delta_S(Y, \hat{\theta}))}{\partial \delta} - \frac{1}{S} \sum_{s=1}^S \frac{\partial \Delta(Y, \delta_S(Y, \hat{\theta}))}{\partial \delta} \right) \\ &\times \left( \frac{\partial \Delta(Y, \delta_S(Y, \hat{\theta}))}{\partial \delta} - \frac{1}{S} \sum_{s=1}^S \frac{\partial \Delta(Y, \delta_S(Y, \hat{\theta}))}{\partial \delta_s} \right)', \end{aligned} \quad (12.41)$$

Once again, we face a circularity involving  $\theta$  and  $\Omega^*$ . A by-now familiar algorithm for dealing with this circularity is as follows. Specify an identity matrix for  $\Omega^0$ , and estimate  $\theta^0$  using (12.32). Next, use  $\theta^0$  to

construct  $\delta_S(Y, \theta^0)$  using (12.30). Finally, use  $\delta_S(Y, \theta^0)$  to construct  $\hat{J}_0^0$  using (12.39) and  $\hat{\Xi}^0$  using (12.41); substitute these estimates into (12.37) to obtain  $\Omega^{*0}$ . Iterate over these steps until  $\hat{\theta}$  converges, and use the final estimate to construct  $W(S, \Omega^*)$  using (12.40) and (12.41).

Having obtained  $\hat{\theta}$  and  $\Omega^*$ , a  $J$ -type test can be used to assess the null hypothesis (12.27). The test statistic in this case is given by

$$J = [\delta(Z) - \delta_S(Y, \hat{\theta})]' \times \Omega^* \times [\delta(Z) - \delta_S(Y, \hat{\theta})], \quad (12.42)$$

which is once again distributed asymptotically as  $\chi^2(r - q)$ .

## 12.3 Implementation in DSGE Models

### 12.3.1 Analyzing Euler Equations

We now discuss four contexts in which the estimation procedures described in section 2 can be used to analyze DSGE models. The first concerns applications involving the set of Euler equations associated with the theoretical model under investigation. Such applications were pioneered by Hansen and Singleton (1982), working in the context of the multi-asset pricing environment introduced in section 3 of chapter 3. For a survey of the enormous literature spawned by this work, see Hansen and West (2002).

As we have seen in Part I the Euler equations associated with a given DSGE model take the form of a collection of first-order nonlinear expectation difference equations. Let these equations be represented by the  $m \times 1$  system

$$E_t \Psi(x_{t+1}, x_t) = 0, \quad (12.43)$$

where the full set of model variables is contained in the  $n \times 1$  vector  $x_t$ , and  $E_t$  is the expectations operator conditional on information available to the model's decision makers at time  $t$ . Assuming that this information includes the collection of variables contained in the  $\ell \times 1$  vector of instruments  $u_t$ , then it will also be the case that

$$E\{[\Psi(x_{t+1}, x_t)]u_t\} = 0, \quad (12.44)$$

where  $E\{\cdot\}$  denotes the unconditional expectations operator. Note that this expression is precisely in terms of the specialization of the GMM orthogonality conditions given in (12.22). Implementation of the GMM estimator in this context is therefore straightforward.

Consider as an example the asset-pricing model analyzed by Mehra and Prescott (1985). Letting  $r^e$  and  $r^f$  denote returns associated with a risky

### 12.3 Implementation in DSGE Models

and risk-free asset, and  $c$  denote consumption, the Euler equations under consideration are given by

$$\beta E_t \left\{ \frac{u'(c_{t+1})}{u'(c_t)} r_{t+1}^f - 1 \right\} = 0 \quad (12.45)$$

$$E_t \left\{ \frac{u'(c_{t+1})}{u'(c_t)} [r_{t+1}^e - r_{t+1}^f] \right\} = 0. \quad (12.46)$$

Under the assumption that preferences follow the CRRA specification

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma},$$

the Euler equations become

$$E_t \left\{ \beta \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} r_{t+1}^f - 1 \right\} = 0 \quad (12.47)$$

$$E_t \left\{ \left( \frac{c_{t+1}}{c_t} \right)^{-\gamma} [r_{t+1}^e - r_{t+1}^f] \right\} = 0. \quad (12.48)$$

In terms of the notation in (12.22),

$$w_t = [c_t/c_{t-1}, r_t^f, r_t^e]'$$

and

$$\theta = [\gamma, \beta]'$$

Valid instruments include a constant, and lagged values of consumption growth and returns. For an early example of an application of this estimation strategy in the context of an RBC framework, see Christiano and Eichenbaum (1992).

### 12.3.2 Analytical Calculations Based on Linearized Models

The second context involves applications to fully specified linearized models. Here, we discuss situations in which implications conveyed by the model regarding the selected empirical targets can be calculated analytically. Recall that the state-space representation we use in this context is given by

$$x_{t+1} = F(\mu)x_t + e_{t+1}, \quad E(e_{t+1}e'_{t+1}) = Q(\mu), \quad (12.49)$$

where the  $k \times 1$  vector  $\mu$  contains the model's structural parameters. One of two measurement equations maps  $x_t$  into observable variables  $X_t$ :

$$X_t = H(\mu)'x_t, \quad (12.50)$$

or

$$X_t = H(\mu)'x_t + u_t, \quad E(u_t u_t') = \Sigma_u, \quad (12.51)$$

where  $u_t$  represents measurement error. Hereafter, the dependence of  $F, Q, H$ , and  $\Sigma_u$  on  $\mu$  is suppressed to simplify notation.

Given this representation, a wide range of empirical targets can be calculated analytically. Chapter 7 described three sets of targets that have received considerable attention in the context of analyzing DSGE models: covariance terms (e.g., Kydland and Prescott, 1982); spectra (e.g., Watson, 1993; Diebold, Ohanian, and Berkowitz, 1998); and impulse response functions (e.g., Christiano, Eichenbaum, and Evans, 2005). The calculation of these targets using vector autoregressive (VAR) representations estimated using actual data was discussed in chapter 7; here we briefly discuss their construction for the theoretical analogues contained in  $X_t$ .

Regarding covariance terms, let

$$\Gamma(0) = E(x_t x_t')$$

denote the contemporaneous variance-covariance matrix of  $x_t$ , and

$$\Gamma(s) = E(x_t x_{t-s}')$$

the  $s^{\text{th}}$ -order covariance matrix. From (12.49),

$$\begin{aligned} \Gamma(0) &= E[(F x_{t-1} + e_t)(F x_{t-1} + e_t)'] \\ &= F\Gamma(0)F' + Q, \end{aligned} \quad (12.52)$$

the solution to which is given by

$$\text{vec}[\Gamma(0)] = [I - F \otimes F]^{-1} \text{vec}[Q], \quad (12.53)$$

where  $\otimes$  denotes the Kronecker product. Further,

$$\begin{aligned} \Gamma(1) &= E(x_t x_{t-1}') \\ &= E((F x_{t-1} + e_t)x_{t-1}') \\ &= F\Gamma(0), \end{aligned}$$

and in general,

$$\begin{aligned} \Gamma(s) &= F\Gamma(s-1) \\ &= F^s\Gamma(0). \end{aligned} \quad (12.54)$$

### 12.3 Implementation in DSGE Models

Thus for the  $m \times 1$  vector of observable variables  $X_t$ , we have

$$\begin{aligned} \Gamma_X(0) &= E(X_t X_t') \\ &= E[(H' x_t + u_t)(H' x_t + u_t)'] \\ &= H'\Gamma(0)H + \Sigma_u, \end{aligned} \quad (12.55)$$

where  $\Sigma_u = 0$  in the absence of measurement error. And since  $\{u_t\}$  is serially uncorrelated,

$$\begin{aligned} \Gamma_X(s) &= E X_t X_{t-s}' \\ &= H'\Gamma(s)H. \end{aligned}$$

Regarding spectra, let  $S(\omega)$  denote the spectral matrix of  $x_t$  evaluated at frequency  $\omega$ . That is, the  $j^{\text{th}}$  diagonal element of  $S(\omega)$  contains the spectrum of the  $j^{\text{th}}$  variable of  $x_t$  at frequency  $\omega$ ,  $s_j(\omega)$ , and the  $(i,j)^{\text{th}}$  element of  $S(\omega)$  contains the cross spectrum between the  $i^{\text{th}}$  and  $j^{\text{th}}$  variables of  $x_t$ ,  $s_{ij}(\omega)$ . Then from Hannan (1970),  $S(\omega)$  is given by

$$S(\omega) = \frac{1}{2\pi} [(I - Fe^{-i\omega})' Q^{-1} (I - Fe^{-i\omega})]^{-1}, \quad (12.56)$$

where  $I$  is the  $n \times n$  identity matrix. And since  $\{u_t\}$  is serially uncorrelated, the spectral matrix of  $X_t$  is proportional to

$$S_X(\omega) = H' S(\omega) H. \quad (12.57)$$

Regarding impulse response functions, given an initial impulse  $e_0$ , the responses of  $x_t$  are given by  $x_0 = e_0$ ,  $x_1 = Fe_0$ ,  $x_2 = F^2 e_0$ , and so on. The corresponding  $j$ -step response of the observables is

$$X_j = H' F^j e_0.$$

Recall that the relationship between  $e_t$  and the structural shocks  $v_t$  is given by

$$e_t = Gv_t;$$

thus the  $j$ -step response to a structural innovation  $v_0$  is given by

$$X_j = H' F^j G v_0.$$

When using impulse response functions as an empirical target, a major challenge involves the identification of an empirical counterpart to  $v_0$ . To illustrate this point, consider a  $p^{\text{th}}$ -order VAR specified for the  $m \times 1$  vector

of observations of actual data  $z_t$  that corresponds with the  $m \times 1$  vector of model variables  $X_t$ :

$$z_t = \Upsilon_1 z_{t-1} + \Upsilon_2 z_{t-2} + \cdots + \Upsilon_p z_{t-p} + \varepsilon_t, \quad E\varepsilon_t \varepsilon'_t = \Sigma_\varepsilon \quad (12.58)$$

$$= \Upsilon_1 z_{t-1} + \Upsilon_2 z_{t-2} + \cdots + \Upsilon_p z_{t-p} + Pv_t, \quad (12.59)$$

$$PP' = \Sigma_\varepsilon, \quad Ev_t v'_t = I, \quad (12.60)$$

where  $\Upsilon_j$  is  $k \times k$ , and  $P$  is a lower-triangular matrix that accomplished the Cholesky decomposition  $PP' = \Sigma_\varepsilon$ . As illustrated in chapter 7, it is straightforward to construct impulse response functions based upon this specification for a given orthogonalized impulse  $v_0$ . The challenge is to construct  $v_0$  in such a way that it is credibly identified with its structural counterpart  $v_0$ . The quest for such an identification involves the construction of what are known as identified VARs, the use of which dates to Sims (1980); see Watson (1994) for a survey, and Hamilton (1994) for a textbook discussion.

As Fernandez-Villaverde et al. (2005) note with understatement, “The process of reverse engineering a subset of economic shocks from the innovations to a VAR is known to be fraught with hazards” (3). However, they have developed a means of determining conditions under which this reverse engineering process can be accomplished. In fact, their procedure illustrates conditions under which a mapping from structural parameters  $\mu$  to the VAR parameters (12.58) can be induced. A related procedure for achieving reverse engineering, or as they characterize it, for solving the “invertibility problem,” has been developed by Sims and Zha (2006).

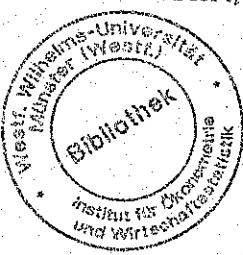
Here we provide a sketch of Fernandez-Villaverde et al.’s reverse engineering procedure. To do so, we must modify the notation used to represent the linearized DSGE model. Let  $s_t$  denote the collection of state variables included in the model, and  $w_t = [v'_t, u'_t]'$  contain the structural shocks and measurement errors associated with the observation of  $X_t$ . Then the representation we seek is given by

$$s_{t+1} = As_t + Bw_t, \quad (12.61)$$

$$X_t = Cs_t + Dw_t. \quad (12.62)$$

Once again, the dependence of  $[A, B, C, D]$  upon  $\mu$  is suppressed to simplify the notation. Measurement errors play no role in determining the law of motion for  $s_t$ , and thus

$$Bw_t = [B_1 \ 0] \begin{bmatrix} v_t \\ u_t \end{bmatrix}.$$



### 12.3 Implementation in DSGE Models

In contrast, both structural shocks and measurement error may play a role in determining the observable variables  $X_t$ , and thus in general,

$$Dw_t = [D_1 \ D_2] \begin{bmatrix} v_t \\ u_t \end{bmatrix}.$$

Now, when  $D$  is square and  $D^{-1}$  exists, we can rearrange (12.62) as

$$w_t = D^{-1}(X_t - Cs_t).$$

Substituting for  $w_t$  in (12.61) then yields

$$s_{t+1} = (A - BD^{-1}C)s_t + BD^{-1}X_t.$$

Representing  $s_t$  as  $Ls_{t+1}$ , where  $L$  is the lag operator, write this equation as

$$[I - (A - BD^{-1}C)L]s_{t+1} = BD^{-1}X_t. \quad (12.63)$$

If the eigenvalues of  $(A - BD^{-1}C)$  are strictly less than unity in modulus, then the inverse of  $[I - (A - BD^{-1}C)L]$  yields a square-summable polynomial in  $L$ , and thus  $s_{t+1}$  can be written as

$$s_{t+1} = \sum_{j=0}^{\infty} (A - BD^{-1}C)^j BD^{-1} X_{t-j}.$$

Lagging this expression by one period, and substituting for  $BD^{-1}X_t$  in (12.62), we obtain

$$X_t = C \sum_{j=0}^{\infty} (A - BD^{-1}C)^j BD^{-1} X_{t-j-1} + Dw_t. \quad (12.64)$$

Comparing (12.58) and (12.64), we see that the representation of  $X_t$  stemming from the model is an infinite-order VAR, with the mapping from structural innovations to VAR disturbances given by

$$Dw_t = Pv_t.$$

To reiterate, the two conditions that yielded this mapping are that  $D$  is square and invertible, and the eigenvalues of  $(A - BD^{-1}C)$  are strictly less than unity in modulus. The largest eigenvalue associated with  $(A - BD^{-1}C)$  determines the rate of decay of the lag coefficients  $(A - BD^{-1}C)^j$  in (12.64), and thus the extent to which the finite-ordered specification (12.58) serves as an adequate approximation of (12.64).

A recent empirical application involving the use of identified VARs involves assessments of the impact of technology shocks in driving the behavior of hours worked (depicted in figure 6.1). Interest in this topic follows Gali (1999), who used an identified VAR to show that hours tend to fall following the realization of positive productivity shocks. Because this evidence poses a severe challenge to the real-business-cycle paradigm, its identification has spawned a large literature.

One response to this evidence has involved investigations of whether various modifications to standard RBC environments are useful in helping take the evidence into account (e.g., Christiano and Todd, 1996; Boldrin, Christiano, and Fisher, 2001; and Francis and Ramey, 2003). An alternative response has involved challenges to the robustness of this evidence to the specification of the VAR upon which it is based. This latter line of investigation has shown the evidence to be fragile along at least two dimensions.

First, Christiano, Eichenbaum, and Vigfusson (2003) have shown that if hours are treated as being stationary in levels rather than in first differences (as Gali assumed), then the direction of the response in hours to productivity shocks is reversed. (For a discussion of difficulties associated with appropriate characterizations of stationarity, see section 1 of chapter 6.)

Second, Chari, Kehoe, and McGrattan (2005) have shown that the evidence can stem spuriously from the specification of a lag structure in the VAR that is not sufficiently rich to capture the dynamic behavior of the data it is meant to characterize. In particular, they show that specification errors associated with technology shocks identified using parsimoniously specified VARs tend to be large. Moreover, in experiments involving artificial data simulated from various DSGE models, they show more generally that VARs can have difficulties in identifying patterns of impulse responses consistent with those present in the models from which the data were obtained. The breadth of the applicability of this critique is surely a topic for future research. We now characterize means by which such simulation exercises can be conducted.

### 12.3.3 Simulations Involving Linearized Models

The third context again involves applications to fully specified linearized models. But in this case, we discuss situations in which implications conveyed by the model regarding the selected empirical targets must be calculated via simulations.

The recursive nature of DSGE models renders the task of model simulation straightforward (for details regarding the accuracy of simulations involving DSGE models, see Santos and Peralta-Alva, 2005). Two inputs are required to generate a simulation of  $\{x_t\}_{t=1}^T$  for a given model parameterization  $\mu$ . First is a specification of the initial state vector  $s_0$ , which

can be taken either as known or as a realization obtained from an underlying distribution. The stationary nature of the model (typically induced by an appropriate transformation of variables) renders this distinction as unimportant in simulations designed to characterize the asymptotic properties of the model. Because the remaining variables in  $x_t$  are a function of the state variable, the specification of  $s_0$  serves to determine  $x_0$ . Second is a drawing of structural shocks  $\{v_t\}_{t=1}^T$ , and if appropriate, a drawing of measurement errors  $\{u_t\}_{t=1}^T$ , from distributions specified for these processes. A random number generator is used to obtain these drawings. Note that unlike estimates obtained using GMM, simulation-based estimates are not distribution-free: inferences will be conditional on the distributions specified for  $v_t$  and  $u_t$ .

Given  $s_0$  and  $\{v_t\}_{t=1}^T$ ,  $\{x_t\}_{t=1}^T$  is constructed via recursive substitution using (12.49):  $x_1 = Fx_0 + v_1$ ,  $x_2 = Fx_1 + v_2$ , and so on, where  $e_t = Gu_t$ . Next,  $\{x_t\}_{t=1}^T$  and  $\{u_t\}_{t=1}^T$  combine in (12.51) to yield a simulated realization  $\{X_t\}_{t=1}^T$ . Finally, the function of interest is computed from this simulated realization, and a single simulation step has been completed. To eliminate the influence of the starting value chosen for  $s_0$ , a burn-in phase can be implemented. This involves merely discarding the first, say, 1,000 artificial realizations obtained for  $\{x_t\}_{t=1}^T$ .

### 12.3.4 Simulations Involving Nonlinear Approximations

The fourth context involves applications to nonlinear approximations to DSGE models. Simulation in this context is critical, as analytical calculations are rarely feasible. Chapter 9 described simulation-based estimation in detail; here we provide a brief synopsis.

To reestablish notation, let  $s_t$  again denote the vector of state variables contained in  $x_t$ , with law of motion determined by

$$s_t = f(s_{t-1}, v_t), \quad (12.65)$$

where  $v_t$  continues to represent structural shocks. Further, let  $c_t$  denote the vector of control variables contained in  $x_t$ . These are determined according to the policy rule

$$c_t = c(s_t), \quad (12.66)$$

which is derived as the solution to the optimization problem facing the decision makers included in the model. Finally, for empirical implementation, (12.65) and (12.66) are mapped into observables  $X_t$  by the observation equation

$$\begin{aligned} X_t &= \tilde{g}(s_t, c_t, v_t, u_t) \\ &\equiv g(s_t, u_t), \end{aligned} \quad (12.67)$$

where  $u_t$  continues to represent measurement error. The parameterizations of  $f(s_{t-1}, v_t)$ ,  $c(s_t)$ , and  $g(s_t, u_t)$ , along with the distributions governing the stochastic behavior of  $v_t$  and  $u_t$ , are determined by the structural parameters  $\mu$ .

Like its linearized counterpart, this nonlinear model representation is recursive, so the task of model simulation remains relatively straightforward. The list of inputs to the simulation process once again includes a specification of the initial state vector  $s_0$ , and distributions specified for  $v_t$  and  $u_t$ . Given these inputs, model simulation proceeds as in the linear case. From (12.65),  $s_0$  and  $\{v_t\}_{t=1}^T$  yield  $\{s_t\}_{t=1}^T$  directly. Using this series as an input to the policy function (12.66) yields  $\{c_t\}_{t=1}^T$ . Then  $\{s_t\}_{t=1}^T$  and  $\{c_t\}_{t=1}^T$  combine to form  $\{x_t\}_{t=1}^T$  ( $x_t \equiv (s'_t, c'_t)$ ), which combined with  $\{u_t\}_{t=1}^T$  maps into  $X_t$  as in (12.67). Functions of interest are computed using the simulated realizations of  $\{X_t\}_{t=1}^T$ , which completes the simulation step.

## 12.4 Empirical Application: Matching RBC Moments

In section 4 of chapter 11, we presented a calibration exercise designed to illustrate the ability of the RBC model introduced in section 1 of chapter 3 to match a collection of moments summarizing the time-series behavior of output, consumption, investment, and labor hours. Here we formalize this illustration statistically by conducting a moment-matching exercise. Specifically, we estimate the parameters of the model using various combinations of the moments reported in table 11.4 as targets, and then compare predictions conveyed by the model estimates regarding the moments with their empirical counterparts. Here we present results obtained using data extending through 2004:IV; as an exercise, you will be asked to work with the data set that extends through 2010:I.

Moments corresponding with the data were measured by estimating a six-lag VAR, and mapping the resulting coefficient estimates into moment measures as indicated in section 1.3 of chapter 7. Associated standard errors were obtained using the delta method, as indicated in section 2 of chapter 7. Moments corresponding with the model were measured as indicated above in section 3.2.

To begin, table 12.1 reports point estimates and associated standard errors of the full collection of moments we consider, along with moments obtained using the baseline model parameterization used in the calibration exercise. To convey the statistical significance of differences between the two sets of moment calculations, table 12.1 also reports associated  $t$ -statistics: differences between the empirical and model moments, divided by associated standard errors.

TABLE 12.1  
Baseline Moment Comparison

Moment	Point estimate	Std. err.	Model moment	$t$ -stat.
$\sigma_y$	0.0177	0.0019	0.0184	-0.41
$\sigma_c$	0.0081	0.0008	0.0085	-0.47
$\sigma_i$	0.0748	0.0070	0.0799	-0.74
$\sigma_n$	0.0185	0.0021	0.0087	4.62
$\varphi_y(1)$	0.8642	0.0199	0.7947	3.50
$\varphi_c(1)$	0.8318	0.0293	0.9601	-4.38
$\varphi_i(1)$	0.7891	0.0307	0.7367	1.71
$\varphi_n(1)$	0.8992	0.0144	0.7290	11.78
$\varphi_{y,c}(0)$	0.8237	0.0379	0.7782	1.20
$\varphi_{y,i}(0)$	0.9502	0.0101	0.9459	0.42
$\varphi_{y,n}(0)$	0.8342	0.0241	0.9107	-3.18
$\varphi_{y,c}(1)$	0.7467	0.0436	0.7591	-0.28
$\varphi_{y,i}(1)$	0.7991	0.0355	0.6790	3.38
$\varphi_{y,n}(1)$	0.6230	0.0396	0.6312	-0.21

Note:  $\varphi(1)$  denotes first-order serial correlation;  $\varphi_{j,y}(l)$  denotes  $l^{\text{th}}$  order correlation between variables  $j$  and  $y$ . Model moments based on the parameterization  $\theta = [\alpha \beta \phi \varphi \delta \rho \sigma]'$  = [0.24 0.99 1.5 0.35 0.025 0.78 0.0067]'. Empirical moments obtained using data extending through 2004:IV.

Immediately apparent from the table is the poor performance of the model in characterizing hours: the  $t$ -statistics associated with  $\sigma_n$ ,  $\varphi_n(1)$ , and  $\varphi_{y,n}(0)$  lie above 3 in absolute value, and range as high as nearly 12. This poor performance was also clearly evident in the calibration exercise. What was not evident in the calibration exercise, but emerges from table 12.1, is that aspects of the model's characterization of each of the additional series are also poor. For example, the  $t$ -statistics associated with  $\varphi_y(1)$ ,  $\varphi_c(1)$ , and  $\varphi_{y,i}(1)$  all lie above 3 in absolute value. The question we now address is, can the model's performance be improved by conducting an estimation step?

We facilitate estimation using the GMM approach outlined in section 2.1. Specifically, in this case

$$\theta = [\alpha \beta \phi \varphi \delta \rho \sigma]',$$

the entries of  $g(Z, \theta)$  contain differences between empirical and model moments, the weighting matrix  $\Omega$  is the inverse of the variance-covariance associated with the empirical moment estimates, and estimates  $\hat{\theta}$  are obtained as the solution to

$$\min_{\theta} \Gamma(\theta) = g(Z, \theta)' \times \Omega \times g(Z, \theta).$$

TABLE 12.2  
Parameter Estimates

Parameter	Calibrated value	All moments	C subset	I subset	N subset
$\alpha$	0.24	0.1860 ( $8.81e-6$ )	0.1792 ( $4.89e-4$ )	0.3224 (0.0013)	0.1791 ( $2.41e-5$ )
$\beta$	0.99 [0.9877, 0.999]	0.9983 ( $1.17e-6$ )	0.9990 ( $1.12e-4$ )	0.9916 ( $7.27e-5$ )	0.9990 ( $5.52e-6$ )
$\delta$	0.025 [0.01, 0.04]	0.0267 ( $3.76e-5$ )	0.0400 ( $5.07e-9$ )	0.0100 ( $3.20e-9$ )	0.0400 ( $4.07e-8$ )
$\phi$	1.5 [0.5, 2.5]	1.184 (0.0089)	2.4999 (0.0031)	2.0512 (0.3812)	1.0206 (13.484)
$\varphi$	0.35	0.3364 ( $2.24e-6$ )	0.3345 ( $1.33e-4$ )	0.3785 ( $4.48e-4$ )	0.3345 ( $6.56e-6$ )
$\rho$	0.78 [0.6, 0.95]	0.8952 (0.0075)	0.6000 (0.5728)	0.8166 (0.0593)	0.9499 (1.4438)
$\sigma$	0.0067 (0, 0.01)	0.0051 ( $2.04e-4$ )	0.0099 (0.0437)	0.0062 (0.0018)	0.0006 (0.0045)

Note: Brackets reported under calibrated parameter values indicate range restrictions. Also, parentheses reported under parameter estimates indicate standard errors. Finally, moments included under J subset are  $\sigma_y$ ,  $\sigma_c$ ,  $\varphi_j(1)$ ,  $\varphi_{y,j}(0)$  and  $\varphi_{y,j}(1)$ ,  $j = c, i, n$ .

Minimization was accomplished using GAUSS's derivative-based software package optmum; details regarding the algorithms it implements are provided in chapter 13.

Table 12.2 reports parameter estimates obtained using four alternative sets of empirical moments as targets: the full set of moments, and three subsets designed to capture the relationship between output and the three remaining variables, each in isolation:  $\sigma_y$ ,  $\sigma_c$ ,  $\varphi_j(1)$ ,  $\varphi_{y,j}(0)$ , and  $\varphi_{y,j}(1)$ ,  $j = c, i, n$ . In all cases, parameter estimates were restricted to the following ranges of values deemed to be reasonable a priori: the discount factor  $\beta \in [0.9877, 0.999]$ , implying a restriction on average real interest rates between 0.1% and 1.25% on a quarterly basis; the CRRA parameter  $\phi \in [0.5, 2.5]$ ; the depreciation rate  $\delta \in [0.01, 0.04]$ , implying annual depreciation rates between 4% and 16%; the AR coefficient  $\rho \in [0.6, 0.95]$ ; and the innovation standard deviation  $\sigma \in (0, 0.01]$ . In addition, the capital-share parameter  $\alpha$ , along with  $\varphi$ , which indicates the importance of consumption relative to leisure in the instantaneous utility specification, were restricted to ensure that steady state values of  $\bar{y}$ ,  $\bar{i}$ , and  $\bar{n}$  match their long-run empirical counterparts (measured as sample averages):

$$\alpha = \left( \frac{\delta + \varphi}{\delta} \right) \frac{\bar{i}}{\bar{y}},$$

#### 12.4 Empirical Application: Matching RBC Moments

$$\varphi = \frac{1}{1 + \frac{2(1-\alpha)}{\bar{i}/\bar{y}}},$$

where  $\varphi = 1/\beta - 1$ . Details regarding the derivation of these restrictions are provided in section 4 of chapter 11, and details regarding the imposition of range restrictions on  $\beta$  and so on are provided in section 4 of chapter 13.

The parameter estimates we obtain are rather unstable across the sets of moments we consider. For example, using the "C Subset," we obtain a relatively high estimate of the CRRA parameter  $\phi$  (near its upper bound of 2.5) and low estimate of the AR parameter  $\rho$  (at its lower bound of 0.6). In contrast, we obtain the reverse pattern of estimates using the "N Subset": the estimate of  $\phi$  we obtain is relatively low (1.02), and the estimate of  $\rho$  is relatively high (0.9499). As another example, while the estimate of  $\alpha$  obtained under the "I Subset" is rather high (0.3224), the remaining estimates in the neighborhood of 0.18, substantially below the calibrated value of 0.24. Given this instability, the results obtained using any given set of moments should be interpreted with caution.

The correspondence between these parameter estimates and their targeted moments is reported in table 12.3. In addition to point estimates, discrepancies with empirical counterparts are again reported using t-statistics.

TABLE 12.3  
Moment Comparisons Using Model Estimates

Moment	All moments	C subset	I subset	N subset
$\sigma_y$	0.0195 (-0.98)	0.0212 (-1.88)	0.0178 (-0.056)	0.0028 (7.96)
$\sigma_c$	0.0109 (-3.27)	0.0084 (-0.36)	XX	XX
$\sigma_i$	0.0739 (0.13)	XX	0.0784 (-0.52)	XX
$\sigma_n$	0.0079 (4.96)	XX	XX	0.0001 (8.39)
$\varphi_y(1)$	0.8952 (-1.56)	XX	XX	XX
$\varphi_c(1)$	0.9801 (-5.05)	0.8427 (-0.37)	XX	XX
$\varphi_i(1)$	0.8430 (-1.76)	XX	0.8031 (-0.46)	XX
$\varphi_n(1)$	0.8334 (4.55)	XX	XX	0.8520 (3.26)
$\varphi_{y,c}(0)$	0.8400 (-0.43)	0.8597 (-2.16)	XX	XX
$\varphi_{y,i}(0)$	0.9262 (2.38)	XX	0.9589 (-0.87)	XX
$\varphi_{y,n}(0)$	0.8691 (-1.45)	XX	XX	0.7313 (4.27)
$\varphi_{y,c}(1)$	0.8408 (-2.16)	0.6961 (-1.16)	XX	XX
$\varphi_{y,i}(1)$	0.7674 (0.89)	XX	0.7623 (1.04)	XX
$\varphi_{y,n}(1)$	0.6870 (-1.87)	XX	XX	0.59700 (0.66)

Note:  $\varphi(1)$  denotes first-order serial correlation;  $\varphi_{j,y}(t)$  denotes  $t^{\text{th}}$  order correlation between variables  $j$  and  $y$ ; t-statistics indicating discrepancies with empirical counterparts provided in parentheses. XX denotes unestimated moment.

**TABLE 12.4**  
Empirical Moments, Extended Data Set

Moment	Point estimate	Std. Err.
$\sigma_y$	0.0177	0.0017
$\sigma_c$	0.0078	0.0008
$\sigma_i$	0.0758	0.0064
$\sigma_n$	0.0190	0.0020
$\varphi_y(1)$	0.8654	0.0184
$\varphi_c(1)$	0.8415	0.0253
$\varphi_i(1)$	0.7931	0.0272
$\varphi_n(1)$	0.9043	0.0127
$\varphi_{y,c}(0)$	0.8241	0.0345
$\varphi_{y,i}(0)$	0.9554	0.0080
$\varphi_{y,n}(0)$	0.8273	0.0207
$\varphi_{y,c}(1)$	0.7492	0.0396
$\varphi_{y,i}(1)$	0.8036	0.0313
$\varphi_{y,n}(1)$	0.6125	0.0345

Note:  $\varphi(1)$  denotes first-order serial correlation;  $\varphi_{j,y}(l)$  denotes  $l^{\text{th}}$  order correlation between variables  $j$  and  $y$ . Estimates obtained using data extending through 2010:I.

Regarding the results obtained using the full collection of moments, the estimation step is seen to have generated little payoff: the performance of the model remains poor across many dimensions. Most notably, the volatility and first-order serial correlation of hours remains problematic, as does the first-order serial correlation of consumption. The model's characterization of hours remains particularly poor: in targeting the moments of hours in isolation, the  $t$ -statistics associated with  $\sigma_n$ ,  $\varphi_n(1)$ , and  $\varphi_{y,n}(0)$  continue to lie above 3. However, targeting the moments of both consumption and investment in isolation, we obtain only one  $t$ -statistic above 2: that associated with  $\varphi_{y,c}(0)$ . This helps explain the instability of the parameter estimates noted in table 12.2: while it is possible to parameterize the model in order to account relatively well for the behavior of consumption and investment in isolation, it is not possible to do so in attempting to account for their joint behavior.

In sum, we draw two conclusions from this exercise. First, although the standard RBC model is well known to have difficulties in characterizing the behavior of hours, its performance is also disappointing along additional dimensions. Second, it is apparent that improvements in the overall performance of the model cannot be attained merely through adjustments in its parameterization: dramatic improvements in performance do not result from an optimization step.

#### 12.4 Empirical Application: Matching RBC Moments

Of course, these results should not be viewed as an indictment of the empirical performance of RBC models in general. State-of-the-art specifications featuring, for example, time-to-build production processes, nonconvexities associated with adjustments to labor and capital inputs, and so on, have been developed to improve upon the performance of the baseline specification we have considered. (The papers in Cooley, 1995, provide a good overview of developments on this front. See also the collection of papers in Sustek, 2005.) Instead, we have offered this exercise as an illustration of the ability of moment-matching exercises to highlight particular dimensions along which the model under investigation succeeds or fails to capture relevant facets of the data it has been designed to characterize. This sort of information is particularly valuable in helping to pinpoint trouble spots, and thus in providing guidance for pursuing fruitful extensions.

#### Exercise 12.2

The empirical moments estimated using the data set extended through 2010:I are reported in table 12.4. Using these moments, reconstruct the results reported in tables 12.2 and 12.3. Is the performance of the model enhanced in working with the extended data set?

# *Chapter 13*

## *Maximum Likelihood*

### **13.1 Overview**

Chapters 11 and 12 presented limited-information methods for evaluating DSGE models. This chapter and the next present full-information alternatives: the classical (chapter 13) and Bayesian (chapter 14) approaches to likelihood analysis. Under these methods, the DSGE model under investigation is viewed as providing a complete statistical characterization of the data, in the form of a likelihood function. As described in chapters 8 and 10, the source of the likelihood function is the corresponding state-space representation of the observable data, coupled with a distributional assumption for the stochastic innovations included in the model.

Two features of full-information analyses distinguish them from the moment-based procedures presented previously. First is the requirement of a distributional assumption for the stochastic innovations: inferences obtained under such analyses are jointly conditional upon the model and the underlying distributional assumption. Second, the basis of inference is the full range of empirical implications conveyed by the model, rather than a subset of implications conveyed by the collection of moments chosen by the researcher.

The distinction between classical and Bayesian approaches to likelihood analysis can be cast succinctly in terms of whether the data or parameters are interpreted by the researcher as random variables. (For more elaborate treatments of this distinction, see, e.g., Berger and Wolpert, 1988; or Poirier, 1995.) Under the classical approach, parameters are interpreted as fixed (but unknown), and the data are interpreted as the realization of a random drawing from the corresponding likelihood function. Given the realization of a drawing of data, parameter estimates are chosen as those that maximize the likelihood function; uncertainty regarding the specific values estimated for the parameters is conveyed by reporting associated standard errors. Uncertainty regarding parameter estimates arises from the fact that the realized data represent one of many possible realizations that could have been obtained from the likelihood function. Because parameter estimates are a function of the corresponding data, randomness in the data imparts randomness in the parameter estimates.

### *13.1 Overview*

Under the classical approach, the empirical plausibility of a given model may be assessed using formal hypothesis-testing procedures. Such procedures begin with the designation of a parameterized version of the model as the null hypothesis, or in other words, as the stochastic process from which the observed data have been realized. The likelihood function corresponding with the null hypothesis is then used to assess the probability that the observed data could have in fact been generated by the null model. If the probability lies above a prespecified threshold, or critical value, the null hypothesis is not rejected, and the model is deemed to have passed the test; otherwise the parameterized model is rejected as a plausible candidate for representing the true data generation process.

Under the Bayesian approach, parameters are interpreted as random and the data as fixed. In estimation, the objective is to make conditional probabilistic statements regarding the parameters of the model, where conditioning is with respect to the observed data. The probabilistic interpretation of the parameters enables the formal incorporation of a priori views regarding their specification. These views are introduced through the specification of a prior distribution; coupling this distribution with the likelihood function via Bayes' Rule yields a posterior distribution for the parameters. Means or modes of the posterior distribution typically provide point estimates of the parameters, and uncertainty associated with these estimates are typically conveyed using posterior standard deviations.

Inferences regarding the empirical support assigned by the data to a particular model involve comparisons of the conditional probability associated with the model relative to probabilities associated with alternative models. As with estimation, conditioning is once again made with respect to the data, and probability calculations are based on posterior distributions. In contrast to the classical approach, no one model is maintained as the null hypothesis to be accepted or rejected as the true data generation process. Instead, judgments regarding the empirical plausibility of competing models are based upon assessments of their relative conditional probabilities.

In sum, the assessment of a given model under the classical approach to likelihood analysis involves the following question: how plausible are the observed data, given the maintained null hypothesis? Under the Bayesian approach, the question is: how plausible is a given model relative to competing alternatives, given the observed data? This chapter provides details regarding the implementation of the classical approach. (For general textbook discussions of the classical approach to likelihood analysis, see, e.g., Amemiya, 1985; Hamilton, 1999; Harvey, 1994; or Greene, 2003.) The key component of implementation involves the maximization of likelihood functions associated with state-space representations. Chapter 14 then provides details regarding the implementation of the Bayesian approach.

### 13.2 Introduction and Historical Background

Chapters 8 and 10 provided details regarding the evaluation of the likelihood function  $L(X|\mu)$  associated with the state-space representation of a given DSGE model. To reestablish notation, recall that the  $k \times 1$  vector  $\mu$  contains the model's structural parameters, and  $X$  represents the  $T$  observations of the  $m \times 1$  vector of observable variables  $X_t$ . The likelihood function is a multiplicative product over the time- $t$  data points  $X_t$ :

$$L(X|\mu) = \prod_{t=1}^T L(X_t|\mu). \quad (13.1)$$

Given this structure, it is convenient in practice to work with the log-likelihood function  $\log L(X|\mu)$ , which is additive over time- $t$  data points:

$$\log L(X|\mu) = \sum_{t=1}^T \log L(X_t|\mu). \quad (13.2)$$

Convenience aside, details regarding implementation apply equally to  $L(X|\mu)$  and  $\log L(X|\mu)$ : Because the logarithmic transform is monotonic, the maximum of the log-likelihood function corresponds to that of the likelihood function.

As noted, the classical approach to estimation involves the maximization of  $\log L(X|\mu)$  with respect to  $\mu$ . Let this maximum-likelihood (ML) estimator be denoted as  $\hat{\mu}$ . Because  $\hat{\mu}$  is a function of  $X$ , it is interpretable as a random variable; thus there is uncertainty associated with the information it conveys regarding the true value of  $\mu$ . This uncertainty is conveyed by reporting associated standard errors.

For a heuristic description of the means by which  $\hat{\mu}$  and associated standard errors may be obtained, note that the log-likelihood function is a particular example of a more general function  $f(\mu)$ . For the case in which  $\mu$  is  $1 \times 1$ , the first-order condition associated with the maximization of  $f(\mu)$  is given by  $f'(\mu) = 0$ , and the second-order condition is  $f''(\mu) < 0$ , implying that the function decreases when moving away from  $\hat{\mu}$ . (Here we take for granted the existence of a single optimizing value  $\hat{\mu}$ ; departures from this case are discussed in section 5.) When  $|f''(\hat{\mu})|$  is large, the function decreases rapidly in moving away from  $\hat{\mu}$ , and thus  $\hat{\mu}$  has been identified with a relatively high degree of precision. For the case in which  $\mu$  is  $k \times 1$ , the same first-order condition applies, except that  $f'(\mu)$  is the  $k \times 1$  gradient vector associated with  $f(\mu)$ ; the second-order condition is that the  $k \times k$  matrix of second derivatives (the Hessian associated with  $f(\mu)$ ) is negative definite when evaluated at  $\hat{\mu}$ . The larger in absolute value

### 13.2 Introduction and Historical Background

is the  $j^{\text{th}}$  diagonal element of the Hessian, the more rapidly  $f(\hat{\mu})$  decreases in moving away from  $\hat{\mu}_j$ .

For the specific case in which the function of interest is the log-likelihood function, the theory of maximum likelihood states that under suitable regularity conditions, the ML estimator  $\hat{\mu}$  is consistent and distributed asymptotically as Normal (e.g., see Amemiya, 1985):

$$\hat{\mu} \rightarrow N(\mu^*, (T\mathfrak{S})^{-1}), \quad (13.3)$$

where  $\mu^*$  is the true but unknown value of  $\mu$  and  $\mathfrak{S}$  is the Hessian associated with  $\log L(X|\mu)$ . The presence of  $T^{-1}$  in the asymptotic covariance matrix offsets the dependence of  $\mathfrak{S}$  on  $T$ :

$$\mathfrak{S}(\mu) = -T^{-1} \frac{\partial^2 \log L(X|\mu)}{\partial \mu \partial \mu'}. \quad (13.4)$$

Substituting (13.4) into (13.3),  $T$  is eliminated, yielding the asymptotic covariance matrix

$$\Sigma_\mu = \left[ \frac{\partial^2 \log L(X|\mu)}{\partial \mu \partial \mu'} \right]^{-1}. \quad (13.5)$$

Evaluating (13.5) at  $\hat{\mu}$  yields an estimate of the covariance matrix  $\widehat{\Sigma}_\mu$  associated with  $\hat{\mu}$ , the diagonal elements of which are associated standard errors,  $\widehat{\sigma}_{\mu_j}$ . Following the logic of the heuristic description, if the log-likelihood function falls sharply in moving away from  $\hat{\mu}_j$ , its corresponding standard error  $\widehat{\sigma}_{\mu_j}$  will be small, and  $\hat{\mu}_j$  will be tightly identified. Before moving to a discussion of techniques for maximizing  $\log L(X|\mu)$  and computing Hessians, we pause to provide historical perspective on their application to DSGE models.

The mapping of DSGE models into likelihood functions specified over  $\mu$  was pioneered by Sargent (1989). A key obstacle that must be overcome in developing any such mapping involves what is known as the stochastic singularity problem: DSGE models carry nontrivial implications for the stochastic behavior of only as many observable variables as there are sources of uncertainty included in the model. With fewer sources of uncertainty than observables, various linear combinations of variables will be predicted to be deterministic, and departures from this prediction will render likelihood analysis untenable (for elaboration, see Ingram, Kocherlakota, and Savin, 1994).

Sargent overcame this obstacle by adding measurement errors in his model's observation equations. This resolves the stochastic singularity problem by attributing departures from the deterministic behavior implied for linear combinations of the observables to measurement errors

associated with their observation. Subsequent studies that used this approach to the problem include Altug (1989), McGrattan (1994), and McGrattan, Rogerson, and Wright (1997).

Ingram, Kocherlakota, and Savin (1994) advocate an alternative resolution to the stochastic singularity problem. Their approach involves the specification of a model featuring as many structural shocks as there are observable variables; this has become known as the multiple-shock approach. Their motivation for this approach, cast in the context of an RBC framework, is that the real world features many more sources of stochastic uncertainty than can be captured through the specification of a catch-all total factor productivity shock. Examples of studies that have used this approach include DeJong, Ingram, and Whiteman (2000a, 2000b), DeJong and Ingram (2001), and Otronk (2001).

An alternative to the multiple-shock approach has been proposed by Ireland (2004b), who recognized the potential for a broader interpretation (beyond measurement error) of the role played by the stochastic terms  $u_t$  that drive a wedge between model predictions and empirical observations in state-space measurement equations. Ireland's contribution lies in the observation that these stochastic terms can play a role in addressing the model-specification issues that have led some researchers to reject the use of statistical formalism in analyzing DSGE models.

A broad discussion of these issues is provided in chapter 11, but the crux of the problem is summarized by Prescott (1986): "The models constructed within this theoretical framework are necessarily highly abstract. Consequently, they are necessarily false, and statistical hypothesis testing will reject them" (10). But as Ireland notes, beyond capturing measurement error,  $u_t$  can be "interpreted more liberally as capturing all of the movements and co-movements in the data that the real business cycle model, because of its elegance and simplicity, cannot explain. In this way, the hybrid model ... combines the power of the DSGE model with the flexibility of a VAR" (1210). Under this interpretation,  $u_t$  plays the role that residuals play in empirical applications involving reduced-form specifications. Their inclusion in the specification of a DSGE model thus provides explicit recognition that the model is not designed to capture the full extent of variation observed in the data.

### 13.3 A Primer on Optimization Algorithms

We now characterize numerical methods for obtaining ML estimates. Several optimization algorithms exist for this purpose; here we discuss two classes: simplex and derivative-based algorithms. (For more comprehensive

textbook overviews of optimization algorithms, see Judd, 1998; Hamilton, 1994; and Harvey, 1999.) In practice, most algorithms are designed to minimize the function of interest. Therefore, hereafter we consider the objective of calculating the  $\hat{\mu}$  that minimizes  $f(\mu) = -\log L(X|\mu)$ .

The algorithms we present require two inputs: a procedure that takes  $\mu$  as an input and delivers  $f(\mu)$  as output, and an initial value specified for  $\mu$ . Given these inputs, simplex algorithms deliver as output parameter estimates  $\hat{\mu}$  and the corresponding value of the likelihood function  $f(\hat{\mu})$ ; in addition, derivative-based methods also deliver  $\hat{\Sigma}_{\mu}$ . (Given the use of a simplex method, an additional step, described below, is required to produce  $\hat{\Sigma}_{\mu}$ .)

In this section, we proceed under the assumption that  $f(\mu)$  is well-behaved. In particular, we assume that it is continuous and monotonic in  $\mu$ , and contains a global minimum. Deviations from this rosy scenario are discussed in section 8.4.

#### 13.3.1 Simplex Methods

Simplex algorithms are designed to minimize  $f(\mu)$ ,  $\mu \in \mathbb{R}^k$  via the use of "triangularization" techniques, which involve the construction and manipulation of simplexes. With  $\mu$  being  $k \times 1$ , a simplex  $\Delta$  is a  $k+1$  collection of specific specifications of  $\mu$ :  $(\mu_1, \dots, \mu_{k+1})$ . The specifications are referred to as vertices. For example, given  $k=2$ ,  $\Delta$  is a triangle; and given  $k=3$ ,  $\Delta$  is a pyramid.

Simplex algorithms are iterative procedures designed to manipulate the shape and location of simplexes over the parameter space, with the ultimate goal of producing a degenerate simplex at the optimal value  $\hat{\mu}$ . Many specific algorithms exist for this purpose; here we present a popular version developed by Nelder and Mead (1965). Further details regarding this algorithm are provided by Lagarias et al. (1998), who detailed its convergence properties in low-dimensional applications; we follow their discussion closely here.<sup>1</sup>

The basic idea behind any simplex algorithm is to search by trial and error for "lower ground" on the likelihood surface. Starting from a given point, the first step in a typical algorithm is to determine a local neighborhood over which an additional move is contemplated. Technically, this amounts to the construction of a simplex. The second step is to make a move within this local neighborhood, from which a new local neighborhood is established and probed. Depending upon the particular conditions

<sup>1</sup> GAUSS code for implementing this algorithm is available through the American University archive at [www.american.edu/academic.depts/cas/econ/gaussres/GAUSSIDX.HTM](http://www.american.edu/academic.depts/cas/econ/gaussres/GAUSSIDX.HTM). It is also available at the GAUSS software archive <http://faculty.washington.edu/rons/>.

encountered at a given point, the size and shape of the selected local neighborhood can vary substantially. The evolution of simplexes that results from the exploration of the likelihood surface is suggestive of the movements of an amoeba gliding through a fluid.

Relative to derivative-based algorithms, simplex algorithms are attractive because they do not require that the likelihood surface be differentiable. In the context of evaluating DSGE models, this is a particularly important consideration because analytical expressions for derivatives are rarely available; and the calculation of numerical derivatives can be inaccurate (particularly in working with numerical approximations of the likelihood function, as with the bootstrap particle filter described in chapter 10).

We begin with a heuristic description of the two-dimensional case. As noted, in this case the simplex is a triangle with three vertices. Let  $\mu = (\beta, \alpha)$ , and denote starting values as  $\mu^0 = (\beta^0, \alpha^0)$ . These map into an initial triangle  $\Delta^0$  as follows:  $\hat{\mu}_1^0 = (\beta_0^0, 0)$ ,  $\hat{\mu}_2^0 = (0, \alpha_0^0)$ , and  $\hat{\mu}_3^0 = (0, 0)$ ; a graphical characterization is provided in figure 13.1.

Using this initial specification, an iteration designed to generate a new triangle that lies closer to the minimum value  $\hat{\mu}$  is performed. Subsequent iterations have the same objective; thus the designation of  $\mu^0$  as the initial specification of  $\mu$  is unimportant; hereafter superscripts denoting iteration

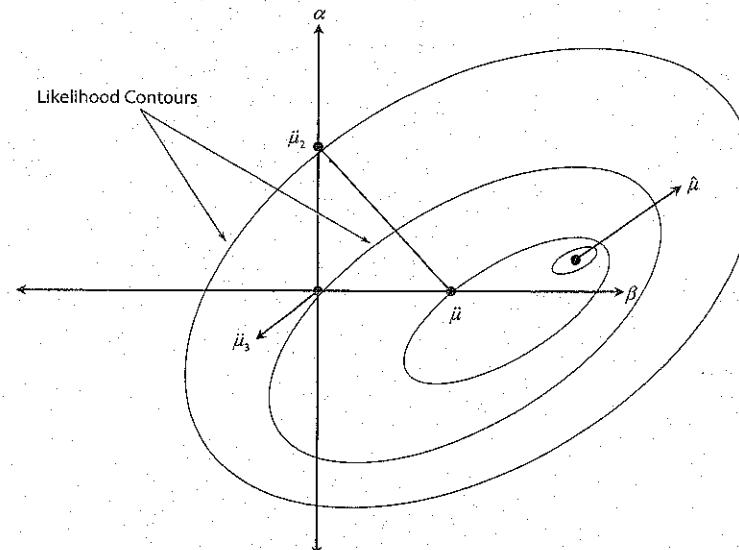


Figure 13.1 Initial  $\Delta^0$  for Nelder-Mead Algorithm

### 13.3 A Primer on Optimization Algorithms

numbers will be suppressed. The first step of the iteration is to relabel (if necessary) the vertices so that  $f(\mu_1) \leq f(\mu_2) \leq f(\mu_3)$ . Note from figure 13.1 that, given the likelihood contours, it is in fact the case that  $f(\hat{\mu}_1) \leq f(\hat{\mu}_3) \leq f(\hat{\mu}_2)$ ; thus the vertices are relabeled as  $\mu_1 = (\beta_0^0, 0)$ ,  $\mu_2 = (0, 0)$ , and  $\mu_3 = (0, \alpha_0^0)$ , achieving  $f(\mu_1) \leq f(\mu_2) \leq f(\mu_3)$ .

Given this rearrangement, a reflection point  $\mu_r$  is calculated:

$$\mu_r = \bar{\mu} + (\bar{\mu} - \mu_3), \quad \bar{\mu} = \frac{\mu_1 + \mu_2}{2}$$

The reflection point, along with the relabeled vertices, is depicted in figure 13.2. (The initial simplex is drawn with dashed lines in all of the subsequent figures in this chapter.)

The next step involves taking either a “shrink step” or a “nonshrink step,” depending upon contingencies involving the reflection point. We first describe contingencies under which a nonshrink step is taken. If these fail, a shrink step is taken.

Several possibilities fall under the heading of a nonshrink step. First, if  $f(\mu_1) \leq f(\mu_r) < f(\mu_2)$ , so that  $\mu_r$  is superior to the next-to-worst vertex associated with the previous iteration, but inferior to the best vertex, the worst point (here  $\mu_3$ ) is discarded and replaced with  $\mu_r$ ; the simplex

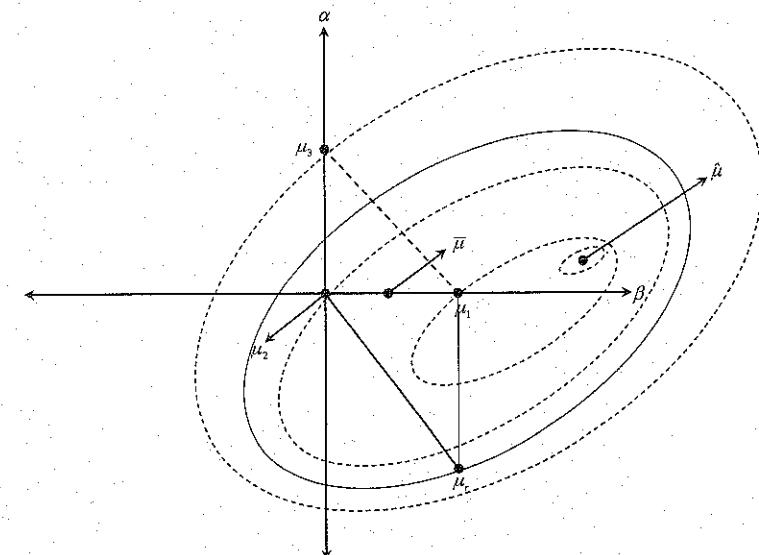


Figure 13.2 Reflection Point

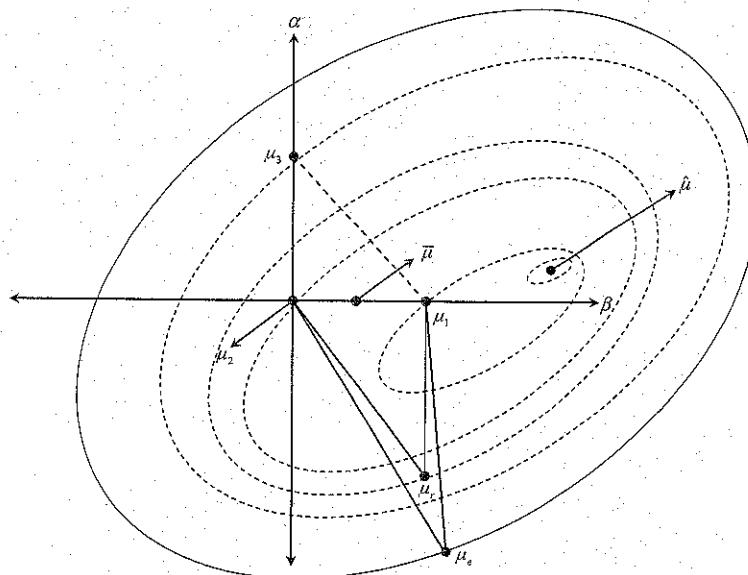


Figure 13.3 Expansion Point

constructed using  $(\mu_1, \mu_2, \mu_r)$  is then passed to the next iteration and the current iteration terminates.

If  $f(\mu_1) \leq f(\mu_r) < f(\mu_2)$  does not hold, as is true with the scenario depicted in figure 13.2, two alternative possibilities arise. If  $f(\mu_r) < f(\mu_1)$ , an expansion step is taken; if instead  $f(\mu_r) \geq f(\mu_2)$ , a contraction step is taken. Both fall under the heading of a nonshrink step. In the former case, the algorithm attempts to improve upon  $\mu_r$  by computing

$$\mu_e = \bar{\mu} + 2(\mu_r - \bar{\mu}).$$

If this indeed turns out to be an improvement,  $\mu_e$  is accepted in favor of  $\mu_r$  and the simplex constructed using  $(\mu_1, \mu_2, \mu_e)$  is passed to the next iteration. Otherwise  $\mu_r$  is retained and the simplex constructed using  $(\mu_1, \mu_2, \mu_r)$  is passed to the next iteration. In the example depicted in figure 13.3, the expansion point would not be accepted.

If  $f(\mu_r) \geq f(\mu_2)$ , the algorithm proceeds to a contraction step, of which there are two types. An outside contraction is conducted if  $f(\mu_3) > f(\mu_r) \geq f(\mu_2)$ ; that is, if  $\mu_r$  is better than the worse vertex but inferior to the next-to-worse vertex. This type of contraction begins by first computing

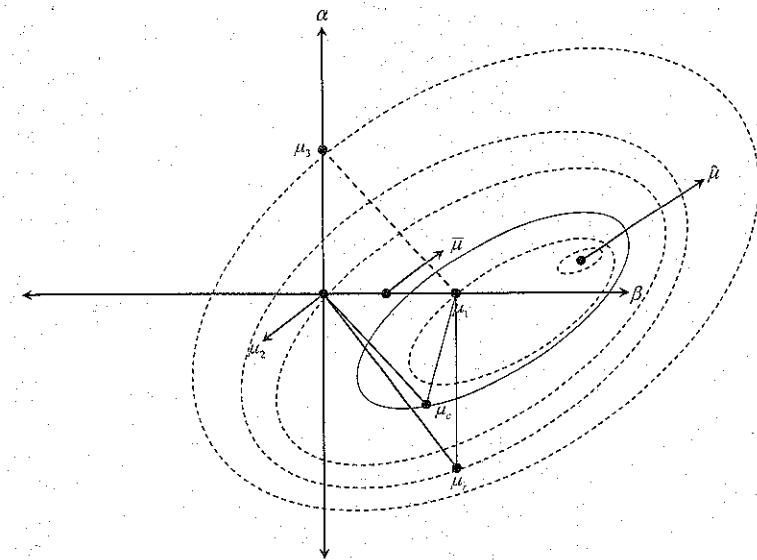


Figure 13.4 Outside Contraction Point

$$\mu_c = \bar{\mu} + \frac{1}{2}(\mu_r - \bar{\mu}).$$

If  $f(\mu_c) \leq f(\mu_r)$ , as is the case in the example depicted in figure 13.4, the new simplex computed using  $(\mu_1, \mu_2, \mu_c)$  is passed to the next iteration. Alternatively, if  $f(\mu_c) > f(\mu_r)$ , then  $\mu_c$  is discarded. In this case the algorithm moves to a shrink step, described below. Note that the rejection of an outside contraction leads to a shrink step even though the reflection point has improved over the worst point  $\mu_3$ , a property of the Nelder-Mead method noted by Lagarias et al. (1998).

An inside contraction is performed for the case in which  $f(\mu_r) \geq f(\mu_3)$ . This is performed by computing

$$\mu_{cc} = \bar{\mu} - \frac{1}{2}(\mu_r - \mu_3).$$

If  $f(\mu_{cc}) < f(\mu_3)$ , the simplex constructed using  $(\mu_1, \mu_2, \mu_{cc})$  is passed to the next iteration; otherwise the algorithm proceeds to a shrink step. An inside contraction point is depicted in figure 13.5.

If the algorithm reaches a shrink step, two new vertices are constructed:

$$\nu_l = \mu_1 + \frac{1}{2}(\mu_l - \mu_1), l = 2, 3.$$

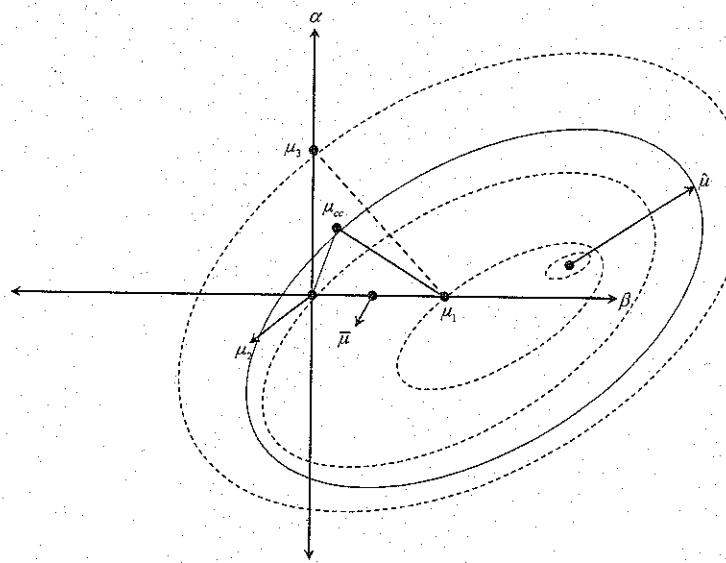


Figure 13.5 Inside Contraction Point

This results in the simplex constructed using  $(\mu_1, \nu_2, \nu_3)$ , which is passed to the next iteration. This simplex is depicted for the example in figure 13.6.

The full range of possibilities outlined above for the completion of a single iteration are illustrated in the flow chart depicted in figure 13.7. As illustrated by the two-dimensional example, the objective of an iteration is to create a new simplex that comes closer to containing  $\bar{\mu}$  as an exclusive element.

We now indicate briefly how the iteration process extends to the  $k$ -dimensional case. Under the first step of the  $i^{\text{th}}$  iteration in the  $k$ -dimensional case, the vertices of the simplex  $\Delta_k^{(i)}$ , denoted as  $\mu_1^{(i)}, \dots, \mu_{k+1}^{(i)}$ , are ordered such that

$$f(\mu_1^{(i)}) \leq \dots \leq f(\mu_{k+1}^{(i)}). \quad (13.6)$$

(Under the first iteration,  $\mu_{k+1}^0$  is typically chosen as the zero vector.) Once again subsuming iteration numbers, the next step involves the computation of a reflection point:

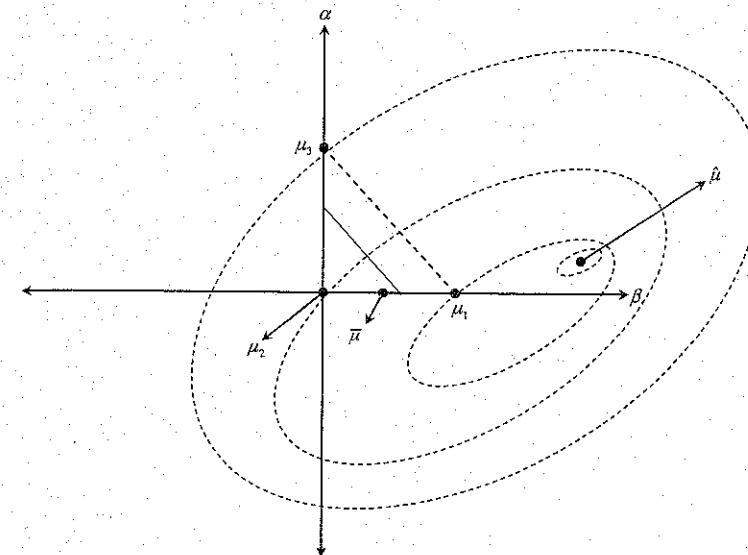


Figure 13.6 Shrink Step

$$\mu_r = \bar{\mu} + \rho(\bar{\mu} - \mu_{k+1}), \quad (13.7)$$

$$\bar{\mu} = \frac{\sum_{j=1}^k \mu_j}{k}, \quad \rho > 0. \quad (13.8)$$

If  $f(\mu_1) \leq f(\mu_r) < f(\mu_k)$ ,  $\mu_{k+1}$  is replaced by  $\mu_r$  and the resulting simplex is passed to the next iteration. Otherwise if  $f(\mu_r) < f(\mu_1)$ , an expansion point is computed as

$$\begin{aligned} \mu_e &= \bar{\mu} + \chi(\mu_r - \bar{\mu}) \\ &= (1 + \rho\chi)\bar{\mu} - \rho\mu_{k+1}, \quad \chi > \rho. \end{aligned} \quad (13.9)$$

If  $f(\mu_e) < f(\mu_r)$ ,  $\mu_e$  replaces  $\mu_r$ ; otherwise  $\mu_r$  is retained and the resulting simplex is passed to the next iteration.

If instead  $f(\mu_r) \geq f(\mu_k)$ , one of two types of a contraction is performed. If  $f(\mu_k) \leq f(\mu_r) < f(\mu_{k+1})$ , an outside contraction point is computed:

$$\begin{aligned} \mu_i &= \bar{\mu} + \gamma(\mu_r - \bar{\mu}) \\ &= (1 + \rho\gamma)\bar{\mu} - \rho\gamma\mu_{k+1}, \quad \gamma \in (0, 1). \end{aligned} \quad (13.10)$$

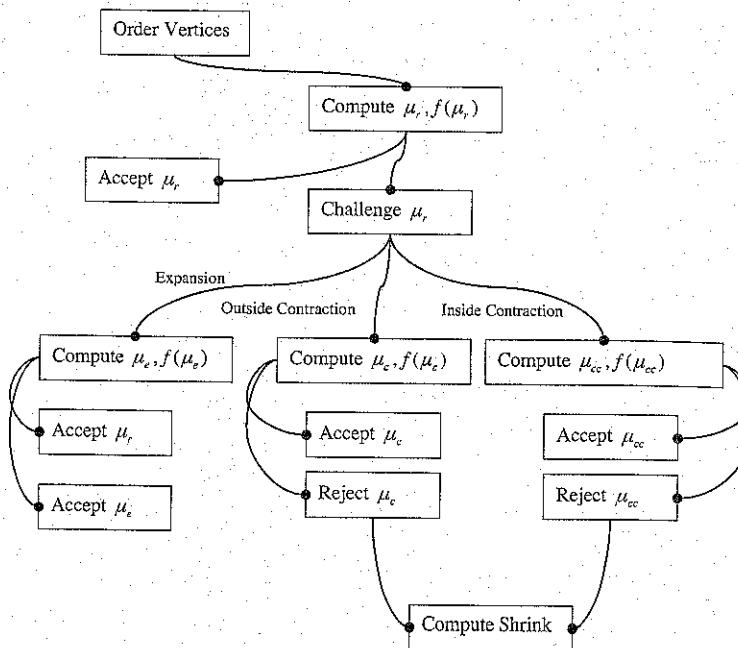


Figure 13.7 Flow Chart for Simplex Method Iteration

If  $f(\mu_c) \leq f(\mu_r)$ ,  $\mu_c$  is accepted; otherwise a shrink step is performed. If instead  $f(\mu_r) \geq f(\mu_{k+1})$ , an inside contraction point is computed:

$$\begin{aligned} \mu_{cc} &= \bar{\mu} - \gamma(\bar{\mu} - \mu_{k+1}) \\ &= (1 - \gamma)\bar{\mu} + \gamma\mu_{k+1} \rightarrow f(\mu_{cc}). \end{aligned} \quad (13.11)$$

If  $f(\mu_{cc}) < f(\mu_{k+1})$ ,  $\mu_{cc}$  is accepted; otherwise a shrink step is performed. If a shrink step is required, a new set of points

$$v_l = \mu_1 + \sigma(\mu_l - \mu_1), \quad \sigma \in (0, 1), \quad l = 1, \dots, k+1$$

is computed, resulting in the simplex constructed using  $(\mu_1, v_2, \dots, v_{k+1})$ . Upon completion of an iteration, a new iteration begins with the reordering of vertices as indicated in (13.6). Note that the simplex procedure we have described for the  $k$ -dimensional case is governed by the set of parameters  $(\rho, \chi, \gamma, \sigma)$ . Values of these parameters corresponding to the Nelder-Mead simplex are given by

$$\rho = 1, \chi = 2, \gamma = \frac{1}{2}, \sigma = \frac{1}{2}.$$

For cases in which ties result in the comparisons described above, Lagarias et al. propose two tie-breaking rules: one for shrink steps and one for nonshrink steps. Under the former, if there is a tie between the best point retained ( $\mu_1$ ) and any one of the new points, the best point is retained for the next iteration. Specifically:

$$f(\mu_1^{(i)}) = \min\{f(v_2^{(i)}), \dots, f(v_{k+1}^{(i)})\} \rightarrow \mu_1^{(i+1)} = \mu_1^{(i)}.$$

Under the latter, the worst vertex  $\mu_{k+1}^{(i)}$  is dropped in favor of an accepted point, say  $v^{(i)}$ , and if there is a tie with any of the remaining vertices, the accepted point takes position  $m$  in the vertices of the simplex  $\Delta_k^{(i+1)}$ , where

$$m = \max_{0 \leq n \leq k} [m | f(v^{(i)}) < f(\mu_{n+1}^{(i)})].$$

The remaining vertices retain their positions.

Regarding convergence, two tolerance measures are set in order to determine whether a minimum has been reached. The first measure is with respect to the "diameter" of the simplex  $\Delta_k^{(i)}$ :

$$\text{diam}(\Delta_k^{(i)}) = \max_{l \neq m} \|\mu_l^{(i)} - \mu_m^{(i)}\|, \quad (13.12)$$

$$\|\mu_l^{(i)} - \mu_m^{(i)}\| = \sqrt{\sum_{j=1}^k (\mu_{l,j}^{(i)} - \mu_{m,j}^{(i)})^2}. \quad (13.13)$$

When this diameter reaches a predetermined level, typically equaling  $10^{-4}$ , the first criterion has been reached. The second measure is with respect to the function values  $f^{(i)}$ . When iterations reach a point such that the difference  $f^{(i+1)}(\mu_1^{i+1}) - f^{(i)}(\mu_1^i)$  reaches a predetermined level, again typically equaling  $10^{-4}$ , the iterations are terminated and the algorithm is declared to have converged. The vertex  $\mu_1$  obtained in the final iteration is taken as the estimate  $\hat{\mu}$ .

The output of non-derivative-based methods is a vector of estimates  $\hat{\mu}$  and the corresponding log-likelihood value  $f(\hat{\mu}) = \log L(X|\hat{\mu})$ . The associated covariance matrix

$$\hat{\Sigma}_{\mu} = \left[ -\frac{\partial^2 \log L(X|\hat{\mu})}{\partial \hat{\mu} \partial \hat{\mu}'} \right]^{-1} \quad (13.14)$$

must typically be computed numerically. This may be achieved using a nested gradient-computation procedure. Specifically, letting  $\iota(\hat{\mu})$  denote the gradient of  $\log L(X|\mu)$  calculated at  $\hat{\mu}$ ,  $\Sigma_{\mu}$  may be obtained by calculating the gradient of  $\iota(\hat{\mu})$  at  $\hat{\mu}$ .<sup>2</sup>

We conclude by noting that because simplex methods do not rely on the calculation of derivatives in seeking a minimum, they tend to converge relatively quickly in comparison to their derivative-based counterparts. Also, they are not as susceptible to problems arising from the presence of discontinuities in the likelihood surface. Although they do not naturally produce standard errors as by-products of their search methodology, this can be overcome using finite-difference methods (see Judd, 1998, for a discussion of numerical approaches to computing gradients and Hessians). Alternatively, Harvey (1999) suggests that such methods can be used to identify attractive initial values for  $\mu$ , which can then be fed into derivative-based methods to obtain estimated standard errors. We now turn to a discussion of these methods.

### 13.3.2 Derivative-Based Methods

Derivative-based methods are patterned after the heuristic characterization of function optimization provided in section 2. (Several versions of derivative-based methods exist. For textbook discussions, see Judd, 1998; Hamilton, 1994; and Harvey, 1999).<sup>3</sup> Recall that if  $\hat{\mu}$  minimizes  $f(\mu)$ , it satisfies the first-order condition

$$\left. \frac{\partial f}{\partial \mu} \right|_{\mu=\hat{\mu}} = \iota(\hat{\mu}) = 0;$$

and the associated Hessian  $\Sigma(\hat{\mu})$  is positive definite (provided  $\iota(\mu)$  and  $\Sigma(\mu)$  exist and are continuous). Now consider the second-order Taylor Series expansion of the function to be minimized around the “true” value  $\hat{\mu}$ :

$$f(\mu) \approx f(\hat{\mu}) + (\mu - \hat{\mu})\iota(\hat{\mu}) + \frac{1}{2}(\mu - \hat{\mu})'\Sigma(\hat{\mu})(\mu - \hat{\mu}).$$

Differentiating this expansion with respect to  $\mu$  and imposing the first-order condition for a minimum, we obtain the following iterative scheme commonly known as Newton’s Method:

$$\mu^{i+1} = \mu^i - \Sigma^{-1}(\mu^i)\iota(\mu^i). \quad (13.15)$$

Iterations on (13.15) will converge to  $\hat{\mu}$  provided that the Hessian is always positive definite, a condition that is in no way guaranteed. If the

<sup>2</sup>The GAUSS procedure `hessp` is available for this purpose.

<sup>3</sup>Code for executing these methods is available in the Maximum Likelihood toolbox of GAUSS.

### 13.3 A Primer on Optimization Algorithms

Hessian is ever singular in a given iteration, the scheme will collapse. Alternatively, if the Hessian ever fails to be positive definite, the scheme can generate increasing rather than decreasing values of  $f(\mu)$ . Indeed, if the step size produced by (13.15) turns out to be excessively large, such “overshooting” can occur even if the Hessian is positive definite. Optimization algorithms that incorporate solutions to these problems fall under the general heading of quasi-Newton methods; we now discuss such methods.

In order to overcome problems of “overshooting,” an additional parameter known as a step length ( $\lambda$ ) is introduced:

$$\mu^{i+1} = \mu^i - \lambda \Sigma^{-1}(\mu^i)\iota(\mu^i). \quad (13.16)$$

If  $\Sigma$  is positive definite and  $\mu^i$  is not a minimum, then, as noted by Harvey (1999), there always exists a positive  $\lambda$  such that  $f(\mu^{i+1}) < f(\mu^i)$ , which helps keep the iterations going until convergence. A common technique for computing  $\lambda$  is the solution to

$$\min_{\lambda} f(\mu^i + \lambda \iota(\mu^i)),$$

which is computed at each iteration.

To avoid problems associated with non-positive definite Hessians, quasi-Newton methods replace  $\Sigma(\mu^i)$  with alternative matrices that mimic the role of the Hessian, but that are guaranteed to be positive definite. Two popular alternatives are due to Broyden, Fletcher, Goldfarb, and Shanno (BFGS) and Davidson, Fletcher, and Powell (DFP). Letting

$$D^i = \Sigma^{-1}(\mu^i), \quad p = (\mu^{i+1} - \mu^i),$$

and

$$q = \iota(\mu^{i+1}) - \iota(\mu^i),$$

the BFGS alternative constructs  $D$  iteratively using

$$D^{i+1} = D^i + \frac{pp'}{p'q} \left[ 1 + \frac{q'D^iq}{p'q} \right] - \frac{Dqp' + pq'D}{p'q}. \quad (13.17)$$

The DFP alternative constructs  $D$  iteratively using

$$D^{i+1} = D^i + \frac{pp'}{p'q} - \frac{Dqq'D}{q'Dq}. \quad (13.18)$$

In both cases, initial specifications of the Hessian are typically chosen as the identity matrix, which is positive definite. Under either alternative, the approximation obtained in the final iteration is used to construct

the estimated variance-covariance matrix associated with  $\hat{\mu}$ , as indicated in (13.5).

With respect to convergence, two criteria are often simultaneously used. Letting  $\mu^i$  represent a proposed minimum, the criteria involve comparisons of the minimum elements of  $(\mu^{i+1} - \mu^i)$  and  $\iota(\mu^i)$  with prespecified thresholds, typically  $10^{-4}$ . Heuristically this translates into a check of whether the sequence of estimates has converged and the first-order condition is satisfied. As noted, because they require the continuous calculations of gradients and Hessians, derivative-based methods tend to be slow to converge relative to simplex methods. However, they do produce estimates of both parameter values and associated standard errors as a by-product of the optimization process, so they are attractive in this regard.

Additional considerations regarding the choice of simplex versus derivative-based methods arise when  $f(\mu)$  is not well-behaved, as we have assumed throughout this section. Problems associated with  $f(\mu)$  fall into three general categories: discontinuities, local minima, and flat surfaces. We now turn to a discussion of these problems, and methods for overcoming them.

### 13.4 Ill-Behaved Likelihood Surfaces: Problems and Solutions

The preceding presentations have taken for granted the existence of well-behaved likelihood surfaces. This of course is not always the case in practice, and as noted, implementation can be particularly challenging in applications involving DSGE models. Specializing to use of log-linear model approximations, the primary source of complication arises from the fact that although optimization is with respect to  $\mu$ , the likelihood function itself is parameterized in terms of  $(F(\mu), G(\mu), Q(\mu), H(\mu))$ , and as we have seen, mappings from  $\mu$  into the parameters of these matrices tend to be complex, highly nonlinear, and analytically intractable. Here we discuss three classes of problems that can arise in minimizing a function: discontinuities, multiple local minima, and identification.

#### 13.4.1 Problems

Loosely, a likelihood surface contains discontinuities if it features “walls” or “cliffs,” so that a small change in a parameter value results in a jump in the value of the function. At such a point the function is not differentiable, thus posing an obvious challenge in the implementation of derivative-based methods. Prepackaged quasi-Newton software routines typically produce

#### 13.4 Ill-Behaved Likelihood Surfaces

as output warnings when discontinuities are encountered. Alternatively, grid plots provide an effective diagnostic tool. In fact, such plots are effective in identifying all three types of problems that can arise in seeking to minimize a function; thus a brief digression is worthwhile.

Grid plots are merely graphical explorations of the likelihood surface. Two-dimensional plots are constructed by fixing all but the  $j^{\text{th}}$  element  $\mu^j$  of the parameters at particular values, and tracing  $f(\mu)$  over a grid of values specified for  $\mu^j$ . Three-dimensional plots are constructed analogously by tracing  $f(\mu)$  over ranges specified for two elements of  $\mu$ . Correlations among the elements of  $\mu$  can pose challenges in effectively constructing and interpreting plots, because they will render particular plots sensitive to values specified for the fixed elements of  $\mu$ . However, initial estimates of the Hessian can be helpful in identifying key dimensions along which correlations exist, and thus dimensions along which the sensitivity of plots can be usefully explored. Even in cases in which convergence to a minimum appears to have been achieved successfully, the construction of grid plots in the neighborhood of the optimum is an advisable practice.

Discontinuities pose less of a challenge in implementing simplex methods, because the calculation of reflection points, expansion points, and so on is unencumbered by their existence. Due to this fact, Christopher Sims has developed a hybrid optimization algorithm that combines the derivative-based BFGS method with a simplex algorithm that is triggered when discontinuities are encountered.<sup>4</sup>

When  $f(\mu)$  features local minima, the obvious possibility arises that the estimate  $\hat{\mu}$  delivered by either class of optimization routine is not in fact a global minimum. Once again, grid plots constructed using generous ranges for parameters are helpful in detecting such instances. In addition, investigations of the sensitivity of  $\hat{\mu}$  to the use of alternative starting values can be used for this purpose.

Finally, an identification problem is said to exist when alternative specifications of  $\mu$  yield the same value of  $f(\mu)$ . Colloquially, identification problems exist if the likelihood function is flat along any dimension of the parameter space. In the context of evaluating state-space representations using the Kalman filter, identification problems are manifested by the inability to invert the matrix  $\Omega_{t|t-1}$ .

#### 13.4.2 Solutions

Just as DSGE models pose special challenges in the context of maximum-likelihood analysis, they also offer special solutions. In particular, it is most

<sup>4</sup> Software facilitating its implementation is available on his website: <http://www.princeton.edu/~sims/>.

often the case that the elements of  $\mu$  are subject to a priori restrictions that, when imposed, can be helpful in avoiding the problems mentioned above. For example, parameters representing discount factors, depreciation rates, capital's share of output, and so on, must all lie between zero and one, and moreover, only fractions of this range warrant serious consideration. In other words, guidance from theory is typically available in imposing restrictions on ranges of the parameter space over which optimization searches are conducted. Moreover, the imposition of such restrictions is straightforward via the use of parameter transformations.

Parameter transformations are implemented by establishing relationships of the form

$$\mu = \mu(\omega), \quad (13.19)$$

where  $\omega$  is  $l \times 1$ , and optimizing  $f(\mu(\omega))$  with respect to  $\omega$ . Given estimates  $(\hat{\omega}, \hat{\Sigma}_\omega)$  generated by optimizing over  $\omega$ , corresponding estimates for  $\mu$  are given by

$$\hat{\mu} = \mu(\hat{\omega}), \quad \hat{\Sigma}_\mu = [\mu'(\hat{\omega})] \hat{\Sigma}_\omega [\mu'(\hat{\omega})]', \quad (13.20)$$

where  $[\mu'(\hat{\omega})]$  is the  $k \times l$  Jacobian matrix evaluated at  $\hat{\omega}$ .

Three goals should be considered in establishing useful parameter transformations. First, the transformations should eliminate the need to impose additional restrictions on ranges for the individual elements  $\omega^j$  over which the search is to be conducted. Second, the transformations should be made so that  $f(\mu(\omega))$  is reasonably responsive to changes in the individual elements  $\omega^j$ . For example, for a one-to-one mapping between  $\mu^j$  and  $\omega^j$ , if  $\partial \mu^j / \partial \omega^j$  is nearly zero in the neighborhood of the minimum, it will be difficult for any optimization routine to pin down the minimizing value. Third, the Hessian associated with  $\omega$  should be "balanced": the sums of its columns (and rows) should be roughly equal. Convergence is greatly enhanced by the use of transformations that help achieve balance.

Parameter transformations of particular relevance are as follows. If it is desired that an individual element  $\mu^j$  be constrained to lie in the range  $(a, b)$ , but no information is available regarding where in this range an optimal value might lie, the following logistic transformation is useful:

$$\mu^j = a + \frac{b - a}{1 + e^{\omega^j}}.$$

A starting value of  $\omega_0^j = 0$  will place  $\mu^j$  at the midpoint between  $a$  and  $b$ . If it turns out that optimization fails given this specification, it may be the case that it is not well suited for meeting the second or third goals. For example, if the optimal value of  $\mu^j$  lies near one of its boundaries, then

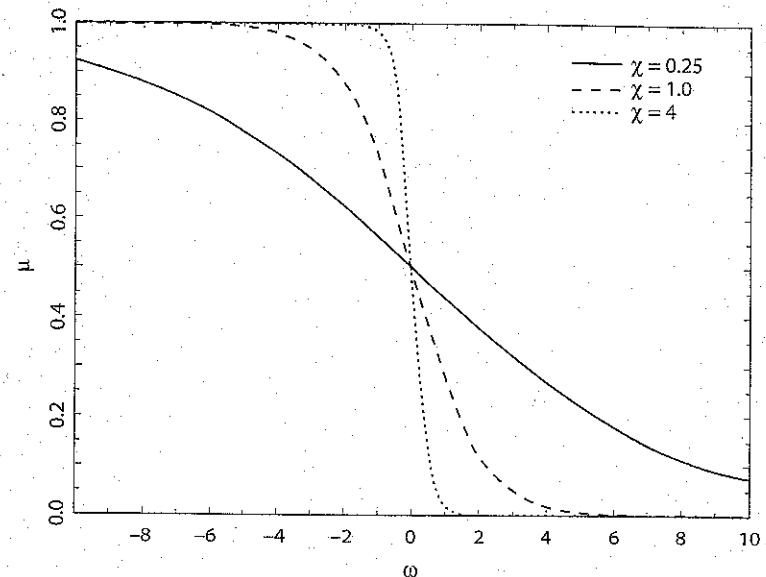


Figure 13.8 Logistic Transformations

$\partial \mu^j / \partial \omega^j$  will be nearly zero. In this case it may be useful to introduce an additional parameter  $\chi$ :

$$\mu^j = a + \frac{b - a}{1 + e^{\chi \omega^j}}.$$

As illustrated in figure 13.8 (constructed using  $a = 0$ ,  $b = 1$ ), as  $\chi$  decreases,  $\partial \mu^j / \partial \omega^j$  increases near the boundaries. Problems with balance can also be addressed given the introduction of  $\chi$ .

Finally, it may also be useful to introduce a centering parameter  $c$ :

$$\mu^j = a + \frac{b - a}{1 + e^{c(\omega^j + c)}}. \quad (13.21)$$

If a good starting value for  $\mu^j$  is known a priori, denoted  $\mu_0^j$ , then zero will be a good starting value for  $\omega^j$  given the specification

$$c = \frac{1}{\chi} \log \left( \frac{b - a}{\mu_0^j - a} - 1 \right); \quad (13.22)$$

moreover,  $\widehat{\omega}^j$  will be approximately zero if the starting value  $\mu_0^j$  turned out to lie near  $\widehat{\mu}^j$ .

For a parameter restricted merely to be positive, a useful transformation is

$$\mu^j = e^{x\omega^j}, \quad (13.23)$$

where again  $x$  can be adjusted to deal with cases under which  $\mu^j$  lies near its boundary, or to help achieve balance.

If any of the problems mentioned above cannot be overcome by restricting ranges over the elements of  $\mu$ , four additional potential remedies are available. First, the outcome of calibration or method-of-moments exercises conducted prior to estimation can help with the specification of effective starting values  $\mu_0$ , or be used to impose additional restrictions for conducting the optimization search. Second, multistep optimization can be used. The idea here is to split the parameter vector  $\mu$  into subvectors (say,  $\mu^1$  and  $\mu^2$  for a two-step procedure), and to optimize over each subvector in turn. Sequential iteration can help lead to overall convergence. Third, a subvector of  $\mu$  can simply be fixed (e.g., at calibrated values), and optimization conducted over the remaining elements.

Finally, quasi-Bayesian methods can be used in order to help induce curvature in the likelihood function. Such methods involve augmenting the log-likelihood function with a prior distribution specified over  $\mu$ , denoted by  $\pi(\mu)$ , and minimizing

$$-\log L(X|\mu) - \log \pi(\mu).$$

In this context,  $\log \pi(\mu)$  can be thought of a penalty function. Ruge-Murcia (2003) evaluated the use of this approach in the context of estimating a DSGE model via maximum likelihood, and found it useful as a gentle means of imposing a priori restrictions in the exercise.

### 13.5 Model Diagnostics and Parameter Stability

Having obtained estimates of  $\mu$  and associated standard errors, model diagnostics can be conducted. Here we discuss four types of diagnostics: tests of restrictions imposed on individual elements of  $\mu$ , tests of restrictions imposed across multiple elements of  $\mu$ , tests of the stability of  $\mu$  over time, and given evidence of temporal parameter instability, a procedure designed to help pinpoint the date at which a structural break is likely to have occurred.

With  $\widehat{\sigma}_{\mu^j}$  denoting the standard error associated with  $\widehat{\mu}^j$ , the statistical significance of the difference between  $\widehat{\mu}^j$  and  $\bar{\mu}^j$  can be assessed using the  $t$ -statistic:

$$t^j = \frac{\widehat{\mu}^j - \bar{\mu}^j}{\widehat{\sigma}_{\mu^j}}. \quad (13.24)$$

The statistic  $t^j$  measures the distance of  $\widehat{\mu}^j$  from  $\bar{\mu}^j$  in units of the estimated standard error  $\widehat{\sigma}_{\mu^j}$ . Under the null hypothesis that  $\mu^j = \bar{\mu}^j$ ,  $t^j$  is distributed as Student's  $t$  with  $T - 1$  degrees of freedom, where  $T$  denotes the sample size. Large values of  $|t^j|$  are assigned low probabilities, or  $p$ -values, under this distribution; when  $p$ -values lie below a prespecified critical value, the null hypothesis is rejected; otherwise it is accepted.

Consider now a collection of restrictions imposed upon  $\mu$  of the form

$$g(\mu) = 0, \quad (13.25)$$

where  $g(\mu)$  is an  $l$ -valued differentiable function, with  $l < k$ . Letting  $\widehat{\mu}_R$  denote the value of  $\mu$  that minimizes  $f(\mu)$  subject to (13.25), the null hypothesis that the restrictions in fact hold can be tested using the likelihood ratio statistic:

$$LR = 2 \log[L(X|\widehat{\mu}) - L(X|\widehat{\mu}_R)].$$

Under the null,  $LR$  is distributed as  $\chi^2(l)$ . As the gap between the unrestricted and restricted likelihood functions widens,  $LR$  increases, and once again its associated  $p$ -value falls. As with the  $t$ -test, if the  $p$ -value falls below a prespecified threshold, the null is rejected; otherwise it is accepted.

Regarding implementation, consider the case in which (13.25) can be rewritten as

$$\mu = r(\omega), \quad (13.26)$$

where  $\omega$  is an  $\ell$ -vector such that  $\ell = k - l$ . Then  $L(X|\widehat{\mu}_R)$  can be calculated by maximizing over  $\omega$ , and setting  $\widehat{\mu}_R = r(\widehat{\omega})$ .

A special adaptation of the  $LR$  test can be used to conduct tests of temporal parameter stability. Let  $\widehat{\mu}_1$  and  $\widehat{\mu}_2$  denote parameter estimates obtained over two disjoint but fully inclusive ranges of the sample period. That is,  $\widehat{\mu}_1$  is estimated over the time span  $[1, t_1]$ , and  $\widehat{\mu}_2$  is estimated over the time span  $[t_1 + 1, T]$ . Letting  $X_1$  and  $X_2$  denote respective samples spanning these time intervals, the  $LR$  test of the null hypothesis that  $\mu_1 = \mu_2$  is given by

$$\mathcal{L} = 2[\log L(X_1|\widehat{\mu}_1) + \log L(X_2|\widehat{\mu}_2) - \log L(X|\widehat{\mu})]. \quad (13.27)$$

For the specification of a single candidate breakpoint  $t_1$ ,  $\mathcal{L}$  will be distributed as  $\chi^2(k)$ , and the null hypothesis can be assessed as described

above. However, if the researcher conducts a search over alternative possible break dates, for example, by searching for the date at which the associated  $p$ -value is minimized, then the resulting distribution of  $\mathcal{L}$  will not be distributed as  $\chi^2(k)$ . Andrews (1993) and Hansen (1997) have developed formal testing procedures that enable specification searches over  $t_1$ , but they are impractical to conduct in this context because they involve calculations of  $L(X_1 | \hat{\mu}_1)$  and  $L(X_2 | \hat{\mu}_2)$  for each candidate break date, and such calculations can be extremely time-consuming in working with DSGE models.

However, progress can be made in this context if in place of a formal test for the occurrence of a break over a range of possible dates, we consider the alternative objective of calculating conditional probabilities associated with the occurrence of a break over a range of dates. For this purpose, a procedure developed by DeJong, Liesenfeld, and Richard (2005) for timing the occurrence of a *known* break can be adopted. Their procedure is designed for implementation after a rejection of the structural-stability hypothesis generated by a preliminary test is attained, for example, using the testing strategy developed by Andrews (1993) and Hansen (1997). We proceed in this context under the caveat that the results of such a pretest will in general be unavailable.

DeJong et al.'s motivation for developing their procedure is that although the structural-stability tests of Andrews and Hansen are effective in identifying the occurrence of breaks, they are ineffective in pinpointing timing. The adoption of the relatively modest objective of calculating conditional probabilities associated with breaks, in place of the objective of seeking to reject the null hypothesis of structural stability, accounts for the relative effectiveness of DeJong et al.'s procedure in pinpointing timing.

The procedure requires two sets of ML estimates:  $\hat{\mu}_1$  calculated over an early portion of the sample  $[1, T_1]$ , and  $\hat{\mu}_2$  calculated over a late portion of the sample  $[T_2 + 1, T]$ . The range of dates over which potential break dates are contained is thus  $[T_1 + 1, T_2]$ . For each  $j \in [T_1 + 1, T_2]$ , consider a division of the entire sample into two intervals:  $\{X_t\}_{t=1}^j$  and  $\{X_t\}_{t=j+1}^T$ . The likelihood value associated with this division of the sample is given by

$$L(j) = L(\{X_t\}_{t=1}^j | \hat{\mu}_1) \times L_2(\{X_t\}_{t=j+1}^T | \hat{\mu}_2). \quad (13.28)$$

Notice that if there is a distinct difference between the alternative likelihood functions parameterized under  $\hat{\mu}_1$  and  $\hat{\mu}_2$  due to a structural break, then  $L(j)$  will be sensitive to changes in  $j$  that improve upon or worsen the alignment of the division of the sample with the date at which the break has occurred.

### 13.6 Empirical Application

Consider now the ratio of  $L(j)$  to the sum of likelihoods calculated for every possible division of the sample between  $[T_1 + 1, T_2]$ :

$$p(j) = \frac{L(j)}{\sum_{\tau=T_1+1}^{T_2} L(\tau)}. \quad (13.29)$$

Given the known occurrence of a break between  $[T_1 + 1, T_2]$ ,

$$\sum_{\tau=T_1+1}^{T_2} L(\tau)$$

represents the likelihood associated with the full range of possible scenarios regarding a break, and thus  $p(j)$  is appropriately interpreted as the probability associated with a break at date  $j$ . Absent the known occurrence of a break, it is still possible to assess the relative size of  $p(j)$  across alternative values of  $j$ . A relatively flat profile signals the absence of a break, whereas abrupt peaks in the profile indicate potential breaks.

### 13.6 Empirical Application: Identifying Sources of Business Cycle Fluctuations

To demonstrate the implementation of ML techniques for estimating DSGE models, we build upon Ireland's (2004a) analysis of the New Keynesian model introduced in section 2 of chapter 3. Recall that this model is an extension of a standard RBC model under which demand shocks, cost-push shocks, and shocks to the central bank's monetary policy function combine with technology shocks in driving business-cycle fluctuations. Using ML estimation techniques, the empirical objective Ireland pursued in estimating this model was to assess the relative importance of these shocks in accounting for fluctuations in output growth, inflation, and the output gap. He pursued this objective by constructing impulse response functions and variance decompositions (introduced in chapter 7), which are designed to measure responses of these series to innovations in the respective shocks.

Here we extend this analysis using the smoothing feature of the Kalman filter introduced in chapter 8. Specifically, conditional on the model estimates obtained by Ireland, we back out the implied time-series observations on innovations to the shocks, along with the shocks themselves, and study the behavior of these measured series over stages of the business cycle. In part, our intent is to depict smoothing as an important by-product of full-information analyses based on the Kalman filter. Examples of

similar exercises in the RBC literature are given by DeJong, Ingram, and Whiteman (2000a) and Chari, Kehoe, and McGrattan (2004). It bears emphasizing that in any such exercise, inferences regarding the behavior of model unobservables are conditional not only on the behavior of the corresponding observable variables, but also on the specification of the model under investigation.

To be clear regarding our terminology, the demand, cost-push, and technology shocks follow AR representations of the form

$$s_t = \rho_s s_{t-1} + \varepsilon_{st}, \quad \varepsilon_s \sim N(0, \sigma_s^2).$$

When we refer to the shocks themselves, we refer to  $s_t$ ; when we refer to innovations to the shocks, we refer to  $\varepsilon_{st}$ . Shocks to the monetary policy function are *iid*, thus there is no distinction between innovations and shocks in this case.

We begin with an overview of Ireland's analysis. Using quarterly data over the period 1948.I–2003.I on output growth, inflation, and short-term nominal interest rates, Ireland sought to obtain ML estimates of the full set of parameters of the model.<sup>5</sup> Denoting the demand, cost-push, technology, and policy shocks as  $(\tilde{a}, \tilde{\varepsilon}, \tilde{z}, \varepsilon_r)$ , the key equations of the model are the IS curve, the Phillips curve, and the Taylor Rule (the full specification of the model is given in section 2 of chapter 3):

$$\tilde{o} = \alpha_o \tilde{o}' + (1 - \alpha_o) E_t \tilde{o}' - (\tilde{r} - E_t \tilde{r}') + (1 - \omega)(1 - \rho_a) \tilde{a} \quad (13.30)$$

$$\tilde{\pi} = \beta \alpha_\pi \tilde{\pi}' + \beta (1 - \alpha_\pi) E_t \tilde{\pi}' + \psi \tilde{o} - \tilde{\varepsilon}, \quad (13.31)$$

$$\tilde{r}' = \rho_r \tilde{r} + \rho_\pi \tilde{\pi}' + \rho_g \tilde{g}' + \rho_o \tilde{o}' + \varepsilon'_r. \quad (13.32)$$

Here, tildes denote logged departures of variables from steady state values,  $x'$  denotes the period- $(t+1)$  observation of  $x$ , and  $x^-$  denotes the period- $(t-1)$  observation of  $x$ . Inflation, nominal interest rates, and the output gap are denoted as  $(\tilde{\pi}, \tilde{r}, \tilde{o})$ , where  $\tilde{o} = \tilde{y} - \omega \tilde{w}$ ;  $\tilde{y}$  denotes output, and  $\omega$  denotes the inverse of the elasticity of labor supply.

Sample means were removed from the data prior to analysis. In addition, sample means were used to pin down the parameters  $z$  and  $\pi$ , which serve to identify steady state values of output growth and inflation:  $z = 1.0048$ , which matches the averaged annualized growth rate observed for output of 1.95%; and  $\pi = 1.0086$ , matching the average annual inflation rate of 3.48%. Given the specification of  $\pi$ , the sample mean of nominal interest rates in principle serves to identify the real steady state interest rate, and thus the discount factor  $\beta$ . However, identifying  $\beta$  in this fashion implies an implausible value in excess of 1, which is a manifestation of the risk-free rate puzzle discussed in section 3 of chapter 3. In response, Ireland

<sup>5</sup> The data and computer code he used are available at [www2.bc.edu/~irelandp/](http://www2.bc.edu/~irelandp/).

### 13.6 Empirical Application

broke the link between nominal interest rates and  $\beta$  by fixing  $\beta$  at 0.99, and removing the mean nominal rate of 5.09% from the data irrespective of this specification. Finally, the policy parameter  $\rho_r$  was fixed at unity; thus deviations from steady state levels of inflation, output growth, and the output gap have persistent affects on interest rates.

Ireland's initial attempts at estimating the remaining parameters resulted in estimates of  $\psi$ , the output-gap parameter in the Phillips curve, that he deemed to be unreasonably small (implying enormous nominal price-adjustment costs). Thus this parameter was also fixed prior to estimating the remaining parameters. The value chosen for this parameter was 0.1, implying that prices on individual goods are reset on average every  $3\frac{3}{4}$  quarters. Estimates of the remaining parameters, obtained conditionally on these specifications, are reported in table 13.1, which replicates Ireland's table 1, p. 928.

Regarding behavioral parameters, the small estimate obtained for  $\omega$  (the inverse of the elasticity of labor supply  $\xi$ ) implies a highly inelastic labor-supply specification. An implication of this estimate is that the efficient level of output is unresponsive to the demand shock. Next, the estimates of  $\alpha_o$  and  $\alpha_\pi$  are both statistically insignificant. Note that these are the parameters associated with lagged values of the output gap in the IS curve, and inflation in the Phillips curve. They were incorporated in the model to investigate whether the allowance for potential backward-looking behavior on the part of decision makers is important for helping characterize the

TABLE 13.1  
ML Estimates

Parameter	Estimate	Std. err.
$\omega$	0.0617	0.0634
$\alpha_o$	0.0836	0.1139
$\alpha_\pi$	0.0000	0.0737
$\rho_\pi$	0.3597	0.0469
$\rho_g$	0.2536	0.0391
$\rho_o$	0.0347	0.0152
$\rho_a$	0.9470	0.0250
$\rho_t$	0.9625	0.0248
$\sigma_a$	0.0405	0.0157
$\sigma_e$	0.0012	0.003
$\sigma_z$	0.0109	0.0028
$\sigma_r$	0.0031	0.003

Note: Source of estimates: Ireland (2004b), table 1, p. 928.

data. The insignificance of these parameters suggests that this allowance turns out to be unimportant. Regarding policy parameters, the estimates obtained for  $\rho_\pi$  and  $\rho_g$  indicate the responsiveness of policy to movements in inflation and output growth, whereas the estimate obtained for  $\rho_\theta$  indicates a relative lack of responsiveness to the output gap.

Turning to estimates regarding the shocks, all four standard deviation parameters are estimated as statistically significant, implying a nontrivial empirical role for each shock. Also, the AR parameters estimated for both the demand- and cost-push shocks are in the neighborhood of 1, suggesting that innovations to these series are extremely persistent (recall that the technology shock was modeled as having a unit root, and that the policy shock was modeled as being *iid*).

Ireland quantified implications of these parameter estimates for the role played by the shocks in driving cyclical fluctuations by calculating impulse response functions and forecast-error variance decompositions. Regarding movements in output growth, he found policy shocks to be most influential, accounting for 37% of the errors associated with forecasts of this series over various horizons; in comparison, technology, demand, and cost-push shocks accounted for 25%, 22%, and 15% of this error. In addition, cost-push shocks were found to have the greatest influence on inflation, and demand shocks on nominal interest rates.

We now underscore these findings by demonstrating the smoothing feature of the Kalman filter. Recall from section 4 of chapter 8 that given a parameterized version of the state-space representation, the time-series observations on the observable variables  $X_t$  can be used to obtain smoothed estimates of the unobservables  $\{\hat{x}_{t|T}\}_{t=1}^T$ , along with innovations to these variables  $\{\hat{e}_{t|T}\}_{t=1}^T$ . These smoothed time-series observations can be thought of as point estimates of a particular set of functions of the collection of structural parameters  $\mu$ . As such, standard errors corresponding to these estimates can be constructed using a gradient-based conversion method. Specifically, let the smoothed observation of the  $j^{\text{th}}$  element of  $\{\hat{x}_{t|T}\}_{t=1}^T$  be given by

$$\{\hat{x}_{t|T}^j\}_{t=1}^T = g(\hat{\mu}).$$

Passing this function along with  $\hat{\mu}$  to a gradient-calculation procedure (such as GAUSS's `gradp`) yields as output  $g'(\hat{\mu})$ , a  $T \times k$  Jacobian matrix evaluated at  $\hat{\mu}$ . The standard errors associated with  $\hat{x}_{t|T}^j$  for each date  $t$  can then be calculated as the square roots of the diagonal elements of the  $T \times T$  matrix  $[g'(\hat{\mu})] \Sigma_\mu [g'(\hat{\mu})]'$ :

$$\{\hat{\sigma}_{t|T}^j\}_{t=1}^T = \sqrt{\text{diag}([g'(\hat{\mu})] \Sigma_\mu [g'(\hat{\mu})]')'}$$

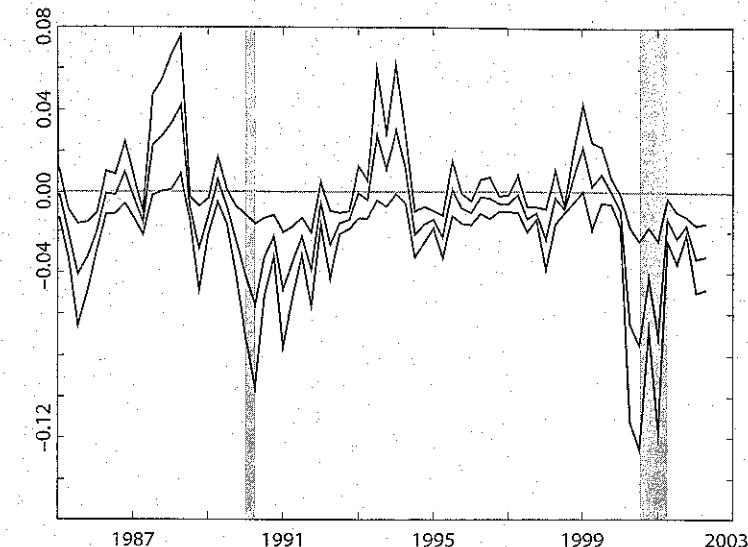


Figure 13.9 Smoothed Innovations to the Demand-Push Shock, and 95% Coverage Intervals

Finally, 95% error bands associated with the observations  $\{\hat{x}_{t|T}^j\}_{t=1}^T$  can be constructed as  $\{\hat{x}_{t|T}^j\}_{t=1}^T \pm 2\{\hat{\sigma}_{t|T}^j\}_{t=1}^T$ .

Representative point estimates and error bands obtained for innovations to the demand-shock process are illustrated in figure 13.9 over the period 1985:1–2003:1. Also depicted are NBER-defined recessions, indicated using vertical lines (beginning at business cycle peaks and ending at troughs). Note, for example, that both recessions observed during this period coincide with the realization of a sequence of negative realizations of these innovations that differ significantly from zero.

The full series of smoothed realizations obtained for the level of the demand shock are depicted in figure 13.10, and their associated innovations are depicted in figure 13.11. Hereafter, standard-error bands are suppressed to aid in the presentation of point estimates, and NBER-dated recessions are illustrated using shaded bars. Note that demand shocks tend to rise systematically as the business cycle moves from trough to peak, and then spike dramatically downwards during recessions. These measurements indicate that, conditional on the estimated model, demand shocks do seem to have played an important role in contributing towards movements over stages of the business cycle.

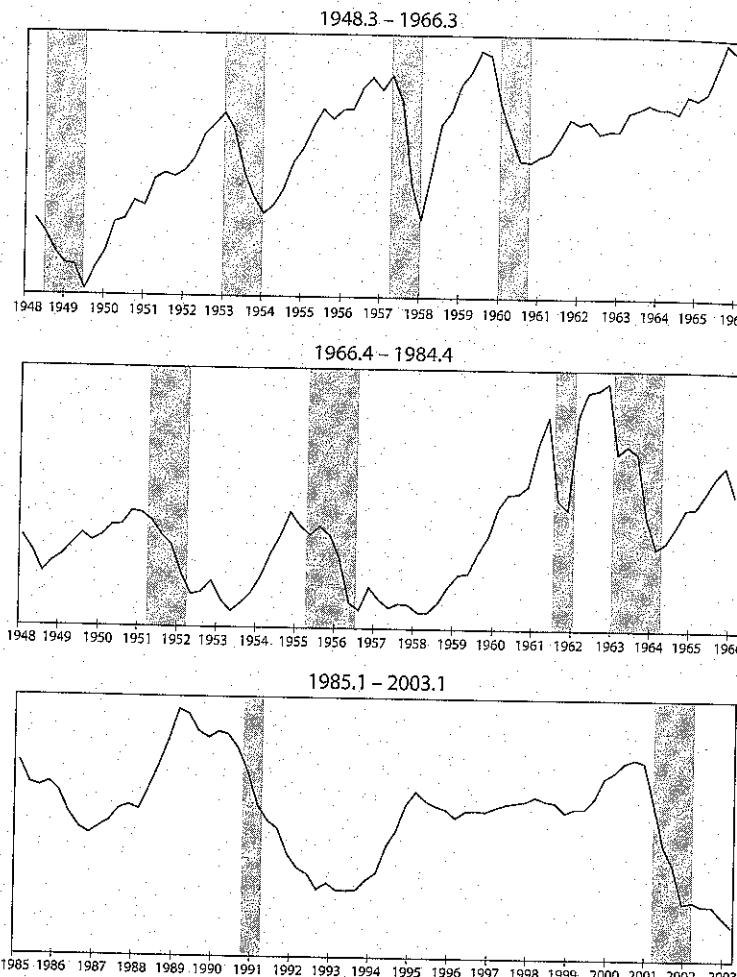


Figure 13.10 Smoothed Demand Shock

Smoothed observations of the cost-push shock and associated innovations are depicted in figures 13.12 and 13.13. Relative to the demand shock, movements in this case are less systematic over phases of the business cycle. However, it is interesting to note that large negative values are observed for this shock around 1974 and 1979–1980, which coincide with the realization of significant oil-price shocks.

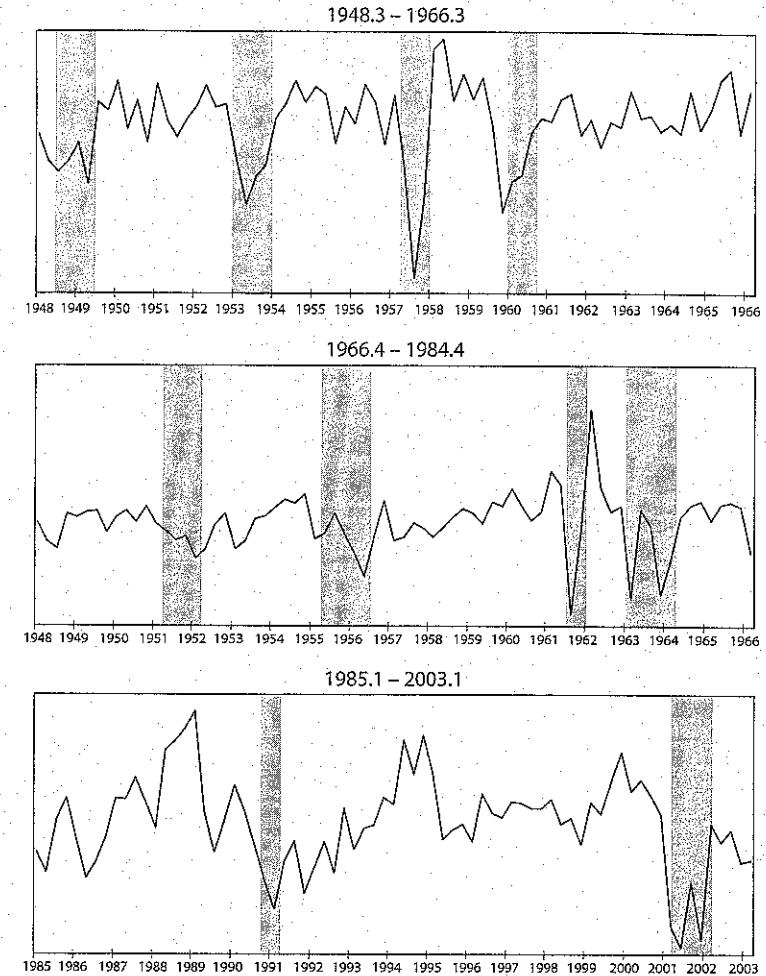


Figure 13.11 Smoothed Innovations to the Demand Shock

We next depict observations on the technology shock and its associated innovations in figures 13.14 and 13.15. Befitting its random-walk specification, this series exhibits wide swings that somewhat obfuscate movements at cyclical frequencies. However, the series does exhibit a clear tendency to drop sharply during recessions. Occasionally these drops slightly predate business cycle peaks; other times the drops coincide almost precisely.

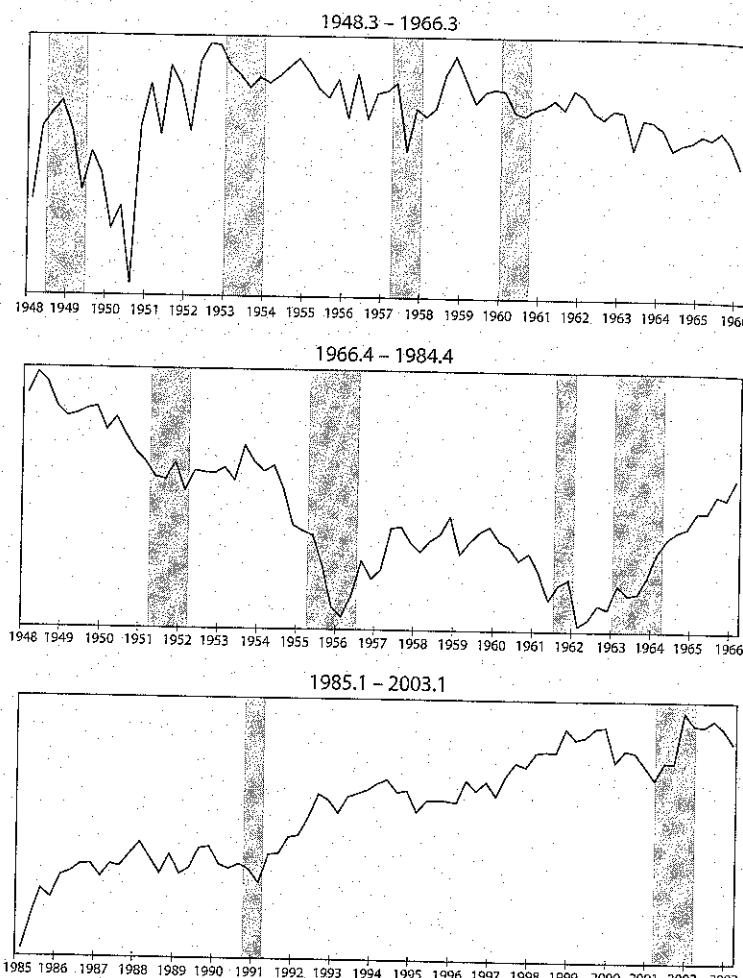


Figure 13.12 Smoothed Cost-Push Shock

Finally, observations on the policy shock are depicted in figure 13.16. Outside of recessions, the volatility of this series is somewhat lower in the early portion of the sample relative to the latter portion. However, throughout the sample the series shows a tendency to spike upwards during recessions, perhaps signaling purposeful responses of policy to these episodes.

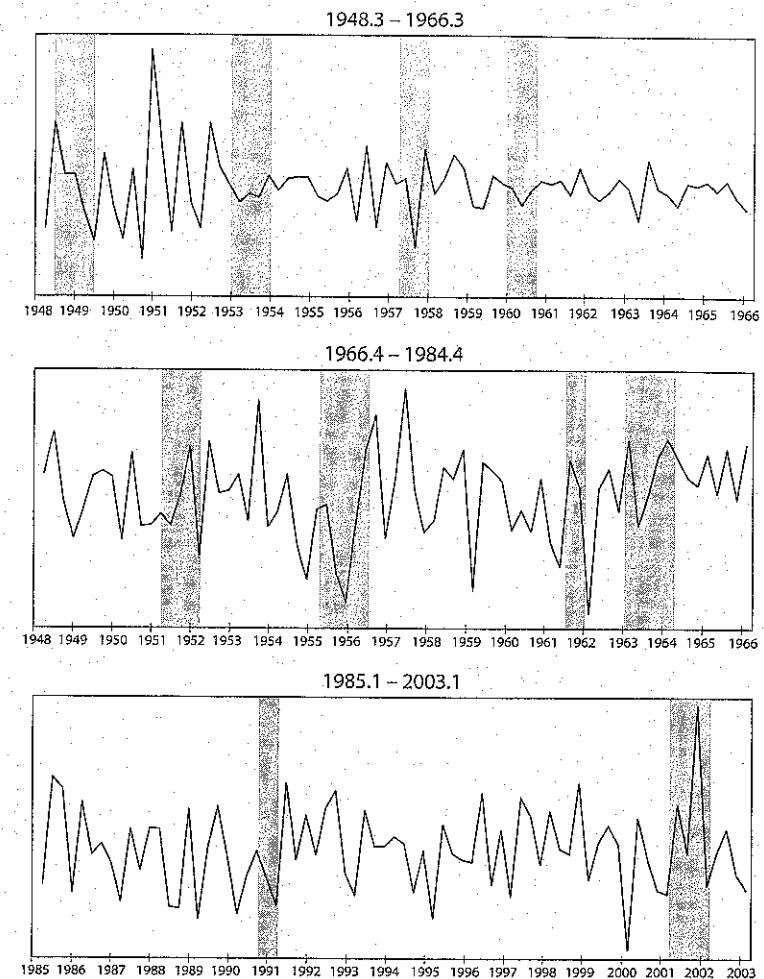


Figure 13.13 Smoothed Innovations to the Cost-Push Shock

To further explore the systematic nature of movements in these series over phases of the business cycle, we conducted a logit analysis under which we used current and lagged values of innovations to each series as explanatory variables used to model the occurrence of NBER-identified recessions. This follows a similar shock-identification analysis conducted

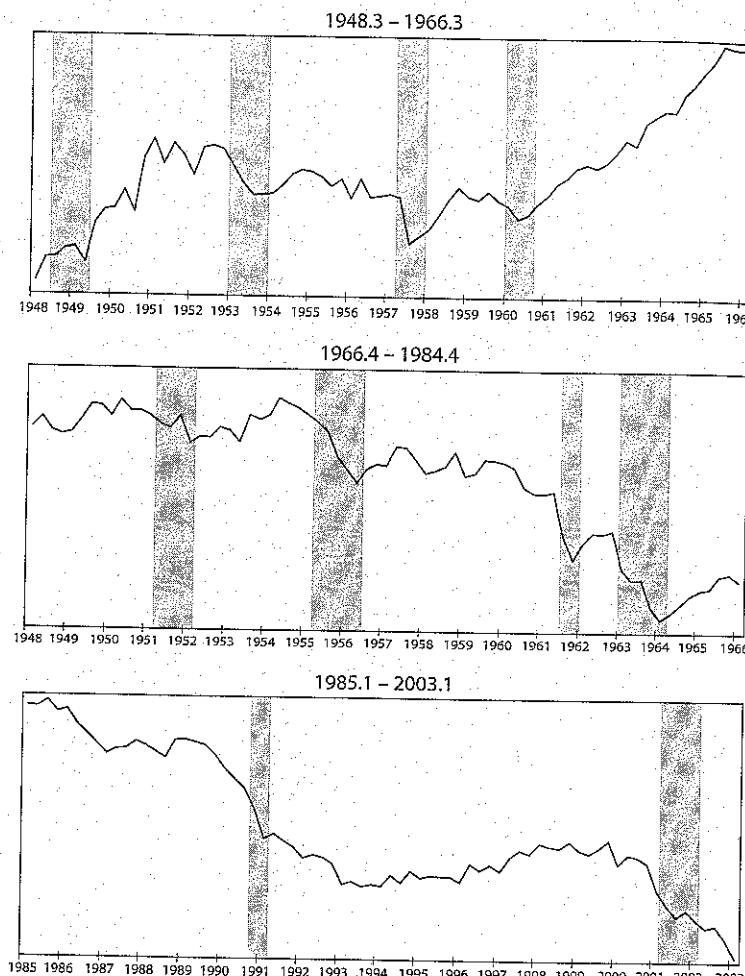


Figure 13.14 Smoothed Technology Shock

by DeJong, Ingram, and Whiteman (2000a). Specifically, we used as a dependent variable an indicator variable that is assigned the value of unity for dates coinciding with NBER-defined recessions and zero otherwise. We then modeled this series using contemporaneous and four lagged values of innovations to each shock. We considered five model specifications: one

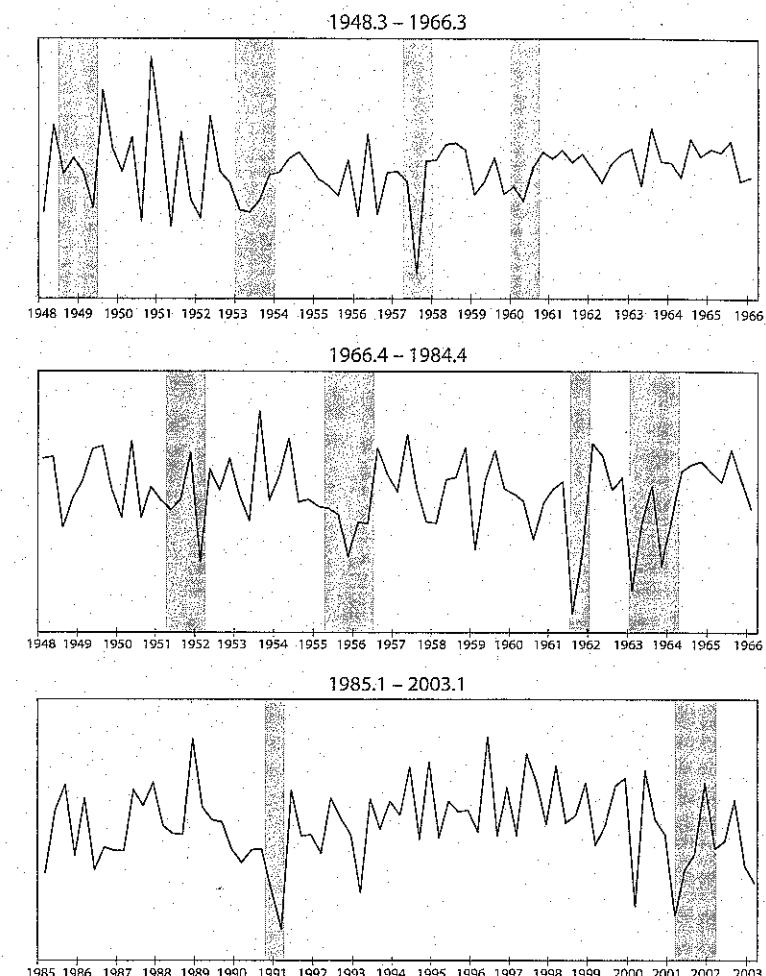


Figure 13.15 Smoothed Innovations to the Technology Shock

each using the individual series as explanatory variables, and one using all four series jointly as explanatory variables. Fitted values of the recession indicator, along with NBER dates, are illustrated in the five panels of figure 13.17.

Using the full set of explanatory variables, we obtained an excellent characterization of the occurrence of recessions. Using 0.75 and 0.5 as ad hoc

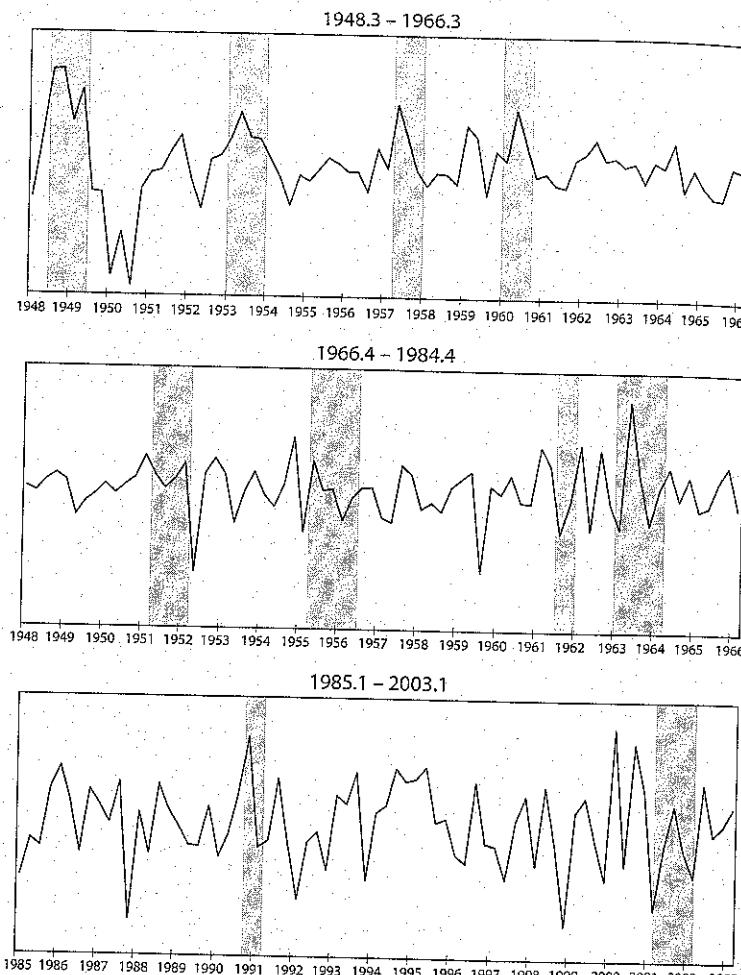


Figure 13.16 Smoothed Policy Shock

trigger probabilities for identifying the onset of a recession, we obtained successful predictions of all 10 recessions observed over the sample period, and obtained no false positive predictions. The demand and technology shocks are most important in contributing towards this performance. Using the former series in isolation, we obtained successful predictions for 7 of the 10 recessions with no false positives using the 0.75 trigger, and 9–10

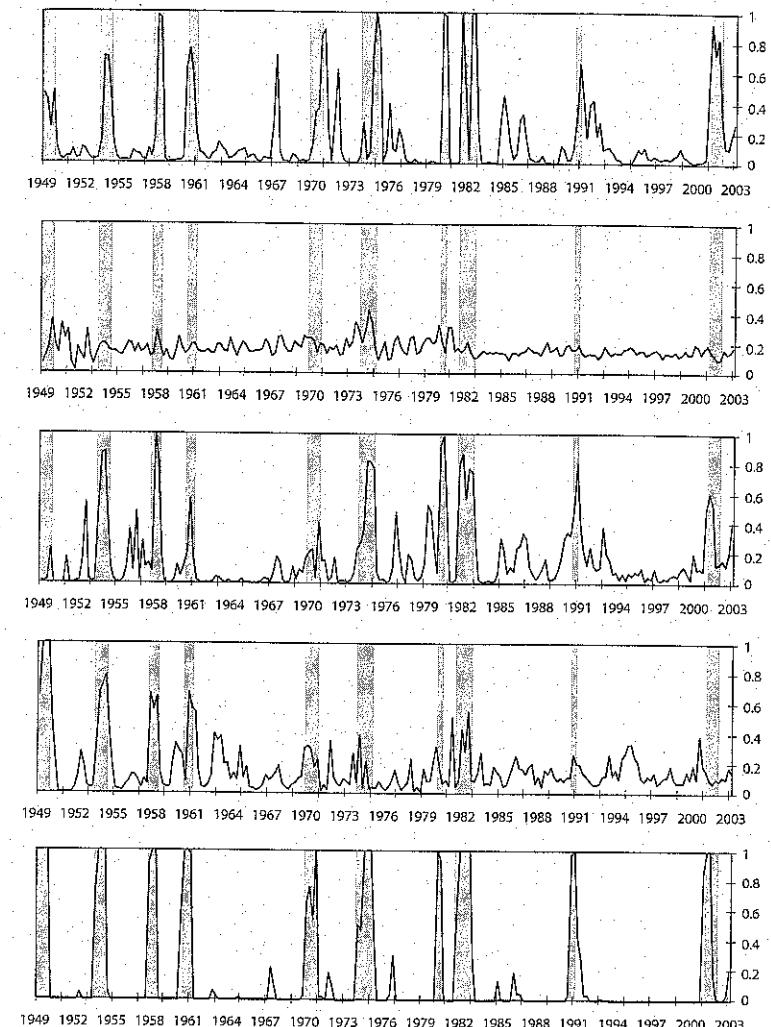


Figure 13.17 Fitted Probabilities of NBER-Dated Recessions. Explanatory variables from top to bottom: demand-push, cost-push, technology, all shocks.

successful predictions with two false positives using the 0.5 trigger. Comparable numbers for the latter series are (6, no false) and (8, two false). In contrast, only the first two recessions were successfully predicted using the policy-shock series and the 0.75 rule, whereas the first four recessions, along with the 1981 recession, were successfully predicted using the 0.5 trigger. This may indicate that policy responses became less distinct over the latter half of the sample. Finally, the cost-push shock never generated predicted probabilities in excess of 0.45.

## Chapter 14

---

### Bayesian Methods

#### 14.1 Overview of Objectives

Chapter 13 described the pursuit of empirical objectives using classical estimation methods. This chapter characterizes and demonstrates an alternative approach to this pursuit: the use of Bayesian methods.

A distinct advantage in using structural models to conduct empirical research is that a priori guidance concerning their parameterization is often much more readily available than is the case in working with reduced-form specifications. The adoption of a Bayesian statistical perspective in this context is therefore particularly attractive, because it facilitates the formal incorporation of prior information in a straightforward manner.

The reason is that from a Bayesian perspective, parameters are interpreted as random variables. In the estimation stage, the objective is to make conditional probabilistic statements regarding the parameterization of the model. Conditioning is made with respect to three factors: the structure of the model, the observed data, and a prior distribution specified for the parameters. The structure of the model and the observed data combine to form a likelihood function. Coupling the likelihood function with a prior distribution using Bayes' Rule yields an associated posterior distribution.

Note that under the Bayesian perspective, both the likelihood function and posterior distribution can be used to assess the relative plausibility of alternative parameterizations of the model. Use of the likelihood function for this purpose gives exclusive voice to the data; use of the posterior distribution gives voice to both the data and the researcher. The point of this observation is that the incorporation of prior information is not what distinguishes classical from Bayesian analysis; the distinguishing feature is the probabilistic interpretation assigned to parameters under the Bayesian perspective.

So it is not necessary to specify prior distributions to gain distinction as a Bayesian. But as noted, in the context of working with structural models, the ability to do so is often advantageous. Indeed, one interpretation of a calibration exercise is of a Bayesian analysis involving the specification of a point-mass prior over model parameters. The incorporation of prior uncertainty into such an analysis through the specification of a diffuse prior gives voice to the data regarding model parameterization, and in so doing enables the formal pursuit of a wide range of empirical objectives.

Broadly defined, Bayesian procedures have been applied to the analysis of DSGEs in pursuit of three distinct empirical objectives; all three are discussed and demonstrated in this chapter. First, they have been used to implement DSGEs as a source of prior information regarding the parameterization of reduced-form models. An important goal in this sort of exercise is to help provide theoretical context for interpreting forecasts generated by popular reduced-form models, most notably vector autoregressions (VARs). Second, they have been used to facilitate the direct estimation of DSGEs, and to implement estimated models in pursuit of a variety of empirical objectives. A leading objective has been the indirect measurement of unobservable facets of macroeconomic activity (e.g., time-series observations of productivity shocks, time spent outside the labor force on home-production and skill-acquisition activities, etc.). Third, Bayesian procedures have been used to facilitate model comparisons. As an alternative to the classical hypothesis-testing methodology, in the Bayesian context model comparison is facilitated via posterior odds analysis. Posterior odds convey relative probabilities assigned to competing models, calculated conditionally given priors, and data. They are straightforward to compute and interpret even in cases where all competing models are known to be false, and when alternatives are nonnested.

The analyses described in this chapter entail the use of computationally intensive numerical integration procedures. Many alternative procedures are available for facilitating their implementation; several were described in detail in chapter 9. While we take knowledge of these procedures as given here, we do briefly resketch certain concepts to keep this chapter relatively self-contained.

Beyond chapter 9, further details regarding numerical integration methods are available from Geweke (1999a, 2005) and Robert and Casella (1999). In addition, Koop (2003) provides a general overview of Bayesian statistical methods with an eye towards discussing computational issues. Finally, Bauwens, Lubrano, and Richard (1999) provide a computationally based overview of the application of Bayesian methods to the study of a broad range of dynamic reduced-form models. Beyond the treatment of computational issues, more general textbook overviews of Bayesian methods are provided by Zellner (1971), Leamer (1978), Poirier (1995), and Lancaster (2004).

## 14.2 Preliminaries

Likelihood functions provide the foundation for both classical and Bayesian approaches to statistical inference. To establish the notation used

### 14.2 Preliminaries

throughout this chapter, let us briefly reestablish their derivation. We begin with a given structural model  $A$  with parameters collected in the vector  $\mu$ . Approximation of the model yields a state-space representation with parameters collected in the vector  $\Lambda(\mu)$ . Coupled with a distributional assumption regarding the disturbances included in the model, the state-space representation yields an associated likelihood function, which can be evaluated using a filtering algorithm. Letting  $X$  denote the sample of observations on the observed variables in the system, the likelihood function is written as  $L(X|\mu, A)$ . Note that the mapping from  $\mu$  to  $\Lambda$  is taken for granted under this notation; this is because  $\mu$  typically serves as the primary focus of the analysis. When working with only a single model, the dependence of the likelihood function on  $A$  will also be taken for granted, and the function will be written as  $L(X|\mu)$ .

From a classical perspective, parameters are interpreted as fixed but unknown entities, and the likelihood function is interpreted as a sampling distribution for the data. The realization of  $X$  is thus interpreted as one of many possible realizations from  $L(X|\mu)$  that could have been obtained. Inferences regarding the specification of  $\mu$  center on statements regarding probabilities associated with the particular observation of  $X$  for given values of  $\mu$ .

From a Bayesian perspective, the observation of  $X$  is taken as given, and inferences regarding the specification of  $\mu$  center on statements regarding probabilities associated with alternative specifications of  $\mu$  conditional on  $X$ . This probabilistic interpretation of  $\mu$  gives rise to a potential avenue for the formal incorporation of a priori views regarding its specification. This is facilitated via the specification of a prior distribution for  $\mu$ , denoted as  $\pi(\mu)$ .

To motivate the incorporation of  $\pi(\mu)$  into the analysis, recall the definition of conditional probability, which holds that the joint probability of  $(X, \mu)$  can be calculated as

$$p(X, \mu) = L(X|\mu)\pi(\mu), \quad (14.1)$$

or reversing the roles of  $\mu$  and  $X$ ,

$$p(X, \mu) = P(\mu|X)p(X). \quad (14.2)$$

In (14.1) the likelihood function is used to perform the conditional probability calculation, and conditioning is made with respect to  $\mu$ ;  $\pi(\mu)$  in turn assigns probabilities to specific values of  $\mu$ . In (14.2),  $P(\mu|X)$  is used to perform the conditional probability calculation, and conditioning is made with respect to  $X$ ;  $p(X)$  in turn assigns probabilities to specific values

of  $X$ . Eliminating  $p(X, \mu)$  by equating (14.1) and (14.2) and solving for  $P(\mu|X)$  yields Bayes' Rule:

$$\begin{aligned} P(\mu|X) &= \frac{L(X|\mu)\pi(\mu)}{p(X)} \\ &\propto L(X|\mu)\pi(\mu), \end{aligned} \quad (14.3)$$

where  $p(X)$  is a constant from the point of view of the distribution for  $\mu$ .

In (14.3),  $P(\mu|X)$  is the posterior distribution. Conditional on  $X$  and the prior  $\pi(\mu)$ , it assigns probabilities to alternative values of  $\mu$ ; this distribution is the central focus of Bayesian analyses.

The typical objective of a Bayesian analysis involves the calculation of the conditional expected value of a function of the parameters  $g(\mu)$ :

$$E[g(\mu)] = \frac{\int g(\mu)P(\mu|X)d\mu}{\int P(\mu|X)d\mu}. \quad (14.4)$$

(The denominator is included in (14.4) to cover the general case in which  $p(X)$  in (14.3) is unknown, and thus  $P(\mu|X)$  does not integrate to 1.) This objective covers a broad range of cases, depending upon the specification of  $g(\mu)$ . For example, if  $g(\mu)$  is simply the identity function, then (14.4) delivers the posterior mean of  $\mu$ . Alternatively, if  $g(\mu)$  is an indicator over a small interval for  $\mu^j$ , the  $j^{\text{th}}$  element of  $\mu$  (e.g.,  $g(\mu) = 1$  for  $\mu^j \in [\underline{\mu}^j, \bar{\mu}^j]$ , and 0 otherwise), then assigning indicators over the support of each element of  $\mu$  would enable the construction of marginal predictive density functions for each structural parameter of the underlying model. (Referring to the predictive density function of  $\mu^j$ , the term *marginal* indicates that the function is unconditional on the values of the additional elements of  $\mu$ . Further elaboration for this important example is provided below in this section.) Marginal predictive density functions for additional functions such as spectra, impulse response functions, predictive densities, and time-series observations of unobservable variables included in the structural model can be constructed analogously. Regardless of the specification of  $g(\mu)$ , (14.4) has a standard interpretation:  $E[g(\mu)]$  is the weighted average of  $g(\mu)$ , with the weight assigned to a particular value of  $\mu$  determined jointly by the data (through the likelihood function) and the prior.

As noted, numerical integration procedures feature prominently in Bayesian analyses. This is due to the fact that in general, it is not possible to calculate  $E[g(\mu)]$  analytically. Instead, numerical methods are used to approximate the integrals that appear in (14.4).

Recall that in the simplest case, it is possible to simulate random drawings of  $\mu$  directly using the posterior distribution  $P(\mu|X)$ . In this

case, (14.4) can be approximated via direct Monte Carlo integration. Let  $\mu_i$  denote the  $i^{\text{th}}$  of a sequence of  $N$  draws obtained from  $P(\mu|X)$ . Then by the law of large numbers (e.g., as in Geweke, 1989):

$$\bar{g}_N = \left( \frac{1}{N} \right) \sum_{i=1}^N g(\mu_i) \rightarrow E[g(\mu)], \quad (14.5)$$

where  $\rightarrow$  indicates convergence in probability. The numerical standard error associated with  $\bar{g}_N$  is  $1/\sqrt{N}$  times the standard deviation of  $g(\mu)$ :

$$\text{s.e.}(\bar{g}_N) = \frac{\sigma(g(\mu))}{\sqrt{N}}. \quad (14.6)$$

Defining  $\bar{\sigma}_N(g(\mu))$  as the sample standard deviation of  $g(\mu)$ ,

$$\bar{\sigma}_N(g(\mu)) = \left[ \left( \frac{1}{N} \right) \sum_{i=1}^N g(\mu_i)^2 - \bar{g}_N^2 \right]^{1/2}, \quad (14.7)$$

$\text{s.e.}(\bar{g}_N)$  is estimated in practice using

$$\text{s.e.}(\bar{g}_N) = \frac{\bar{\sigma}_N(g(\mu))}{\sqrt{N}}. \quad (14.8)$$

A typical choice for  $N$  is 10,000, in which case  $1/\sqrt{N} = 1\%$ .

With these preliminaries established, we turn in section 3 to a discussion of an important class of empirical applications under which integrals of interest can be approximated via direct Monte Carlo integration. Under this class of applications, structural models are used as sources of prior information to be combined with likelihood functions associated with reduced-form models, typically in pursuit of forecasting applications. Judicious choices of prior distributions in this case can yield posterior distributions from which sample realizations of  $\mu$  can be simulated. However, in working with state-space representations directly, this is no longer the case. Thus section 4 returns to the problem of approximating  $E[g(\mu)]$ , and introduces more sophisticated numerical integration techniques.

### 14.3 Using Structural Models as Sources of Prior Information for Reduced-Form Analysis

When available, the choice of pursuing an empirical question using either a structural or reduced-form model typically involves a trade-off: reduced-form models are relatively easy to work with, but generate results relatively lacking in theoretical content. Here we discuss a methodology that strikes a

compromise in this trade-off. The methodology involves the use of structural models as sources of prior distributions for reduced-form models. This section presents a general characterization of a popular implementation of this methodology, and section 6 presents a forecasting application based on an autoregressive model.

Let  $\Theta$  denote the collection of parameters associated with the reduced-form model under investigation. The goal is to convert a prior distribution  $\pi(\mu)$  specified over the parameters of a corresponding structural model into a prior distribution for  $\Theta$ . If the mapping from  $\mu$  to  $\Theta$  was one-to-one and analytically tractable, the distribution over  $\Theta$  induced by  $\pi(\mu)$  could be constructed using a standard change-of-variables technique. But for cases of interest here, with  $\mu$  representing the parameters of a nonlinear expectational difference equation, analytically tractable mappings will rarely be available. Instead, numerical mappings are used.

Numerical mappings are accomplished using simulation techniques. A simple algorithm, presented in the context of a log-linear model representation, is as follows. Given a parameterization  $\mu$ , a corresponding parameterization  $\Lambda(\mu)$  of the associated state-space representation is obtained using the model-solution methods described in chapter 4. Next, a simulated sequence of structural shocks  $\{v_t\}$  is obtained from their corresponding distribution, and then fed through the parameterized state-space model to obtain a corresponding realization of model variables  $\{x_t\}$ . Next,  $\{x_t\}$  is fed into the observer equation  $X_t = H(\mu)'x_t$  to obtain a corresponding realization of observable variables  $\{X_t\}$ . Finally, the artificially obtained observables are used to estimate the reduced-form parameters  $\Theta$ . Repeating this process for a large number of drawings of  $\mu$  from  $\pi(\mu)$  yields a corresponding prior distribution for  $\Theta$ ,  $\pi(\Theta)$ .

As an important leading example, outlined by Zellner (1971) in full detail, consider the case in which the reduced-form model under investigation is of the form

$$Y = XB + U, \quad (14.9)$$

where  $Y$  is  $T \times n$ ,  $X$  is  $T \times k$ ,  $B$  is  $k \times n$ ,  $U$  is  $T \times n$ , and the rows of  $U$  are  $iidN(0, \Sigma)$ . (Note the slight abuse of notation: in the remainder of this section,  $(Y, X)$  will constitute the observable variables of the structural model.) The likelihood function for

$$\Theta \equiv (B, \Sigma) \quad (14.10)$$

is given by

$$L(Y, X | B, \Sigma) \propto |\Sigma|^{-T/2} \exp \left\{ -\frac{1}{2} \text{tr}[(Y - XB)'(Y - XB)\Sigma^{-1}] \right\}, \quad (14.11)$$

### 14.3 Using Structural Models as Sources of Prior Information

where  $\text{tr}[\cdot]$  denotes the trace operation. In this case, it is not only possible to induce a prior over  $\Theta$  using  $\pi(\mu)$ , but it is possible to do so in such a way that the resulting posterior distribution  $P(\Theta | Y, X)$  can be implemented directly as a sampling distribution in pursuit of calculations of  $E[g(\Theta)]$ .

To explain how this is accomplished, it is useful to note before proceeding that the likelihood function in (14.11) can itself be interpreted as a posterior distribution for  $\Theta$ . Specifically, combining (14.11) with a prior distribution that is proportional to  $|\Sigma|^{-(n+1)/2}$ , and thus uninformative over  $B$  (the so-called Jeffreys prior), the resulting posterior distribution for  $\Theta$  can be partitioned as  $P(B|\Sigma)P(\Sigma)$ , where the dependency on  $(Y, X)$  has been suppressed for ease of notation. The conditional distribution  $P(B|\Sigma)$  is multivariate Normal, and the marginal distribution  $P(\Sigma)$  is inverted-Wishart. Specifically, letting  $\widehat{B}$  denote the ordinary least squares (OLS) estimate of  $B$ ,

$$\widehat{H} = (Y - X\widehat{B})'(Y - X\widehat{B}),$$

and defining

$$\widehat{\beta} = \text{vec}(\widehat{B}),$$

so that  $\widehat{\beta}$  is  $(kn \times 1)$ , the distributions of  $P(B|\Sigma)$  and  $P(\Sigma)$  are given by

$$P(B|\Sigma) \sim N(\widehat{\beta}|\Sigma \otimes (X'X)^{-1}) \quad (14.12)$$

$$P(\Sigma) \sim IW(\widehat{H}, \tau, n), \quad (14.13)$$

where  $IW(\widehat{H}, \tau, n)$  denotes an inverted-Wishart distribution with  $n \times n$  parameter matrix  $\widehat{H}$  and  $\tau$  degrees of freedom, and  $\otimes$  denotes the Kronecker product. In this case,  $\tau = T - k + n + 1$ .

Continuing with the example, the attainment of sample drawings of  $\Theta \equiv (B, \Sigma)$  directly from the Normal/inverted-Wishart (N-IW) distribution given in (14.12) and (14.13) for the purpose of computing  $E[g(\Theta)]$  via Monte Carlo integration can be accomplished using the following algorithm. First, calculate  $\widehat{\beta}$  and

$$\begin{aligned} \widehat{S} &= \left( \frac{1}{T-k} \right) (Y - X\widehat{B})'(Y - X\widehat{B}) \\ &\equiv \left( \frac{1}{\tau - n - 1} \right) \widehat{H} \end{aligned}$$

via OLS. Next, construct  $\widehat{S}^{-1/2}$ , the  $n \times n$  lower triangular Cholesky decomposition of  $\widehat{S}^{-1}$ ; that is,

$$\widehat{S}^{-1} = (\widehat{S}^{-1/2})(\widehat{S}^{-1/2})'$$

Next, calculate the Cholesky decomposition of  $(X'X)^{-1}$ , denoted as  $(X'X)^{-1/2}$ . The triple

$$(\hat{\beta}, \hat{S}^{-1/2}, (X'X)^{-1/2})$$

constitute inputs into the algorithm.

To obtain a sample drawing of  $\Sigma$ ,  $\Sigma_i$ , draw  $\xi_i$ , an  $n \times (T - k)$  matrix of independent  $N(0, (T - k)^{-1})$  random variables, and construct  $\Sigma_i$  using

$$\Sigma_i = [(\hat{S}^{-1/2}\xi_i)(\hat{S}^{-1/2}\xi_i)' ]^{-1}. \quad (14.14)$$

To obtain a sample drawing of  $\beta$ ,  $\beta_i$ , calculate the Cholesky decomposition of the drawing  $\Sigma_i$ , denoted as  $\Sigma_i^{1/2}$ , and construct  $\beta_i$  using

$$\beta_i = \hat{\beta} + (\Sigma_i^{1/2} \otimes (X'X)^{-1/2}) w_i, \quad (14.15)$$

where  $w_i$  is a sample drawing of  $(nk) \times 1$  independent  $N(0, 1)$  random variables. The drawing

$$\beta_{i+1} = \hat{\beta} - (\Sigma_i^{1/2} \otimes (X'X)^{-1/2}) w_i$$

is an antithetic replication of this process;  $(\beta_i, \beta_{i+1})$  constitute an antithetic pair (see Geweke, 1988, for a discussion of the benefits associated with the use of antithetic pairs in this context).<sup>1</sup>

A similar distributional form for the posterior distribution will result from the combination of an informative N-IW prior specified over  $(B, \Sigma)$ , coupled with  $L(Y, X|B, \Sigma)$  in (14.11). In particular, letting the prior be given by

$$P(\beta|\Sigma) \sim N(\beta^*|\Sigma \otimes N^{*-1}) \quad (14.16)$$

$$P(\Sigma) \sim IW(H^*, \tau^*, n), \quad (14.17)$$

the corresponding posterior distribution is given by

$$P(\beta|\Sigma) \sim N(\beta^P|\Sigma \otimes (X'X + N^*)^{-1}) \quad (14.18)$$

$$P(\Sigma) \sim IW(H^P, \tau + \tau^*, n), \quad (14.19)$$

where

$$\beta^P = [\Sigma^{-1} \otimes (X'X + N^*)]^{-1} [(\Sigma^{-1} \otimes X'X)\hat{\beta} + (\Sigma^{-1} \otimes N^*)\beta^*] \quad (14.20)$$

$$H^P = \tau^* H^* + \hat{H} + (\hat{B} - B^*)' N^* (X'X + N^*)^{-1} (X'X) (\hat{B} - B^*). \quad (14.21)$$

<sup>1</sup>The GAUSS procedure niw.prc is available for use in generating sample drawings from Normal-Inverted Wishart distributions.

### 14.3 Using Structural Models as Sources of Prior Information

Note that the conditional posterior mean  $\beta^P$  is a weighted average of the OLS estimate  $\hat{\beta}$  and the prior mean  $\beta^*$ . Application of the algorithm outlined above to (14.18) and (14.19) in place of (14.12) and (14.13) enables the calculation of  $E[g(\Theta)]$  via direct Monte Carlo integration.

Return now to the use of  $\pi(\mu)$  as the source of a prior distribution to be combined with  $L(Y, X|B, \Sigma)$  in (14.11). A methodology that has been used frequently in pursuit of forecasting applications involves a restriction of the prior over  $(B, \Sigma)$  to be of the N-IW form, as in (14.16) and (14.17). (Examples include DeJong, Ingram, and Whiteman, 1993; Ingram and Whiteman, 1994; Sims and Zha, 1998; and Del Negro and Schorfheide, 2004.) Once again, this enables the straightforward calculation of  $E[g(\Theta)]$  via direct Monte Carlo integration, using (14.18) and (14.19). The only complication in this case involves the mapping of implications of  $\pi(\mu)$  for  $\beta^*$ ,  $N^*$ , and  $H^*$ .

A simple algorithm for performing this mapping is inspired by the problem of pooling sample information obtained from two sources. In this case, the two sources are the likelihood function and the prior. The algorithm is as follows. As described above, begin by obtaining a drawing  $\mu_i$  from  $\pi(\mu)$ , and simulate a realization of observable variables  $\{Y_i, X_i\}$  from the structural model. Using these simulated variables, estimate the reduced-form model (14.9) using OLS, and use the estimated values of  $\beta_i$  and  $H_i$  for  $\beta^*$  and  $\tau^* H^*$  in (14.16) and (14.17). Also, use the simulated value of  $(X'_i X_i)$  in place of  $N^*$ , and denoting the sample size of the artificial sample as  $T^*$ , note that  $\tau^* = T^* - k + n + 1$ .<sup>2</sup> Using

$$(\beta_i, H_i, (X'_i X_i), T^* - k + n + 1)$$

in place of

$$(\beta^*, \tau^* H^*, N^*, \tau^*)$$

in (14.18)–(14.21), construct the corresponding posterior distribution, and obtain simulated drawings of  $\Theta$  using the algorithm outlined above. Note that at this stage, the posterior distribution is conditional on  $\mu_i$ . As a suggestion, obtain 10 drawings of  $\Theta_i$  for each drawing of  $\mu_i$ . Finally, repeating this process by drawing repeatedly from  $\pi(\mu)$  enables the approximation of  $E[g(\Theta)]$ .

By varying the ratio of the artificial to actual sample sizes in simulating  $\{Y_i, X_i\}$ , the relative influence of the prior distribution can be adjusted.

<sup>2</sup>For cases in which the dimensionality of  $X$  is such that the repeated calculation of  $(X'X)^{-1}$  is impractical, shortcuts for this step are available. For example, DeJong, Ingram, and Whiteman (1993) use a fixed matrix  $C$  in place of  $(X'X)^{-1}$ , designed to assign harmonic decay to lagged values of coefficients in a VAR specification.

Equal sample sizes ( $T = T^*$ ) assign equal weight to the prior and likelihood functions; increasing the relative number of artificial drawings assigns increasing weight to the prior. Note (in addition to the strong resemblance to sample pooling) the strong resemblance of this algorithm to the mixed-estimation procedure of Theil and Goldberger (1961). For a discussion of this point, see Leamer (1978).

#### 14.4 Implementing Structural Models Directly

We return to the problem of calculating (14.4), focusing on the case in which the posterior distribution is unavailable as a direct source of sample drawings of  $\mu$ . In the applications of interest here, this will be true in general due to the complexity of likelihood functions associated with state-space representations, and because likelihood functions are specified in terms of the parameters  $\Lambda(\mu)$ , whereas priors are specified in terms of the parameters  $\mu$ . However, it remains the case that we can calculate the likelihood function and posterior distribution for a given value of  $\mu$ , with aid from a filtering algorithm; these calculations are critical components of any approximation procedure.

Given unavailability of the posterior as a sampling distribution, we are faced with the problem of choosing a stand-in density for obtaining sample drawings  $\mu_i$ . Recall that effective samplers have two characteristics, accuracy and efficiency; they deliver close approximations of  $E[g(\mu)]$ , and do so using a minimal number of draws. The benchmark for comparison is the posterior itself; when available, it represents the pinnacle of accuracy and efficiency.

One option as a stand-in density is the importance sampling density. Specializing the notation of chapter 9 to the specific context of posterior analysis, we denote this density as  $I(\mu; \alpha)$ , where  $\alpha$  denotes its parameterization. Introducing  $I(\mu; \alpha)$  into (14.4) as

$$E[g(\mu)] = \frac{\int g(\mu) \frac{P(\mu|X)}{I(\mu; \alpha)} I(\mu; \alpha) d\mu}{\int \frac{P(\mu|X)}{I(\mu; \alpha)} I(\mu; \alpha) d\mu},$$

$E[g(\mu)]$  can be approximated via Monte Carlo integration as

$$\bar{g}_N = \frac{\sum_{i=1}^N g(\mu_i) w(\mu_i)}{\sum_{i=1}^N w(\mu_i)} \rightarrow E[g(\mu)], \quad (14.22)$$

$$w(\mu_i) = \frac{P(\mu_i|X)}{I(\mu_i; \alpha)}. \quad (14.23)$$

#### 14.5 Model Comparison

The numerical standard error of the approximation is

$$s.e.(\bar{g}_N)_I = \sqrt{\frac{\sum_{i=1}^N (g(\mu_i) - \bar{g}_N)^2 w(\mu_i)}{\left[\sum_{i=1}^N w(\mu_i)\right]^2}} \quad (14.24)$$

Another option as a stand-in density is a Metropolis-Hastings MCMC sampler. We denote such a sampler as  $i(\mu; \mu_{i-1}, H)$ , where  $\mu_{i-1}$  denotes the  $(i-1)^{\text{st}}$  draw of  $\mu$ , and  $(\mu_{i-1}, H)$  characterize the centering and spread/rotation of  $i()$ , respectively. Let  $\mu_i^*$  denote a drawing obtained from  $i(\mu; \mu_{i-1}, H)$  that serves as a candidate to become the next successful drawing  $\mu_i$ . Then under the Metropolis-Hastings Algorithm, this will be the case according to the probability

$$q(\mu_i^* | \mu_{i-1}) = \min \left[ 1, \frac{P(\mu_i^* | X) i(\mu_{i-1}; \mu_i^*, H)}{P(\mu_{i-1} | X) i(\mu_{i-1}^*; \mu_{i-1}, H)} \right]. \quad (14.25)$$

Letting  $\{\mu_i\}$  denote the sequence of accepted drawings,  $E[g(\mu)]$  is approximated via

$$\bar{g}_N = \frac{1}{N} \sum_1^N g(\mu_i);$$

its associated numerical standard error is given by

$$s.e.(\bar{g}_N)_G = \sqrt{\frac{1}{N} \left( \bar{\gamma}_0 + 2 \sum_{l=1}^{N-1} \bar{\gamma}_l \frac{N-l}{N} \right)}, \quad (14.26)$$

where  $\bar{\gamma}_l$  denotes the numerical estimate of the  $l^{\text{th}}$ -order autocovariance of  $g(\mu)$  obtained from the simulated  $\mu_i$ 's.

#### 14.5 Model Comparison

Just as posterior distributions can be used to assess conditional probabilities of alternative values of  $\mu$  for a given model, they can also be used to assess conditional probabilities associated with alternative model specifications. The assessment of conditional probabilities calculated over a set of models lies at the heart of the Bayesian approach to model comparison. (For examples of applications to DSGE models, see Geweke, 1999b; Schorfheide, 2000; Fernandez-Villaverde and Rubio-Ramirez, 2004a; and DeJong and Ripoll, 2004.)

Relative conditional probabilities are calculated using a tool known as a posterior odds ratio. To derive this ratio for the comparison of two models  $A$  and  $B$ , we begin by rewriting the posterior distribution as represented in (14.3). We do so to make explicit that the probability assigned to a given value of  $\mu$  is conditional not only on  $X$ , but also on the specified model. For model  $A$ , the posterior is given by

$$P(\mu_A|X, A) = \frac{L(X|\mu_A, A)\pi(\mu_A|A)}{p(X|A)}, \quad (14.27)$$

where we use  $\mu_A$  to acknowledge the potential for  $\mu$  to be specific to a particular model specification. Integrating both sides with respect to  $\mu_A$ , and recognizing that the left-hand side integrates to 1, we have

$$p(X|A) = \int L(X|\mu_A, A)\pi(\mu_A|A)d\mu_A. \quad (14.28)$$

This is the marginal likelihood associated with model  $A$ : it is the means by which the quality of the model's characterization of the data is measured. The measure for model  $B$  is analogous.

Just as Bayes' Rule can be used to calculate the conditional probability associated with  $\mu_A$ , it can also be used to calculate the conditional probability associated with model  $A$ :

$$p(A|X) = \frac{p(X|A)\pi(A)}{p(X)}, \quad (14.29)$$

where  $\pi(A)$  indicates the prior probability assigned to model  $A$ . Substituting (14.28) into (14.29) yields

$$p(A|X) = \frac{[\int L(X|\mu_A, A)\pi(\mu_A|A)d\mu_A]\pi(A)}{p(X)}. \quad (14.30)$$

Finally, taking the ratio of (14.30) for models  $A$  and  $B$  (a step necessitated by the general inability to calculate  $p(X)$ ) yields the posterior odds ratio

$$PO_{A,B} = \frac{[\int L(X|\mu_A, A)\pi(\mu_A|A)d\mu_A]\pi(A)}{[\int L(X|\mu_B, B)\pi(\mu_B|B)d\mu_B]\pi(B)}. \quad (14.31)$$

The ratio of marginal likelihoods is referred to as the Bayes factor, and the ratio of priors as the prior odds ratio. Extension to the analysis of a larger collection of models is straightforward.

Approaches to model comparison based on posterior-odds ratios have several attractive features. First, all competing models are treated symmetrically; there is no null model being compared to an alternative, just

a collection of models for which relative conditional probabilities are computed. Relatedly, no difficulties in interpretation arise for the case in which all models are known to be approximations of reality (i.e., false). The question is simply: which among the collection of models has the highest conditional probability? Finally, the models need not be nested to facilitate implementation.

Regarding implementation, this can be achieved using importance-sampling or MCMC methods. The only complication relative to the calculation of  $E[g(\mu)]$  is that in this case, proper densities must be used in calculating  $w(\mu)$ . Given proper densities,

$$\int L(X|\mu_A, A)\pi(\mu_A|A)d\mu_A$$

can be approximated via importance sampling using

$$\bar{w} = \sum_{i=1}^N w(\mu_i).$$

Using MCMC in the independence chain Metropolis-Hastings Algorithm, with  $i(\mu; \mu_{i-1}, H)$  serving as the proposal density, and  $\mu_i^*$  representing the proposal drawn on the  $i^{\text{th}}$  iteration, define

$$w(\mu_i^*) = \frac{L(X|\mu_i^*)\pi(\mu_i^*)}{i(\mu; \mu_{i-1}, H)}.$$

In this case,  $\int L(X|\mu_A, A)\pi(\mu_A|A)d\mu_A$  can be approximated using

$$\bar{w} = \sum_{i=1}^N w(\mu_i, \mu_i^*).$$

Geweke (1999a) provides details for both approximations.

In addition to these traditional means of calculating marginal likelihoods, Geweke (1999b) proposes a posterior-odds measure similar in spirit to a methods-of-moments analysis. The construction of the measure begins with the specification of a reduced-form model for the observable variables under investigation. Using this specification and an uninformative prior, the corresponding posterior distribution over a collection of moments  $m$  is estimated. Finally, the overlap between this distribution and distributions associated with competing structural models is measured, and converted into a posterior-odds measure. The construction of all posteriors is straightforward, involving the calculation of  $E[g(\mu)]$  as described above, Schorfheide (2000) proposes a related procedure under which a loss function is specified over  $m$ , enabling the researcher to adjust the relative weights assigned to specific elements of  $m$  in judging overlap.

Let  $P_A(m)$  denote the posterior distribution of  $m$  associated with model  $A$ , and  $V(m)$  the posterior distribution of  $m$  associated with the reduced-form model. The posterior odds in favor of model  $A$  relative to  $B$  are defined in this case as

$$PO_{A,B} = \frac{[\int P_A(m)V(m)dm]\pi(A)}{[\int P_B(m)V(m)dm]\pi(B)}; \quad (14.32)$$

thus  $PO_{A,B}$  provides a characterization of the relative overlap between the reduced-form density over  $m$  and the posterior densities associated with models  $A$  and  $B$ . The greater the overlap exhibited by model  $A$  relative to model  $B$ , the greater will be  $PO_{A,B}$ .

Evaluation of the integrals required to calculate  $PO_{A,B}$  can be facilitated by kernel density approximation. Let  $m_A(i)$  denote the  $i^{\text{th}}$  of  $M$  drawings of  $m$  obtained from  $P_A(m)$  (or given the use of an importance-sampling approximation, a suitably weighted drawing from the importance density associated with  $P_A(m)$ ), and  $m_V(j)$  denote the  $j^{\text{th}}$  of  $N$  drawings obtained from  $V(m)$ . Then the numerator of  $PO_{A,B}$  can be approximated using

$$\frac{1}{MN} \sum_{i=1}^M \sum_{j=1}^N K(m_A(i), m_V(j)), \quad (14.33)$$

and likewise for the denominator.  $K(m_A(i), m_V(j))$  denotes a density kernel used to judge the distance between  $m_A(i)$  and  $m_V(j)$ . For details on kernel density approximation, see, for example, Silverman (1986).

Guidance for interpreting the “weight of evidence” conveyed by the data for model  $A$  against model  $B$  through a posterior odds calculation is provided by Jeffreys (1961). Odds ranging from 1:1–3:1 constitute “very slight evidence” in favor of  $A$ ; odds ranging from 3:1–10:1 constitute “slight evidence” in favor of  $A$ ; odds ranging from 10:1–100:1 constitute “strong to very strong evidence” in favor of  $A$ ; and odds exceeding 100:1 constitute “decisive evidence” in favor of  $A$ .

## 14.6 Using an RBC Model as a Source of Prior Information for Forecasting

To demonstrate the use of structural models as a source of prior information for reduced-form analysis, here we pursue a forecasting application. The goal is to forecast output using an autoregressive specification featuring a unit root with drift. Versions of this workhorse specification provide a common benchmark against which alternative forecasting models are often

compared, thus it serves as a useful example for this demonstration (e.g., see Pesaran and Potter, 1997; and DeJong, Liesenfeld, and Richard, 2006).

One motivation for introducing a structural model into such an analysis is that autoregressive models, along with their multivariate vector-autoregressive counterparts, tend to produce large forecast errors. Following Doan, Litterman, and Sims (1984), a leading response to this problem involves the use of a prior distribution as a gentle means of imposing parameter restrictions on the forecasting model. The restrictions serve to enhance forecasting precision. Working in the context of a VAR model, the so-called “Minnesota prior” used by Doan et al. is centered on a parameterization under which the data are modeled as independent unit-root processes. As noted in section 3, the extensions to this work pursued by DeJong, Ingram, and Whiteman (1993); Ingram and Whiteman (1994); Sims and Zha (1998); and Del Negro and Schorfheide (2004) have enabled the use of structural models as an alternative source of prior information that have also proven effective in enhancing forecasting precision. Here we use the RBC model introduced in section 1 of chapter 3 as a source of prior information.

As a measure of output we consider the post-war quarterly series introduced in chapter 6, spanning 2004:IV. We use the data spanning 1947:I–1997:IV for estimation, and compare the remaining observations through 2004:IV with 1- through 28-step-ahead forecasts. Letting  $\Delta y_t$  denote the first difference of logged output, the model is given by

$$\Delta y_t = \underline{\rho}_0 + \underline{\rho}_1 \Delta y_{t-1} + \underline{\rho}_2 \Delta y_{t-2} + \varepsilon_t, \quad \varepsilon_t \sim iidN(0, \sigma^2); \quad (14.34)$$

that is,  $\Delta y_t$  follows an AR(2) specification. We chose this particular specification following DeJong and Whiteman (1994a), who examined its forecasting performance in application to the 14 macroeconomic time series originally studied by Nelson and Plosser (1982), and extended by Schotman and van Dijk (1991). As DeJong and Whiteman illustrated, although point-forecasts (e.g., means of predictive densities) generated by this model tend to be accurate, they also tend to be imprecise (e.g., as indicated by coverage intervals of predictive densities, described below).

Three alternative representations of (14.34) will prove to be convenient for the purposes of estimation and forecasting. Adding  $y_{t-1}$  to both sides of (14.34), we obtain an expression for the level of  $y_t$  given by

$$\begin{aligned} y_t &= \underline{\rho}_0 + (1 + \underline{\rho}_1)y_{t-1} + (-\underline{\rho}_1 + \underline{\rho}_2)y_{t-2} - \underline{\rho}_2 y_{t-3} + \varepsilon_t \\ &\equiv \rho_0 + \rho_1 y_{t-1} + \rho_2 y_{t-2} + \rho_3 y_{t-3} + \varepsilon_t. \end{aligned} \quad (14.35)$$

Thus we see that the AR(2) specification for  $\Delta y_t$  implies a restricted AR(3) specification for  $y_t$ . The restrictions

$$\rho_1 = (1 + \underline{\rho}_1), \rho_2 = (-\underline{\rho}_1 + \underline{\rho}_2), \rho_3 = -\underline{\rho}_2 \quad (14.36)$$

imply that the polynomial

$$\rho(L) = (1 - \rho_1 L - \cdots - \rho_3 L)$$

in the lag operator  $L$  has a unit root, and thus  $\rho_0$  captures the growth rate of  $y_t$  (see chapter 7 for a discussion).

For the purpose of forecasting it is convenient to express (14.35) in companion form:

$$\begin{bmatrix} y_t \\ y_{t-1} \\ y_{t-2} \\ 1 \end{bmatrix} = \begin{bmatrix} \rho_1 & \rho_2 & \rho_3 & \rho_0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{t-1} \\ y_{t-2} \\ y_{t-3} \\ 1 \end{bmatrix} + \begin{bmatrix} \varepsilon_t \\ 0 \\ 0 \\ 0 \end{bmatrix},$$

or more compactly,

$$z_t = Az_{t-1} + u_t,$$

where

$$z_t = [y_t \ y_{t-1} \ y_{t-2} \ 1]',$$

and

$$u_t = [\varepsilon_t \ 0 \ 0 \ 0].$$

Then letting  $e$  be a  $5 \times 1$  row vector with a 1 in the first row and zeros elsewhere, so that  $e' z_t = y_t$ ,  $y_{T+J}$  is given by

$$y_{T+J} = e' A^J z_T + e' \sum_{j=0}^{J-1} A^j u_{T+J-j}. \quad (14.37)$$

Finally, for the purpose of estimation, it is convenient to express the AR(2) model (14.34) for  $\Delta y_t$  in terms of standard regression notation:

$$y = X\underline{\beta} + v, \quad (14.38)$$

where

$$y = [\Delta y_2 \ \Delta y_3 \ \dots \ \Delta y_T]',$$

$$\underline{\beta} = [\underline{\rho}_0 \ \underline{\rho}_1 \ \underline{\rho}_2],$$

and so on. This is a special case of the reduced-form specification (14.9) involving a single variable rather than a collection of  $n$  variables. As such,

its associated posterior distribution is a slight modification of the Normal/inverted-Wishart distributions for  $(B, \Sigma)$  associated with (14.9), which are given in (14.12) and (14.13). Specifically, the conditional posterior distribution for  $\underline{\beta}$  is once again Normal:

$$P(\underline{\beta} | \sigma^2) \sim N(\widehat{\underline{\beta}}, \sigma^2(X'X)^{-1}), \quad (14.39)$$

where  $\widehat{\underline{\beta}}$  is the OLS estimate of  $\underline{\beta}$ . Then defining

$$\widehat{b} = (y - X\underline{\beta})'(y - X\underline{\beta}),$$

the distribution of  $\sigma^2$  is now inverted-Gamma rather than inverted-Wishart:

$$P(\sigma^2) \sim IG(\widehat{b}, \tau), \quad (14.40)$$

where  $\tau$  once again represents degrees of freedom:  $\tau = T - k$ .

The algorithm described in section 3 for obtaining artificial drawings of  $\Sigma$  from an inverted-Wishart distribution can also be used to obtain artificial drawings of  $\sigma^2$  from an inverted-Gamma distribution. The lone difference is that in this case

$$\widehat{S} = \left( \frac{1}{T - k} \right) (y - X\underline{\beta})'(y - X\underline{\beta})$$

is a single value rather than a matrix. Thus given OLS estimates  $\widehat{\underline{\beta}}$  and  $\widehat{b}$ , calculations of  $E(g(\underline{\beta}))$  can be obtained via the direct-sampling method described in sections 2 and 3: obtain a drawing of  $\sigma_i^2$  from  $IG(\widehat{b}, \tau)$ , obtain a drawing of  $\underline{\beta}_i$  from  $P(\underline{\beta} | \sigma^2)$ , calculate  $g(\underline{\beta}_i)$ , and approximate  $E(g(\underline{\beta}))$  by computing the sample average  $\bar{g}_N$  over  $N$  replications.

The object of interest in this application is a predictive density. To ease notation, let

$$\Theta = [\underline{\beta}' \ \sigma^2]$$

collect the full set of parameters associated with (14.34), so that

$$P(\Theta) \propto P(\underline{\beta} | \sigma^2)P(\sigma^2).$$

Also, let

$$\tilde{y} = [y_{T+1} \ y_{T+2} \ \dots \ y_{T+J}]'$$

denote a  $J$ -vector of future observations of  $y_t$ . By Bayes' Rule, the joint posterior distribution of  $\tilde{y}$  and  $\Theta$  is given by

$$P(\tilde{y}, \Theta) = P(\tilde{y} | \Theta)P(\Theta), \quad (14.41)$$

and the predictive density of  $\tilde{y}$  is obtained by integrating (14.41) with respect to  $\Theta$ :

$$P(\tilde{y}) = \int_{\Theta} P(\tilde{y}|\Theta)P(\Theta)d\Theta. \quad (14.42)$$

Note that by integrating over  $\Theta$ , the resulting predictive density is unconditional on a given parameterization of the model; the only conditioning is with respect to the data.

In the present application, the algorithm for approximating (14.42) works as follows. First, (14.34) is estimated via OLS, yielding  $\hat{\beta}$  and  $\hat{h}$ . Along with  $\tau = T - k$ , these parameterize the Normal/inverted-Gamma distributions over  $\sigma^2$  and  $\beta$  defined in (14.39) and (14.40), which are used to generate artificial drawings  $\sigma_i^2$  and  $\beta_i$  as described above. For a given drawing, a sequence of disturbances  $[u_{iT+1} \ u_{iT+2} \dots \ u_{iT+j}]$  is obtained from a  $N(0, \sigma_i^2)$  distribution, and  $\beta_i$  is mapped into  $[\rho_{i0} \ \rho_{i1} \ \rho_{i2} \ \rho_{i3}]$  as indicated in (14.36). Finally, these are fed into (14.37) to obtain a sample drawing of  $\tilde{y}$ . Histograms compiled over 10,000 replications of this process for each element of  $\tilde{y}$  approximate the predictive densities we seek. Below, means of these densities provide point forecasts, and associated measures of precision are provided by upper and lower boundaries of 95% coverage intervals.

Before presenting forecasts obtained in this manner, we describe the method employed for introducing prior information into the analysis, using the RBC model as a source. Recall briefly the structural parameters included in the model: capital's share of output  $\alpha$ ; the subjective discount factor  $\beta = \frac{1}{1+\varrho}$ , where  $\varrho$  is the subjective discount rate; the degree of relative risk aversion  $\phi$ ; consumption's share (relative to leisure) of instantaneous utility  $\varphi$ ; the depreciation rate of physical capital  $\delta$ ; the AR parameter specified for the productivity shock  $\rho$ ; and the standard deviation of innovations to the productivity shock  $\sigma$ .

Collecting these structural parameters into the vector  $\mu$ , the methodology begins with the specification of a prior distribution  $\pi(\mu)$ . The objective of the methodology is to map the prior over  $\mu$  into a prior over the parameters  $\Theta$  of the reduced-from model. For a given drawing  $\mu_i$  from  $\pi(\mu)$ , this is accomplished as follows. First the model is solved, yielding a parameterized specification for the model variables  $x_t$ :

$$x_t = F_i x_{t-1} + e_t, \quad e_t \sim N(0, Q_i). \quad (14.43)$$

Next, a simulated realization  $\{e_{it}\}_{t=1}^{T^*}$  is obtained from  $N(0, Q_i)$ , which along with  $x_0 = 0$  is fed into (14.43) to produce a simulated realization of  $\{x_{it}\}_{t=1}^{T^*}$ . The simulated realization of output  $\{y_{it}\}_{t=1}^{T^*}$  we seek is obtained

by applying an appropriately defined observation matrix  $H$  to the realized sequence of  $x_{it}$ :  $y_{it} = H' x_{it}$ .

Given  $\{y_{it}\}_{t=1}^{T^*}$ , the next step is to use these observations to estimate the AR(2) model (14.34) via OLS. Casting the model in the standard OLS form (14.38), as output from this step the following terms are collected:

$$\begin{aligned} \hat{\beta}_i &= (X_i' X_i)^{-1} X_i' y_i \\ \hat{h}_i &= (y - X \hat{\beta})'(y - X \hat{\beta}). \end{aligned}$$

Coupled with  $\tau^* = T^* - k$ , these terms serve to parameterize the Normal/inverted-Gamma prior distributions associated with  $\mu_i$ :

$$P(\beta|\sigma^2) = N(\hat{\beta}_i|\sigma^2(X_i' X_i)^{-1}), \quad (14.44)$$

$$P(\sigma^2) \sim IG(\hat{h}_i, \tau^*). \quad (14.45)$$

Combining these distributions with the Normal/inverted-Gamma distributions (14.39) and (14.40) associated with the actual data, we obtain the modified distributions

$$P(\beta|\sigma^2) \sim N(\beta^P|\sigma^2(X' X + X_i' X_i)^{-1}) \quad (14.46)$$

$$P(\sigma^2) \sim IG(h^P, \tau + \tau^*), \quad (14.47)$$

where

$$\beta^P = [\sigma^{-2}(X' X + X_i' X_i)]^{-1}[(\sigma^{-2} X' X)\hat{\beta} + (\sigma^{-2} X_i' X_i)\hat{\beta}_i] \quad (14.48)$$

$$H^P = \hat{h}_i + \hat{H} + (\hat{\beta} - \hat{\beta}_i)'(X_i' X_i)(X' X + X_i' X_i)^{-1}(X' X)(\hat{\beta} - \hat{\beta}_i). \quad (14.49)$$

Predictive densities can be constructed numerically using these modified distributions precisely as described above.

Recall that (14.46)–(14.49) are constructed from a single realization of structural parameters  $\mu_i$  from the prior distribution  $\pi(\mu)$ . In the results presented below, for each of 10,000 realizations of  $\mu_i$  we obtained, we generated 10 realizations of  $\Theta$ ; from (14.46) to (14.49), each of which was used to construct a realization of  $\tilde{y}$ .

Regarding the specification of  $\pi(\mu)$ , this was patterned closely after the calibration exercise conducted in section 4 of chapter 11. Its foundation is a sequence of independent Normal distributions specified over  $\beta$ ,  $\delta$ , and  $\phi$ ; the respective means (standard deviations) used to produce the results reported below are 0.99 (0.005), 0.025 (0.01), 1.5 (1). The distribution over  $\beta$  was truncated from above at 0.995 and from below at 0.98, and the distributions over  $\delta$  and  $\phi$  were truncated from below at 0 (the imposition of truncation is described below).

Given a drawing of  $(\beta, \delta, \phi)$ , the remaining parameters were constructed using the following sequence of steps. Using  $\beta = 1/\varrho - 1$  and  $\delta, \alpha$  was constructed as described in chapter 11 as the value necessary for matching the steady state investment/output ratio with its empirical counterpart:

$$\begin{aligned}\alpha &= \left( \frac{\delta + \varrho}{\delta} \right) \bar{y} \\ &= \left( \frac{\delta + \varrho}{\delta} \right) 0.175.\end{aligned}\quad (14.50)$$

In addition,  $\alpha$  was restricted to lie between 0.15 and 0.4. Likewise,  $\varphi$  was constructed as the value necessary for matching the steady state consumption/output ratio with its empirical counterpart:

$$\begin{aligned}\varphi &= \frac{1}{1 + \frac{2(1-\alpha)}{\bar{c}/\bar{y}}} \\ &= \frac{1}{1 + \frac{2(1-\alpha)}{0.825}}.\end{aligned}\quad (14.51)$$

Next,  $\delta$  was used to construct a realization of the implied trajectory of the physical capital stock  $\{k_t\}_{t=1}^T$  using the perpetual inventory method. Along with the realization of  $\alpha$ , this in turn was used to generate a realization of the Solow residual  $\{\log z_t\}_{t=1}^T$ . The growth rate of  $\log z_t$ , the parameter  $g_y$ , was then calculated and used to construct the growth rate of the artificial realization of output:  $g_y = g/(1 - \alpha)$ . To maintain consistency with the unit-root specification for output, the AR parameter  $\rho$  was fixed at 1. Finally,  $\sigma$  was constructed as the standard deviation of  $\Delta \log z_t$ .

To summarize, a drawing  $\mu_i$  from  $\pi(\mu)$  was obtained by drawing  $(\beta, \delta, \phi)$  from their respective Normal densities, and then constructing  $(\alpha, \varphi, g_y, \sigma)$  as described above. The truncation restrictions noted above were imposed by discarding any  $\mu_i$  that included an element that fell outside of its restricted range. Approximately 15% of the 100,000 total drawings of  $\mu_i$  we obtained were discarded in this application.

Given a successful drawing of  $\mu_i$ ,  $\{y_{it}\}_{t=1}^T$  was simulated from (14.43) as described above. Recall that this series is measured in terms of logged deviations of the level of output from steady state. Thus a final adjustment is necessary to convert  $y_{it}$  to a measure of the corresponding level of output. The steady state trajectory of output is given by

$$\bar{y}_t = \bar{y}_0(1 + g_y)^t,\quad (14.52)$$

where  $\bar{y}_0$  represents an arbitrary initial steady state value (subsequent first-differencing eliminates it from the analysis). Let the level of output corresponding with  $y_{it}$  be given by  $Y_{it}$ . Then

$$y_{it} = \log \left( \frac{Y_{it}}{\bar{y}_0(1 + g_y)^t} \right),\quad (14.53)$$

and thus the theoretical characterization of the level of output is given by

$$Y_{it} = \bar{y}_0(1 + g_y)^t e^{y_{it}}.\quad (14.54)$$

It is the first difference of the log of  $Y_{it}$  that is used to estimate (14.34) in the numerical integration algorithm described above.

The results of this exercise are presented in table 14.1 and figure 14.1. The table presents posterior means and standard deviations of  $\Theta$ ; the figure illustrates the actual data, point forecasts, and 95% coverage intervals. Five sets of estimates are presented: those obtained in the absence of theoretical restrictions imposed by  $\pi(\mu)$ , and four sets obtained using alternative specifications of  $\tau^*/\tau$ : 1, 2, 5, 10. (Point forecasts obtained given the restrictions imposed by  $\pi(\mu)$  are virtually indistinguishable from those obtained in the absence of  $\pi(\mu)$ , thus only the latter are illustrated in the figure.) Recall that  $\tau^*/\tau$  measures the ratio of artificial to actual degrees of freedom used in the construction of  $P(\beta|\sigma^2)$  and  $P(\sigma^2)$  in (14.46) and (14.47). The greater is this ratio, the greater is the influence of  $\pi(\mu)$  in the analysis. (Results obtained using  $\tau^*/\tau = 100$  closely resemble those obtained using  $\tau^*/\tau = 10$ ; thus this case appears to approximate a limiting scenario.)

TABLE 14.1  
Posterior Estimates of the AR(2) Model

$\tau^*/\tau$	$\rho_0$	$\rho_1$	$\rho_2$	$\sigma$
<b>Means</b>				
0	0.0030	0.384	0.041	0.0096
1	0.0036	0.242	0.073	0.0095
2	0.0038	0.191	0.076	0.0081
5	0.0042	0.139	0.073	0.0072
10	0.0043	0.113	0.071	0.0069
<b>Std. dev.</b>				
0	0.0008	0.072	0.071	0.0005
1	0.0007	0.063	0.061	0.0006
2	0.0006	0.050	0.049	0.0005
5	0.0004	0.037	0.037	0.0005
10	0.0003	0.030	0.030	0.0006

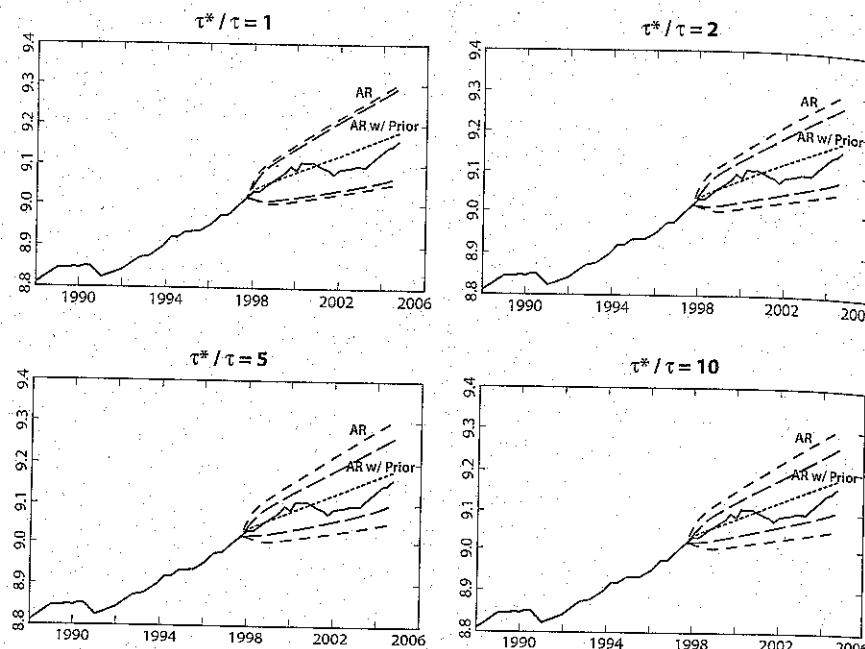


Figure 14.1 Predictive Densities

Three aspects of the changes in parameter estimates obtained in moving from  $\tau^*/\tau = 0$  to  $\tau^*/\tau = 10$  are notable. First, posterior means of  $\rho_1$  are reduced by a factor of 3, implying that fluctuations in  $\Delta y_t$  become increasingly less persistent as the influence of the RBC model is increased. Second, the posterior standard deviations associated with  $(\rho_0, \rho_1, \rho_2)$  are reduced by a factor of 2. Third, posterior means of  $\sigma$  fall from 0.0096 to 0.0069, approaching the standard deviation of  $\log z_t$  of 0.0067 obtained under the benchmark parameterization of the model in the calibration exercise conducted in chapter 11 (summarized in table 11.4). All three factors help account for the increased precision of forecasts that results from increases in  $\tau^*/\tau$ , as illustrated in figure 14.1.

To quantify this shrinkage, table 14.2 reports the percentage difference in the levels of output that define upper- and lower-95% coverage intervals at 1-, 4-, 8-, and 12-quarter forecast horizons for all five sets of forecasts. Shrinkage is noticeable even at the 1-quarter horizon, falling from 4% to 3% in moving from  $\tau^*/\tau = 0$  to  $\tau^*/\tau = 10$ . Respective shrinkages at the

TABLE 14.2  
Coverage-Interval Spreads (%)

$\tau^*/\tau$	Horizon:			
	1	4	8	12
0	4.00	9.75	13.00	15.75
1	4.00	8.75	12.00	14.25
2	3.50	7.50	10.30	12.00
5	3.25	6.50	8.75	10.50
10	3.00	6.00	8.30	10.00

4-, 8-, and 12-quarter horizons are from 9.75% to 6%, 13% to 8.3%, and 15.75% to 10%.

### Exercise 14.1

Explore the sensitivity of the results of this exercise to changes in  $\pi(\mu)$ . Are there particular elements of  $\mu$  that are particularly influential? Why might this be so?

## 14.7 Estimating and Comparing Asset-Pricing Models

To demonstrate the use of Bayesian methods for model estimation and comparison, we conclude this chapter by analyzing two versions of Lucas' (1978) one-tree model of asset-pricing behavior, introduced in section 3 of chapter 3. Both versions feature a representative household and a single asset. The price of the asset reflects the future dividend stream it is expected to generate, weighted by the household's marginal rate of substitution. Consumption is financed by dividend payments and an exogenous endowment; innovations to these stochastic processes drive stock prices.

The analysis centers on a puzzle first noted by LeRoy and Porter (1981) and Shiller (1981). In aggregate data on stock prices and corresponding dividends, prices and dividends are closely correlated, but prices are far more volatile. The excess volatility of stock prices over dividends is difficult to reconcile with standard models of asset-pricing behavior, hence the puzzle.

The joint behavior of prices and dividends is illustrated in figure 14.2 for an extended version of the annual Standard and Poor's 500 data analyzed by Shiller. The data span 1871–2004, and are contained in the file S&PData.txt, available at the textbook website.

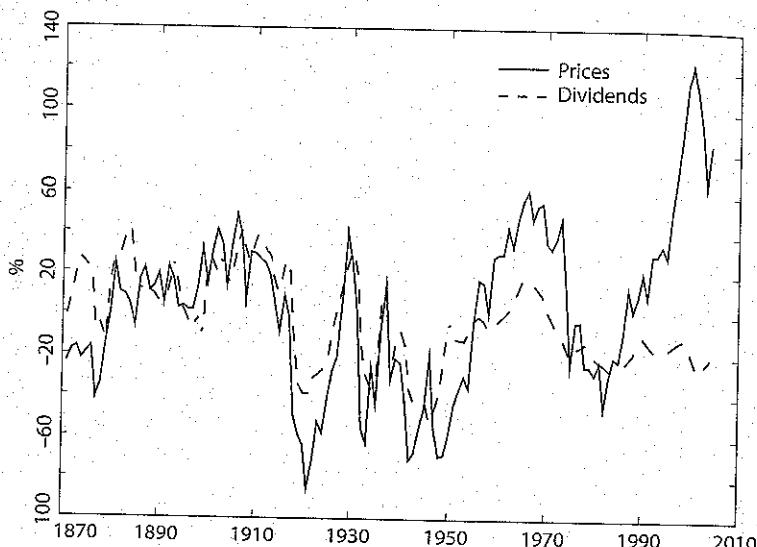


Figure 14.2 Stock Prices and Dividends (logged deviations from trend)

The data depicted in the figure are measured as logged deviations from a common trend (details regarding the trend are provided below). Prices are far more volatile than dividends; respective standard deviations are 40.97% and 22.17%. The correlation between the series is 0.409. As discussed below, absent this correlation virtually any preference specification can account for the observed pattern of volatility in Lucas' one-tree environment, if coupled with a sufficiently volatile endowment process. The problem is that endowment innovations weaken the link between price and dividend movements. Accounting for both the high volatility and correlation patterns observed in the data represents the crux of the empirical challenge facing the model.

The analysis presented here follows that of DeJong and Ripoll (2004), which is based on the data measured through 1999. The first version of the model features CRRA preferences; the second features preferences specified following Gul and Pesendorfer (2004), under which the household faces a temptation to deviate from its intertemporally optimal consumption plan by selling its entire holding of shares and maximizing current-period consumption. This temptation imposes a self-control cost that affects the household's demand for share holdings. DeJong and Ripoll's analysis centers on the relative ability of these models to account for the joint behavior of prices and dividends.

Each version of the model reduces to a system of three equations: a pricing kernel that dictates the behavior of  $p_t$ , laws of motion for dividends ( $d_t$ ), and the household's endowment ( $q_t$ ), which are taken as exogenous. Dropping time subscripts, so that  $p_t$  and  $p_{t+1}$  are represented as  $p$  and  $p'$ , and so on, and letting  $MRS$  denote the marginal rate of substitution between periods  $t$  and  $t+1$ , the equations are given by

$$p = \beta E_t[MRS \cdot (d' + p')], \quad (14.55)$$

$$\log d' = (1 - \rho_d) \log(\bar{d}) + \rho_d \log(d) + \varepsilon_{dt} \quad (14.56)$$

$$\log q' = (1 - \rho_q) \log(\bar{q}) + \rho_q \log(q) + \varepsilon_{qt}, \quad (14.57)$$

with  $|\rho_i| < 1$ ,  $i = d, q$ , and

$$\begin{bmatrix} \varepsilon_{dt} \\ \varepsilon_{qt} \end{bmatrix} \sim iidN(0, \Sigma).$$

Under CRRA preferences, (14.55) is parameterized as

$$p = \beta E_t \left\{ \frac{(d' + q')^{-\gamma}}{(d + q)^{-\gamma}} (d' + p') \right\}, \quad (14.58)$$

where  $\gamma > 0$  measures the degree of relative risk aversion. Regarding the stock-price volatility puzzle, notice that a relatively large value of  $\gamma$  intensifies the household's desire for maintaining a smooth consumption profile, leading to an increase in the predicted volatility of price responses to exogenous shocks. Thus any amount of price volatility can be captured through the specification of a sufficiently large value of  $\gamma$ . However, increases in  $\gamma$  entail the empirical cost of decreasing the predicted correlation between  $p_t$  and  $d_t$ , because increases in  $\gamma$  heighten the role assigned to  $q_t$  in driving price fluctuations. This is the puzzle in a nutshell.

Under self-control preferences, the momentary and temptation utility functions are parameterized as

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma}, \quad (14.59)$$

$$v(c) = \lambda \frac{c^\phi}{\phi}, \quad (14.60)$$

with  $\gamma > 0$ ,  $\lambda > 0$ , and  $1 > \phi > 0$ . These imply the following specification for the pricing kernel:

$$p = \beta E \left[ \frac{\frac{(d' + q')^{-\gamma}}{(d + q)^{-\gamma}} + \lambda(d + q)^\gamma [(d' + q')^{\phi-1} - (d' + q' + p')^{\phi-1}]}{1 + \lambda(d + q)^{\phi-1+\gamma}} \times [d' + p'] \right] \quad (14.61)$$

Notice that when  $\lambda = 0$ , there is no temptation, and the pricing equation reduces to the CRRA case.

To consider the potential for temptation in helping to resolve the stock-price volatility puzzle, suppose  $\lambda > 0$ . Then given a positive shock to either  $d$  or  $q$ , the term

$$\lambda(d+q)^\gamma[(d'+q')^{\phi-1} - (d'+q'+p')^{\phi-1}],$$

which appears in the numerator of the pricing kernel, increases. However, if  $\phi - 1 + \gamma > 0$ , that is, if the risk-aversion parameter  $\gamma > 1$ , then the denominator of the kernel also increases. If the increase in the numerator dominates that in the denominator, then higher price volatility can be observed under temptation than in the CRRA case.

To understand this effect, note that the derivative of the utility cost of self-control with respect to wealth is positive given  $\phi < 1$ , so that  $v'(\cdot)$  is concave:

$$v'(d'+q') - v'(d'+q'+p') > 0.$$

That is, as the household gets wealthier, self-control costs become lower. A reduction in self-control costs serves to heighten the household's incentive to save, and thus its demand for asset holdings. Thus an exogenous shock to wealth increases share prices beyond any increase resulting from the household's incentives for maintaining a smooth consumption profile. This explains why it might be possible to get higher price volatility in this case. Whether this extension is capable of capturing the full extent of the volatility observed in stock prices, while also accounting for the close correlation observed between prices and dividends, is the empirical question to which we now turn.

The first empirical objective is to obtain parameter estimates for each model. This is accomplished using as observables  $X_t = (p_t \ d_t)'$ , which as noted are measured as logged deviations from a common trend. Regarding the trend, Shiller's (1981) analysis was based on the assumption that dividends and prices are trend stationary; DeJong (1992) and DeJong and Whiteman (1991, 1994b) provided subsequent empirical support for this assumption. Coupled with the restriction imposed by the asset-pricing kernel between the relative steady state levels of dividends and prices, this assumption implies a restricted trend-stationarity specification for prices.

Regarding steady states, since  $\{d_t\}$  and  $\{q_t\}$  are exogenous, their steady states  $\bar{d}$  and  $\bar{q}$  are simply parameters. Normalizing  $d$  to 1 and defining  $\eta = \frac{\bar{q}}{d}$ , so that  $\eta = \bar{q}$ , the steady state value of prices implied by (14.58) for the CRRA specification is given by

$$\begin{aligned}\bar{p} &= \frac{\beta}{1-\beta} \bar{d} \\ &= \frac{\beta}{1-\beta}.\end{aligned}\quad (14.62)$$

Letting  $\beta = 1/(1+r)$ , where  $r$  denotes the household's discount rate, (14.62) implies  $\bar{p}/\bar{d} = 1/r$ . Thus as the household's discount rate increases, its asset demand decreases, driving down the steady state price level. Empirically, the average price/dividend ratio observed in the data serves to pin down  $\beta$  under this specification of preferences.

Under the temptation specification, the steady state specification for prices implied by (14.61) is given by

$$\bar{p} = \beta(1+\bar{p}) \left[ \frac{(1+\eta)^{-\gamma} + \lambda(1+\eta)^{\phi-1} - \lambda(1+\eta+\bar{p})^{\phi-1}}{(1+\eta)^{-\gamma} + \lambda(1+\eta)^{\phi-1}} \right]. \quad (14.63)$$

The left-hand side of this expression is a 45-degree line. The right-hand side is strictly concave in  $\bar{p}$ , has a positive intercept, and positive slope that is less than one at the intercept. Thus (14.63) yields a unique positive solution for  $\bar{p}$  for any admissible parameterization of the model. An increase in  $\lambda$  causes the function of  $\bar{p}$  on the right-hand side of (14.63) to shift down and flatten; thus  $\bar{p}$  is decreasing in  $\lambda$ . The intuition for this is that an increase in  $\lambda$  represents an intensification of the household's temptation to liquidate its asset holdings. This drives down its demand for asset shares, and thus  $\bar{p}$ . Note the parallel between this effect and that generated by an increase in  $r$ , or a decrease in  $\beta$ , which operates analogously in both (14.62) and (14.63).

Returning to the problem of detrending, given the value of  $\bar{p}/\bar{d}$  implied under either (14.62) or (14.63), the assumption of trend-stationarity for dividends, coupled with this steady state restriction, yields:

$$\begin{aligned}\ln(\bar{d}_t) &= \phi_0 + \phi_1 t \\ \ln(\bar{p}_t) &= \left[ \phi_0 + \ln\left(\frac{\bar{p}}{\bar{d}}\right) \right] + \phi_1 t.\end{aligned}\quad (14.64)$$

Thus detrending is achieved by computing logged deviations of prices and dividends from the restricted trajectories given in (14.64). Because the restrictions are a function of the collection of deep parameters  $\mu$ , the raw data must be retransformed for every reparameterization of the model. Specifically, given a particular specification of  $\mu$ , which implies a particular value for  $\bar{p}/\bar{d}$ , logged prices and dividends are regressed on a constant and linear time trend, given the imposition of the parameter restrictions indicated in (14.64). Deviations of the logged variables from their respective trend

TABLE 14.3  
Parameter Estimates

	$\beta$	$\gamma$	$\lambda$	$\phi$	$\rho_d$	$\rho_q$	$\sigma_{ed}$	$\sigma_{eq}$	$\text{corr}(e_d, e_q)$	$\frac{\bar{q}}{d}$
<b>Means</b>										
Prior	0.960	2.000	0.0000	0.4	0.900	0.900	UN	UN	UN	10.00
CRRA	0.958	2.914	NA	NA	0.877	0.914	0.114	0.096	0.311	10.75
Self-con.	0.967	2.04	0.00294	0.32	0.879	0.910	0.114	0.149	0.320	7.54
<b>Std. dev.</b>										
Prior	0.020	1.000	0.01	0.2	0.050	0.050	UN	UN	UN	5.00
CRRA	0.004	0.826	NA	NA	0.033	0.027	0.007	0.034	0.130	4.44
Self-control	0.008	0.598	0.00330	0.16	0.029	0.023	0.007	0.043	0.124	3.65
<b>Posterior correlations</b>										
	$\gamma, \sigma_{eq}$	$\gamma, \text{corr}(e_d, e_q)$	$\sigma_{eq}, \text{corr}(e_d, e_q)$	$\text{corr}(e_d, e_q), \frac{\bar{q}}{d}$						
CRRA	-0.839	-0.337	0.201	0.557						
Self-control	-0.850	-0.298	0.164	0.556						

Note: UN denotes “uninformative prior”; NA denotes “not applicable”; and  $\text{corr}(x, y)$  denotes correlation between  $x$  and  $y$ .

specifications constitute the values of  $X_t$  that correspond with the specific parameterization of  $\mu$ .<sup>3</sup>

Under the CRRA specification, the parameters to be estimated are the discount factor  $\beta$ ; the risk-aversion parameter  $\gamma$ ; the AR parameters  $\rho_d$  and  $\rho_q$ ; the standard deviations of innovations to  $d$  and  $q$ , denoted as  $\sigma_{ed}$  and  $\sigma_{eq}$ ; the correlation between these innovations  $\text{corr}(e_d, e_q)$ ; and the steady state ratio  $\eta = \frac{\bar{q}}{d}$ . Under the temptation specification, the curvature parameter  $\phi$  and the temptation parameter  $\lambda$  are added to this list.

In specifying prior distributions, DeJong and Ripoll placed an emphasis on standard ranges of parameter values, while assigning sufficient prior uncertainty to allow the data to have a nontrivial influence on their estimates. The priors they used for this purpose are summarized in table 14.3. In all cases, prior correlations across parameters are zero, and each of the informative priors we specify is Normally distributed (but truncated when appropriate).

Consider first the parameters common across both models. The prior mean (standard deviation) of the discount factor  $\beta$  is 0.96 (0.02), implying an annual discount rate of 4%. (This compares with an average dividend/price ratio of 4.6% in the data.) The prior for the risk aversion

<sup>3</sup> The GAUSS procedure ct.prc is available for use in removing a common trend from a collection of variables.

#### 14.7 Estimating and Comparing Asset-Pricing Models

parameter  $\gamma$  is centered at 2 with a standard deviation of 1. A noninformative prior is specified over the covariance matrix of the shocks  $\Sigma$ , proportional to  $\det(\Sigma)^{-(m+1)/2}$  (where  $m$  is the dimension of  $\Sigma$ , or 2 in this case), and the priors over the persistence parameters ( $\rho_d, \rho_q$ ) are centered at 0.9, with standard deviations of 0.05. The prior over  $\eta = \frac{\bar{q}}{d}$  is more problematic, because neither the literature nor a prior guidance is available to aid in its specification. DeJong and Ripoll considered two alternative specifications of prior means (standard deviations) for  $\eta$ : 5 (3) and 10 (5). In so doing, they found that the data are relatively uninformative regarding the location of  $\eta$ , so that the prior dominates the posterior along this dimension. They also found that the prior over  $\eta$  had little influence on either the fit of the competing models or estimates of the additional parameters. Here we report results obtained using the (10, 5) specification. Finally, the priors over  $(\beta, \rho_d, \rho_q)$  are truncated from below at zero and from above at 1, and the prior over  $\eta$  is truncated from below at zero.

Priors for the self-control parameters  $\lambda$  and  $\phi$  were specified by focusing on the implications they carry for the steady state relationship between prices and dividends, as indicated in (14.63). Note that along with  $\lambda$  and  $\phi$ , the parameters  $\beta$ ,  $\gamma$ , and  $\eta$  also influence this relationship. When  $\lambda = 0$ , the model reverts to the CRRA case, under which the prior over  $\beta$  is centered on 0.96, implying an annual rate of return of 4%, or a steady state price/dividend ratio of 24. As noted, the average rate of return in the data is 4.6%, implying an average price/dividend ratio of approximately 22. Thus parameterizations of  $\lambda$  and  $\phi$  were chosen with this ratio as an approximate target. Fixing  $\beta$ ,  $\gamma$ , and  $\eta$  at their prior means,  $(\lambda, \phi)$  combinations in the respective ranges  $[[0, 0.001], [0, 0.75]]$  deliver this behavior. Increasing  $\beta$  by one prior standard deviation, up to 0.98, moves these ranges to  $[[0, 0.006], [0, 0.8]]$ ; recentering  $\beta$  at 0.96 and decreasing  $\gamma$  by one standard deviation to 1 moves these ranges to  $[[0, 0.014], [0, 0.8]]$ ; and recentering  $\gamma$  and decreasing  $\eta$  by one standard deviation to 5 moves these ranges to  $[[0, 0.004], [0, 0.7]]$ . Thus notice that small changes in  $\lambda$  can have large impacts on steady state price/dividend values, whereas changes in  $\phi$  have relatively small impacts. Recall the intuition behind the impact of  $\lambda$ : as the household's temptation to consume increases, its demand for asset holdings decreases, thus driving down  $\bar{p}$ . In light of these calculations, DeJong and Ripoll specified a Normal distribution for  $\lambda$  centered and truncated from below at 0, with a standard deviation of 0.01; and a Normal distribution for  $\phi$  centered at 0.4 with a standard deviation of 0.2, truncated from below at 0 and from above at 1.

The models were estimated using the importance sampling procedure described in section 4. A multivariate- $t$  distribution was specified for  $I(\mu; \alpha)$ , centered initially on prior means specified for the individual elements of  $\mu$ . The initial covariance matrix specified for  $I(\mu; \alpha)$  was diagonal,

with standard deviations specified as two times larger than the standard deviations specified under the prior. A second specification of  $I(\mu; \alpha)$  was then constructed using the estimated posterior mean and covariance matrix obtained for  $\mu$  based on drawings from the initial sampler. After several rounds, moment calculations converged (subject to numerical sampling error) to those reported below.

Due to the method used to detrend the data, the algorithm used to facilitate integration works as follows. Obtain a drawing of  $\mu$  from  $I(\mu; \alpha)$ . Calculate  $\bar{p}/\bar{d}$  based on this drawing, and detrend the raw data using (14.64). Solve the model, and evaluate the likelihood function associated with the corresponding state-space representation. Calculate the weight  $w(\mu)$  associated with the drawing, along with functions of interest  $g(\mu)$ . Finally, compute weighted averages  $\bar{g}_N$  as in (14.22) given  $N$  replications of this process.

#### 14.7.1 Estimates

Posterior means and standard deviations obtained in this manner for both model specifications are presented in table 14.3. For the CRRA specification, marginal posterior distributions of parameters are graphed in figure 14.3; for the temptation specification, these distributions are graphed in figures 14.4 and 14.5.

For the CRRA specification, the posterior distribution of  $\beta$  is centered very close to the prior (with respective means of 0.957 and 0.96), but is much more tightly distributed (its standard deviation of 0.004 is five times less than the prior's). In contrast, the posterior and prior dispersions of  $\gamma$  are similar (respective standard deviations are 0.814 and 1), whereas the posterior distribution is moderately right-shifted relative to the prior (respective means are 2.884 and 2). Thus the data indicate a somewhat higher degree of risk aversion than was embodied in the prior, but not by a dramatic amount: the means differ by less than one prior standard deviation. Regarding the shocks, the persistence of the endowment shock is greater than that of the dividend shock (posterior means of  $\rho_d$  and  $\rho_q$  are 0.876 and 0.913, with standard deviations of 0.033 and 0.026), and the shocks are positively correlated (the posterior mean of  $\text{corr}(\varepsilon_d, \varepsilon_q)$  is 0.313, with posterior standard deviation of 0.124). Regarding the size of the shocks, the posterior distributions of  $\sigma_{\varepsilon_d}$  and  $\sigma_{\varepsilon_q}$  have similar means (0.114 and 0.097, respectively), but the standard deviation of  $\sigma_{\varepsilon_q}$  is five times that of  $\sigma_{\varepsilon_d}$ , and is skewed substantially to the right. Finally, as noted above, the prior and posterior distributions of  $\eta = \bar{q}/\bar{d}$  are virtually indiscernible, indicating that the data are uninformative regarding the location of this ratio.

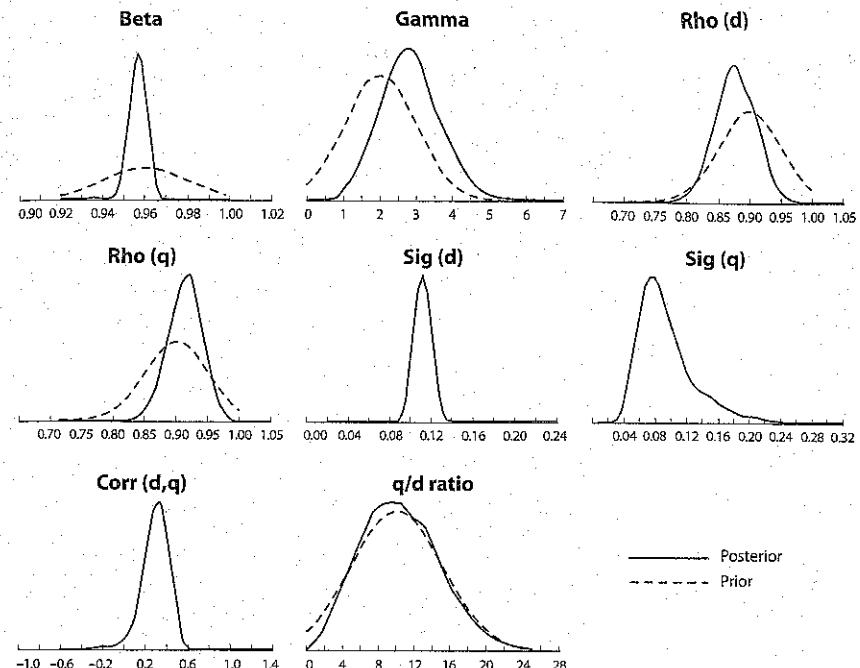


Figure 14.3 Parameter Estimates, CRRA Preferences

To gain intuition for these results, it is important to keep in mind that the empirical problem confronting the model is largely two-dimensional: simultaneously account for the relatively high volatility of stock prices, and the close correlation observed between prices and dividends. Four parameters play a critical role in confronting this problem. First, large values of  $\sigma_{\varepsilon_q}$  and  $\eta$  are useful in helping the model along the former dimension: by respectively increasing the volatility and importance of the household's endowment in the budget constraint, these parameters serve to increase the volatility of stock prices without requiring an increase in the volatility of dividends. But these effects are harmful along the latter dimension: they serve to weaken the correlation between movements in prices and dividends. However, this harm can be offset given a large corresponding value of  $\text{corr}(\varepsilon_d, \varepsilon_q)$ , because this boosts the correlation between dividends and the endowment process, and thus between dividends and prices. Finally, large values of  $\gamma$  provide help along the former dimension by boosting the

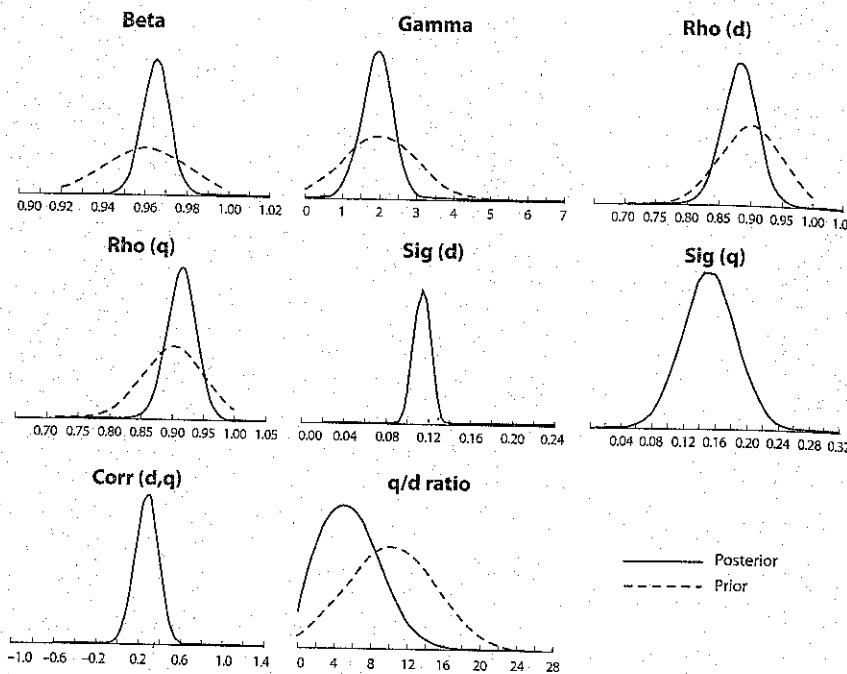


Figure 14.4 Parameter Estimates, Self-Control Preferences

volatility of the household's marginal rate of substitution for a given volatility of consumption. But this again serves to weaken the correlation between prices and dividends, because dividend fluctuations do not constitute the exclusive source of fluctuations in consumption.

These considerations are helpful in understanding, for example, why the posterior distribution of  $\gamma$  is only moderately right-shifted relative to the prior, and why the data clearly favor a positive value for  $\text{corr}(\varepsilon_d, \varepsilon_q)$ . They are also helpful in interpreting the posterior correlations estimated between the parameters of the model. Among the eight parameters featured in this version of the model, there are four nontrivial posterior correlations (reported in table 14.3): between  $\gamma$  and  $\sigma_{eq}$  (-0.836),  $\gamma$  and  $\text{corr}(\varepsilon_d, \varepsilon_q)$  (-0.302),  $\sigma_{eq}$  and  $\text{corr}(\varepsilon_d, \varepsilon_q)$  (0.165), and  $\text{corr}(\varepsilon_d, \varepsilon_q)$  and  $\eta$  (0.520). So for example, given a relatively large value of  $\sigma_{eq}$ ,  $\text{corr}(\varepsilon_d, \varepsilon_q)$  is also relatively large, and  $\gamma$  is relatively small: adjustments in these latter parameters help combat the decreased dividend/price correlation associated with the large value of  $\sigma_{eq}$ .

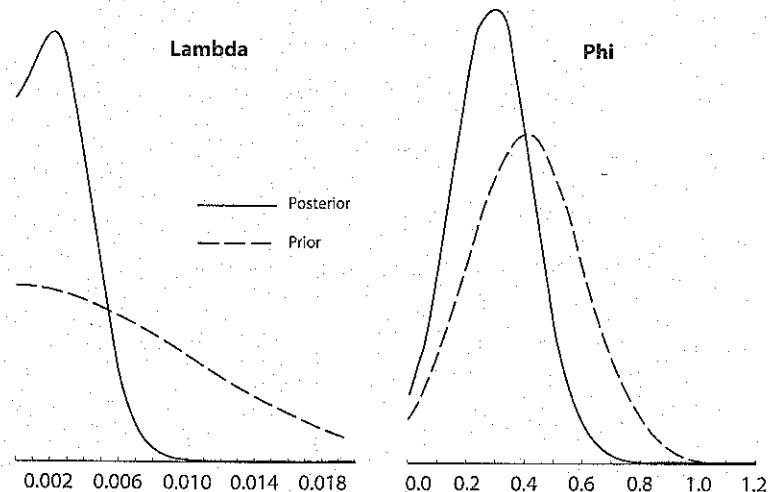


Figure 14.5 Parameter Estimates, Self-Control Preferences, Continued

Turning to results obtained using the temptation specification, the estimates obtained for the parameters common to both the self-control and CRRA specifications are relatively distinct. Most notably, the posterior mean of  $\beta$  increases by approximately one standard deviation under the self-control specification; the means of  $\gamma$  and  $\eta$  decrease by approximately one standard deviation; and the mean of  $\sigma_{eq}$  increases by more than 1.5 standard deviations. Anticipating the discussion of fit to follow, we note here that decreases in  $\gamma$  and  $\eta$  will serve to increase the correlation between prices and dividends in the model, while dampening the volatility of prices; the increase in  $\sigma_{eq}$  will have the opposite effect. This raises the question of why these offsetting factors arise. The key lies in reconsidering the relationship between steady state prices and dividends given in (14.63). Movements away from zero by the self-control parameters  $\lambda$  and  $\phi$  cause sharp declines in the steady state price/dividend ratio; decreases in  $\gamma$  and  $\eta$ , and increases in  $\beta$ , can serve to offset these effects. Thus it appears that the altered estimates of  $\gamma$ ,  $\eta$ , and  $\beta$  arise from an empirical preference for nonzero values of  $\lambda$  and  $\phi$ ; the posterior estimates of these self-control parameters confirm this observation.

The posterior mean of  $\phi$  is 0.3, roughly one posterior standard deviation below its prior mean, but still significantly higher than zero. And the posterior mean of  $\lambda$  is 0.00286, which coincides closely with its mode, and lies approximately 1.5 posterior standard deviations above the prior mode of

zero. To appreciate the quantitative significance of this estimate, consider two alternative calculations. First, returning to (14.63), if  $\gamma$ ,  $\eta$ ,  $\beta$ , and  $\phi$  are fixed at their posterior means, an increase of  $\lambda$  from zero to 0.00286 implies a decrease in the steady state price/dividend ratio from approximately 28 to 20.46, implying an increase in annual average returns from 3.57% to 4.89%. In other words, a household endowed with the self-control preferences associated with these parameter estimates would demand an annual "return premium" over a household endowed with CRRA preferences of 1.32%, representing a 36% difference. Second, consider the quantity of steady state consumption a household endowed with the temptation specification we estimate would be willing to sacrifice in order to be rid of this temptation. This is the value  $x$  such that

$$u(\bar{c} - x) = u(\bar{c}) + v(\bar{c}) - v(\bar{c}),$$

where here  $\bar{c}$  denotes the steady state value of temptation consumption. Using posterior means of parameter estimates, the implied value of  $x$  amounts to 5.25% of  $\bar{c}$ . Thus these results support the presence of a quantitatively significant temptation effect in the data.

### 14.7.2 Model Comparison

To assess the relative ability of the competing models to account for the behavior of dividends and prices, DeJong and Ripoll conducted a moment-based posterior-odds comparison following Geweke (1999b), as described in section 5. Four moments were chosen for comparison: the standard deviations of prices and dividends  $\sigma_p$  and  $\sigma_d$ , the ratio of standard deviations  $\sigma_d/\sigma_p$ , and the correlation between prices and dividends  $\text{corr}(p, d)$ . In order to characterize implications for these moments conveyed by the data, posterior distributions were constructed using a six-lag VAR specified for the data. Posterior means and standard deviations obtained using the data and both versions of the model are reported in table 14.4. Corresponding posterior-odds calculations were obtained by approximating the integrals in (14.32) using the kernel density approximation given by (14.33), with a density kernel provided by the Normal distribution. These are also reported in table 14.4. Finally, marginal posterior distributions over these statistics obtained under the CRRA and self-control specifications are depicted in figures 14.6 and 14.7. For comparison, corresponding distributions obtained using the VAR specification are also depicted in both figures.

Regarding the distributions obtained using the VAR specification, note that the distributions over  $\sigma_p$  and  $\text{corr}(p, d)$  exhibit considerable skewness. As a result, for example, the posterior mean of  $\sigma_p$  (0.62) lies considerably above its sample average of 0.406 (computed using the data measured through 1999).

TABLE 14.4  
Summary Statistics

	$\sigma_p$	$\sigma_d$	$\frac{\sigma_d}{\sigma_p}$	$\text{Corr}(p, d)$
<b>Means</b>				
VAR	0.620	0.268	0.448	0.672
CRRA	0.434	0.244	0.571	0.536
Self-control	0.418	0.248	0.608	0.564
<b>Std. Dev.</b>				
VAR	0.183	0.065	0.095	0.178
CRRA	0.067	0.041	0.101	0.079
Self-control	0.069	0.035	0.125	0.101

Note: VAR statistics summarize flat-prior posterior distributions associated with a six-lag vector autoregression;  $\sigma_x$  denotes the standard deviation of the logged deviation of  $x$  from its steady state value; and  $\text{Corr}(x, y)$  denotes the correlation between logged deviations from steady state values of  $x$  and  $y$ .

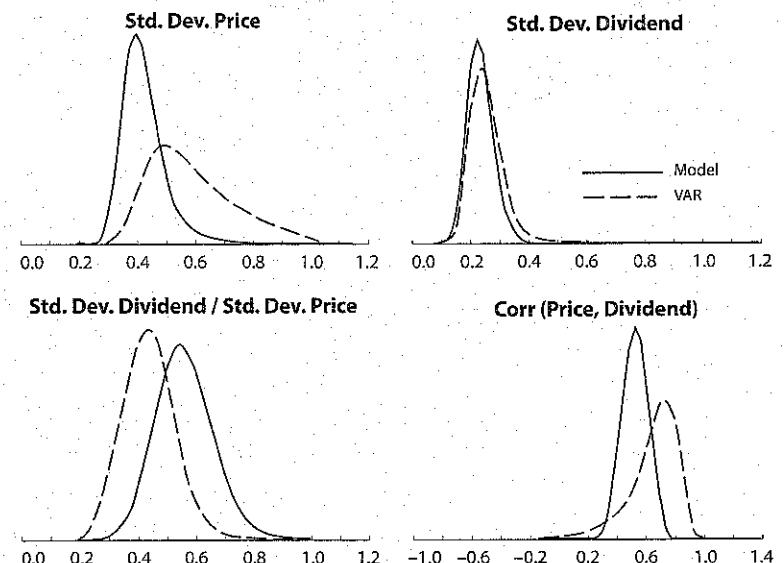


Figure 14.6 Summary Statistics, CRRA Preferences

The relative performance of the competing models in matching the data along these dimensions is very similar. Along the  $\sigma_p$  dimension, respective posterior means (standard deviations) are 0.62 (0.183) in the data, 0.434 (0.067) under the CRRA specification, and 0.418 (0.069) under

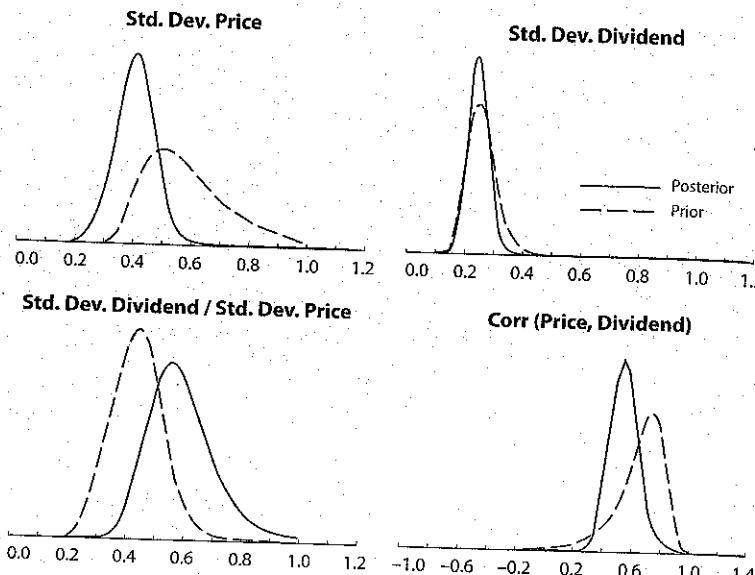


Figure 14.7 Summary Statistics, Self-Control Preferences

the self-control specification. Along the  $\text{corr}(p, d)$  dimension, respective means (standard deviations) are 0.672 (0.178), 0.536 (0.079), and 0.564 (0.101). These differences translate into a posterior-odds ratio of 2.4:1 in favor of the self-control specification, which constitutes “very slight evidence” in favor of the self-control specification according to Jeffreys’ (1961) guidelines.

In conclusion, although the parameter estimates reported above support the presence of a nontrivial self-control effect, the effect fails to translate into a significant improvement over the account of the stock-price volatility puzzle provided by CRRA preferences. Moreover, in an extension of this analysis to the two-asset environment introduced in section 3.2 of chapter 3, DeJong and Ripoll (2007) obtained a similar result given the joint investigation of the volatility, risk-free rate, and equity premium puzzles.

## References

- Adda, J., and R. Cooper, 2003. *Dynamic Economics*. Cambridge: MIT Press.
- Akaike, H., 1974. “A New Look at the Statistical Model Identification,” *IEEE Transactions on Automatic Control*, AC 19, 716–23.
- Akashi, H., and H. Kumamoto, 1977. “Random Sampling Approach to State Estimation in Switching Environment,” *Automatica*, 13, 429–34.
- Altug, S., 1989. “Time-to-Build and Aggregate Fluctuations: Some New Evidence,” *International Economic Review*, 30, 889–920.
- Amemiya, T., 1985. *Advanced Econometrics*, Cambridge: Harvard University Press.
- Andrews, D.W.K., 1993. “Tests for Parameter Instability and Structural Change with Unknown Change Point,” *Econometrica*, 61, 821–56.
- An, S., and F. Schorfheide, 2007. “Bayesian Analysis of DSGE Models,” *Econometric Reviews*, 26, 113–72.
- Andrews, D.W.K., and R. C. Fair, 1988. “Inference in Nonlinear Econometric Models with Structural Change,” *Review of Economic Studies*, 55, 615–40.
- Aruoba, S. B., J. Fernandez-Villaverde, and J. Rubio-Ramirez, 2006. “Comparing Solution Methods for Dynamic Equilibrium Economies,” *Journal of Economic Dynamics and Control*, 30, 2477–2508.
- Azariadis, C., 1993. *Intertemporal Macroeconomics*, Oxford: Blackwell.
- Bai, J., and S. Ng, 2002. “Determining the Number of Factors in Approximate Factor Models,” *Econometrica*, 70, 191–221.
- , 2008. “Large-Dimensional Factor Analysis,” *Foundations and Trends in Econometrics*, 3, 89–163.
- Bank of Sweden, 1989. Press Release: “The Sveriges Riksbank (Bank of Sweden) Prize in Economic Sciences in Memory of Alfred Nobel for 1989,” <http://nobelprize.org/economics/laureates/1989/press.html>.
- , 2004. Press Release: “The Sveriges Riksbank (Bank of Sweden) Prize in Economic Sciences in Memory of Alfred Nobel for 2004,” <http://nobelprize.org/economics/laureates/2004/press.html>.
- Barillas, F., and J. Fernandez-Villaverde, 2007. “A Generalization of the Endogenous Grid Method,” *Journal of Economic Dynamics and Control*, 31, 2698–2712.
- Barro, R. J., 1989. *Modern Business Cycle Theory*, Cambridge: Harvard University Press.
- Barro, R. J., and X. Sala-i-Martin, 2004. *Economic Growth*, 2nd ed., Cambridge: MIT Press.
- Bauwens, L., and M. Lubrano, 1998. “Bayesian Inference on GARCH Models using the Gibbs Sampler,” *Econometrics Journal*, 1, C23–C46.
- Bauwens, L., M. Lubrano, and J.-F. Richard, 1999. *Bayesian Inference in Dynamic Econometric Models*, Oxford: Oxford University Press.
- Baxter, M., and R. G. King, 1999. “Measuring Business Cycles: Approximate Band-Pass Filters for Economic Time Series,” *Review of Economics and Statistics*, 81, 575–93.

- Bell, W. R., and B. C. Monsell, 1992. "X-11 Symmetric Linear Filters and Their Transfer Functions," *Bureau of the Census Statistical Research Division Report Series*, RR-92/15.
- Benassy, J., 2002. *The Macroeconomics of Imperfect Competition and Nonclearing Markets*, Cambridge: MIT Press.
- Benhabib, J., R. Rogerson, and R. Wright, 1991. "Homework in Macroeconomics: Household Production and Aggregate Fluctuations," *Journal of Political Economy*, 99, 1166–87.
- Berger, J. O., and R. L. Wolpert, 1988. *The Likelihood Principle*, 2nd ed., Hayward, CA: Institute of Mathematical Statistics.
- Bernanke, B. S., J. Boivin, and P. Eliasz, 2005. "Measuring the Effects of Monetary Policy: A Factor-Augmented Vector Autoregressive (FAVAR) Approach," *Quarterly Journal of Economics*, 120, 387–422.
- Blake, A., and M. Isard, 1998. *Active Contours*, New York: Springer.
- Blanchard, O. J., and S. Fischer, 1998. *Lectures on Macroeconomics*, Cambridge: MIT Press.
- Blanchard, O. J., and C. M. Kahn, 1980. "The Solution of Linear Difference Models under Rational Expectations," *Econometrica*, 48, 1305–11.
- Bobrovsky, B. Z., and M. Zakai, 1975. "A Lower Bound on the Estimation Error for Markov Processes," *IEEE Transactions on Automatic Control*, 20, 785–88.
- Boldrin, M., L. J. Christiano, and J. Fisher, 2001. "Asset Pricing Lessons for Modeling Business Cycles," *American Economic Review*, 91, 149–66.
- Bollerslev, T., 1986. "Generalized Autoregressive Conditional Heteroscedasticity," *Journal of Econometrics*, 31, 307–27.
- Bollerslev, T., R. Y. Chou, and K. F. Kroner, 1992. "ARCH Modelling in Finance: A Review of the Theory and Empirical Evidence," *Journal of Econometrics*, 52, 5–59.
- Brooks, C., S. Berke, and G. Persand, 2003. "Multivariate GARCH Models: Software Choice and Estimation Issues," ISMA Centre Discussion Papers in Finance, University of Reading.
- Brown, J. W., and R. V. Churchill, 2003. *Complex Variables and Applications*, 7th ed., Boston: McGraw-Hill.
- Burnside, C., and M. Eichenbaum, 1997. "Factor Hoarding and the Propagation of Business-Cycle Shocks," *American Economic Review*, 86, 1154–74.
- Burnside, C., M. Eichenbaum, and S. Rebelo, 1993. "Labor Hoarding and the Business Cycle," *Journal of Political Economy*, 101, 245–73.
- Campbell, J. Y., A. W. Lo, and A. C. MacKinlay, 1997. *The Econometrics of Financial Markets*, Princeton: Princeton University Press.
- Canova, F., 1995. "Sensitivity Analysis and Model Evaluation in Simulated Dynamic General Equilibrium Models," *International Economic Review*, 36, 477–501.
- , 2007. *Methods for Applied Macroeconomic Research*. Princeton: Princeton University Press.
- Cappé, O., S. J. Godsill, and E. Moulines, 2007. "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo," *Proceedings of the IEEE*, 95, 899–924.

- Carroll, C. D., 2006. "The Method of Endogenous Gridpoints for Solving Dynamic Stochastic Optimization Problems," *Economics Letters*, 91, 312–20.
- Chan, K., J. C. Hayya, and K. Ord, 1977. "A Note on Trend Removal Methods: The Case of Polynomial versus Variate Differencing," *Econometrica*, 45, 737–44.
- Chari, V. V., P. J. Kehoe, and E. R. McGrattan, 2004. "Business Cycle Accounting," Working Paper No. 10351, National Bureau of Economic Research, Cambridge, MA.
- , 2005. "A Critique of Structural VARs Using Business Cycle Theory," Research Department Staff Report 364, Federal Reserve Bank of Minneapolis.
- , 2009. "New Keynesian Models: Not Yet Useful for Policy Analysis," *American Economic Journal: Macroeconomics*, 1, 242–66.
- Chib, S., and E. Greenberg, 1995. "Understanding the Metropolis-Hastings Algorithm," *American Statistician*, 49, 327–35.
- Chib, S., and S. Ramamurthy, 2009. "Tailored Randomized-Block MCMC Methods with Application to DSGE Models," working paper, Washington University of St. Louis.
- Christiano, L. J., 2002. "Solving Dynamic Equilibrium Models by a Method of Undetermined Coefficients," *Computational Economics*, 20, 21–55.
- Christiano, L., and M. Eichenbaum, 1992. "Current Real-Business-Cycle Theories and Aggregate Labor Market Fluctuations," *American Economic Review*, 82, 430–50.
- Christiano, L., M. Eichenbaum, and C. L. Evans, 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. Vigfusson, 2003. "What Happens after a Technology Shock?" International Finance Discussion Paper Number 768, Board of Governors of the Federal Reserve System.
- Christiano, L. J., and T. J. Fitzgerald, 1999. "The Band Pass Filter," Federal Reserve Bank of Cleveland Working Paper.
- Christiano, L. J., and R. M. Todd, 1996. "Time to Plan and Aggregate Fluctuations," *Federal Reserve Bank of Minneapolis Quarterly Review*, Winter, 14–27.
- Clarida, R., J. Gali, and M. Gertler, 2000. "Monetary Policy Rules and Macroeconomic Stability: Evidence and Some Theory," *Quarterly Journal of Economics*, 115, 147–80.
- Cochrane, J. H., 2001. *Asset Pricing*, Princeton: Princeton University Press.
- Cogley, T., and J. M. Nason, 1995. "Effects of the Hodrick-Prescott Filter on Trend and Difference Stationary Time Series: Implications for Business Cycle Research," *Journal of Economic Dynamics and Control*, 19, 253–78.
- Cogley, T., and T. J. Sargent, 2001. "Evolving Post-World War II Inflation Dynamics," *NBER Macroeconomic Annual*, 16, 331–73.
- , 2005. "Drifts and Volatilities: Monetary Policies and Outcomes in the Post-World War II U.S.," *Review of Economic Dynamics*, 8, 262–302.
- Cooley, T. F., 1995. *Frontiers of Business Cycle Research*, Princeton: Princeton University Press.

- Cooley, T. F., and E. C. Prescott, 1976. "Estimation in the Presence of Stochastic Parameter Variation," *Econometrica*, 44, 167–84.
- , 1995. "Economic Growth and Business Cycles," in T. F. Cooley, ed., *Frontiers of Business Cycle Research*, Princeton: Princeton University Press.
- DeJong, D. N., 1992. "Co-Integration and Trend-Stationarity in Macroeconomic Time Series," *Journal of Econometrics*, 52, 347–70.
- DeJong, D. N., H. Dharmarajan, R. Liesenfeld, G. Moura, and J.-F. Richard, 2010. "Efficient Likelihood Evaluation of State-Space Representations," working paper, University of Pittsburgh.
- DeJong, D. N., and B. F. Ingram, 2001. "The Cyclical Behavior of Skill Acquisition," *Review of Economic Dynamics*, 4, 536–61.
- DeJong, D. N., B. F. Ingram, and C. H. Whiteman, 1993. "Evaluating VARs with Monetary Business Cycle Priors," *Proceedings of the American Statistical Association, Bayesian Statistical Science Section*, 1993, 160–69.
- , 1996. "A Bayesian Approach to Calibration," *Journal of Business and Economic Statistics*, 14, 1–9.
- , 2000a. "Keynesian Impulses versus Solow Residuals: Identifying Sources of Business Cycle Fluctuations," *Journal of Applied Econometrics*, 15, 311–29.
- , 2000b. "A Bayesian Approach to Dynamic Macroeconomics," *Journal of Econometrics*, 98, 203–23.
- DeJong, D. N., R. Liesenfeld, and J.-F. Richard, 2005. "A Nonlinear Forecasting Model of GDP Growth," *Review of Economics and Statistics*, 87, 697–708.
- , 2006. "Timing Structural Change: A Conditional Likelihood Approach," *Journal of Applied Econometrics*, 21, 175–90.
- DeJong, D. N., and M. Ripoll, 2004. "Self-Control Preferences and the Volatility of Stock Prices," working paper, University of Pittsburgh.
- , 2007. "Do Self-Control Preferences Help Explain the Puzzling Behavior of Asset Prices?" *Journal of Monetary Economics*, 54, 1035–50.
- DeJong, D. N., and C. H. Whiteman, 1991a. "The Temporal Stability of Dividends and Prices: Evidence from the Likelihood Function," *American Economic Review*, 81, 600–617.
- , 1991b. "Trends and Random Walks in Macroeconomic Time Series: A Reconsideration Based on the Likelihood Principle," *Journal of Monetary Economics*, 28, 221–54.
- , 1993. "Unit Roots in U.S. Macroeconomic Time Series: A Survey of Classical and Bayesian Methods," in D. Brillinger et al., eds., *New Directions in Time Series Analysis, Part II*, Berlin: Springer-Verlag.
- , 1994a. "The Forecasting Attributes of Trend- and Difference-Stationary Representations for Macroeconomic Time Series," *Journal of Forecasting*, 13, 279–97.
- , 1994b. "Modeling Stock Prices without Pretending to Know How to Induce Stationarity," *Econometric Theory*, 10, 701–19.
- Del Moral, P., 1996. "Nonlinear Filtering: Interacting Particle Solution," *Markov Processes and Related Fields*, 2, 555–79.
- Del Negro, M., and F. Schorfheide, 2004. "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45, 643–73.

- Diebold, F. X., L. E. Ohanian, and J. Berkowitz, 1998. "Dynamic Equilibrium Economies: A Framework for Comparing Model and Data," *Review of Economic Studies*, 65, 433–51.
- Doan, T., R. Litterman, and C. Sims, 1984. "Forecasting and Conditional Projection Using Realistic Prior Distributions," *Econometric Reviews*, 3, 1–100.
- Doucet, A., N. de Freitas, and N. Gordon, 2001. *Sequential Monte Carlo Methods in Practice*, Berlin: Springer-Verlag.
- Duffie, D., and K. Singleton, 1993. "Simulated Moment Estimation of Markov Models of Asset Prices," *Econometrica*, 61, 929–52.
- Durbin, J., and S. J. Koopman, 1997. "Monte Carlo Maximum Likelihood Estimation for Non-Gaussian State Space Models," *Biometrika*, 84, 669–84.
- Eichenbaum, M. S., 1983. "A Rational Expectations Equilibrium Model of Inventories of Finished Goods and Employment," *Journal of Monetary Economics*, 12, 71–96.
- Einarsson, T., and M. H. Marquis, 1997. "Home Production with Endogenous Growth," *Journal of Monetary Economics*, 39, 551–69.
- Engle, R. F., 1982. "Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of United Kingdom Inflation," *Econometrica*, 50, 987–1007.
- Fernandez-Villaverde, J., P. Guerron-Quintana, and J. Rubio-Ramirez, 2010. "Fortune or Virtue: Time-Variant Volatilities versus Parameter Drifting in US Data," working paper, University of Pennsylvania.
- Fernandez-Villaverde, J., and J. Rubio-Ramirez, 2004a. "Comparing Dynamic Equilibrium Models to Data: A Bayesian Approach," *Journal of Econometrics*, 123, 153–87.
- , 2004b. "Sequential Monte Carlo Filtering: An Example," Working Paper, University of Pennsylvania.
- , 2005. "Estimating Dynamic Equilibrium Economies: Linear versus Nonlinear Likelihood," *Journal of Applied Econometrics*, 20, 821–910.
- , 2007. "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74, 1059–87.
- Fernandez-Villaverde, J., J. Rubio-Ramirez, T. J. Sargent, and M. Watson, 2005. "A, B, C's (and D)'s for Understanding VARs," *American Economic Review*, 97, 1021–26.
- Person, W. E., and G. M. Constantinides, 1991. "Habit Formation and Durability in Aggregate Consumption: Empirical Tests," *Journal of Financial Economics*, 29, 199–240.
- Francis, N., and V. A. Ramey, 2003. "Is the Technology-Driven Real Business Cycle Hypothesis Dead? Shocks and Aggregate Fluctuations Revisited," working paper, University of California at San Diego.
- Friedland, B., 2005. *Control System Design: An Introduction to State-Space Methods*, Mineola, MN: Dover.
- Frisch, R., 1933a. "Editorial," *Econometrica*, 1, 1–5.
- , 1933b. "Propagation Problems and Impulse Problems in Dynamic Economics," in *Economic Essays in Honour of Gustav Cassel, October 20th, 1933*, London: Allen & Unwin.

- Frisch, R., 1970. "Econometrics in the World Today," in W. A. Eltis, M.F.G. Scott, and J. N. Wolf, eds., *Induction, Growth and Trade: Essays in Honor of Sir Roy Harrod*, Oxford: Clarendon Press.
- Gali, J., 1999. "Technology, Employment, and the Business Cycle: Do Technology Shocks Explain Aggregate Fluctuations?" *American Economic Review*, 89, 249–71.
- Geweke, J., 1977. "The Dynamic Factor Analysis of Economic Time Series," in D. J. Aigner and A. S. Goldberger, eds., *Latent Variables in Socio-Economic Models*, Amsterdam: North Holland.
- , 1988. "Antithetic Acceleration of Monte Carlo Integration in Bayesian Inference," *Journal of Econometrics*, 38, 73–89.
- , 1989. "Bayesian Inference in Econometric Models Using Monte Carlo Integration," *Econometrica*, 57, 1317–39.
- , 1992. "Evaluating the Accuracy of Sampling-Based Approaches to the Calculation of Posterior Moments," in J. M. Bernardo, J. O. Berger, A. P. David, and A. F. M. Smith, eds., *Bayesian Statistics*, 4, Oxford: Clarendon Press.
- , 1999a. "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18, 1–126.
- , 1999b. "Computational Experiments and Reality," working paper, University of Iowa.
- , 2005. *Contemporary Bayesian Econometrics and Statistics*, Hoboken, NJ: Wiley.
- Ghez, G. R., and G. S. Becker, 1975. *The Allocation of Time and Goods over the Life Cycle*, New York: Columbia University Press.
- Gordon, N. J., D. J. Salmon, and A.F.M. Smith, 1993. "A Novel approach to Non-linear and Non-Gaussian Bayesian State Estimation," *IEEE Proceedings F*, 140, 107–13.
- Gourieroux, C., and A. Monfort, 1996. *Simulation-Based Econometric Methods*, Oxford: Oxford University Press.
- Gourieroux, C., A. Monfort, and E. Renault, 1993. "Indirect Inference," *Journal of Applied Econometrics*, 8, S85–S118.
- Granger, C.W.J., 1969. "Investigating Causal Relations by Econometric Models and Cross-Spectral Methods," *Econometrica*, 37, 424–38.
- Greene, W. H., 2003. *Econometric Analysis*, 5th ed., Upper Saddle River, NJ: Prentice Hall.
- Greenwood, J., and Z. Hercowitz, 1991. "The Allocation of Capital and Time over the Business Cycle," *Journal of Political Economy*, 99, 1188–1214.
- Gul, F., and W. Pesendorfer, 2004. "Self-Control, Revealed Preference, and Consumption Choice," *Review of Economic Dynamics*, 7, 243–64.
- Haavelmo, T., 1944. "The Probability Approach in Econometrics," *Econometrica*, 12 (supplement), iii–vi and 1–115.
- Hamilton, J. D., 1989. "A New Approach to the Economic Analysis of Nonstationary Time Series and the Business Cycle," *Econometrica*, 57, 357–84.
- , 1994. *Time Series Analysis*, Princeton: Princeton University Press.
- Handschin, J., 1970. "Monte Carlo Techniques for Prediction and Filtering of Non-linear Stochastic Processes," *Automatica*, 6, 555–63.

- Handschin, J., and D. Mayne, 1969. "Monte Carlo Techniques to Estimate the Conditional Expectation in Multi-stage Non-linear Filtering," *International Journal of Control*, 9, 547–59.
- Hannan, E. J., 1970. *Multiple Time Series*, New York: Wiley.
- Hansen, B. E., 1997. "Approximate Asymptotic p-values for Structural Change Tests," *Journal of Business and Economic Statistics*, 15, 60–67.
- Hansen, B. E., and K. West, 2002. "Generalized Method of Moments and Macroeconomics," *Journal of Business & Economic Statistics*, 20, 460–69.
- Hansen, G. D., 1985. "Indivisible Labor and the Business Cycle," *Journal of Monetary Economics*, 16, 309–27.
- Hansen, L. P., 1982. "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50, 1029–54.
- Hansen, L. P., and J. J. Heckman, 1996. "The Empirical Foundations of Calibration," *Journal of Economic Perspectives*, 10, 87–104.
- 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.
- , 1981. "Exact Linear Rational Expectations Models: Specification and Estimation," *Federal Reserve Bank of Minneapolis Staff Report*, 71.
- , 2005. "Recursive Models of Dynamic Linear Economies," manuscript, New York University.
- , 2008. *Robustness*, Princeton: Princeton University Press.
- Hansen, L. P., and K. J. Singleton, 1982. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models," *Econometrica*, 50, 1269–86.
- , 1983. "Stochastic Consumption, Risk Aversion, and the Temporal Behavior of Asset Prices," *Journal of Political Economy*, 91, 249–65.
- Harvey, A. C., 1993. *Time Series Models*, Cambridge: MIT Press.
- , 1999. *The Econometric Analysis of Time Series*, Cambridge: MIT Press.
- Harvey, A. C., and A. Jaeger, 1993. "D detrending, Stylized Facts and the Business Cycle," *Journal of Econometrics*, 8, 231–47.
- Hastings, W. K., 1970. "Monte Carlo Sampling Methods Using Markov Chains and Their Application," *Biometrika*, 57, 97–109.
- Heaton, J., 1985. "An Empirical Investigation of Asset Pricing with Temporally Dependent Preference Specifications," *Econometrica*, 63, 681–717.
- Heer, B., and A. Maussner, 2005. *Dynamic General Equilibrium Modelling: Computational Methods and Applications*, Berlin: Springer.
- Heiss, F., and V. Winschel, 2008. "Likelihood Approximation by Numerical Integration on Sparse Grids," *Journal of Econometrics*, 144, 62–80.
- Houthakker, H., 1956. "The Pareto Distribution and the Cobb-Douglas Production Function," *Review of Economic Studies*, 23, 27–31.
- Hugget, M., and S. Ospina, 2001. "Aggregate Precautionary Savings: When Is the Third Derivative Irrelevant?" *Journal of Monetary Economics*, 48, 373–96.
- Ingram, B. F., N. R. Kocherlakota, and N. E. Savin, 1994. "Explaining Business Cycles: A Multiple-Shock Approach," *Journal of Monetary Economics*, 34, 415–28.

- 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.
- Ireland, P., 2004a. "Technology Shocks in the New Keynesian Model," *Review of Economics and Statistics*, 86, 923–36.
- , 2004b. "A Method for Taking Models to Data," *Journal of Economic Dynamics and Control*, 28, 1205–26.
- Jeffreys, H., 1961. *Theory of Probability*, 3rd ed., Oxford: Clarendon Press.
- Judd, K. L., 1992. "Projection Methods for Solving Aggregate Growth Models," *Journal of Economic Theory*, 58, 410–52.
- , 1998. *Numerical Methods in Economics*, Cambridge: MIT Press.
- Judd, K. L., and S. M. Guu, 1997. "Asymptotic Methods for Aggregate Growth Models," *Journal of Economic Dynamics and Control*, 21, 1025–42.
- Judge, G. G., W. E. Griffiths, R. Carter Hill, H. Lutkepohl, and T.-C. Lee, 1985. *The Theory and Practice of Econometrics*, New York: Wiley.
- Kaiser, R., and A. Maravall, 2001. *Measuring Business Cycles in Economic Time Series*, Berlin: Springer-Verlag.
- King, R. G., C. I. Plosser, and S. T. Rebelo, 1988. "Production, Growth and Business Cycles. I. The Basic Neoclassical Model," *Journal of Monetary Economics*, 21, 195–232.
- King, R. G., and M. W. Watson, 2002. "System Reduction and Solution Algorithms for Solving Linear Difference Systems under Rational Expectations," *Computational Economics*, 20, 57–86.
- Kitagawa, G., 1996. "Monte Carlo Filter and Smoother for Non-Gaussian Non-linear State-Space Models," *Journal of Computational and Graphical Statistics*, 5, 1–25.
- Klein, P., 2000. "Using the Generalized Schur Form to Solve a Multivariate Linear Rational Expectations Model," *Journal of Economic Dynamics and Control*, 24, 1405–23.
- Kloek, T., and H. K. van Dijk, 1978. "Bayesian Estimates of Equation System Parameters: An Application of Integration by Monte Carlo," *Econometrica*, 46, 1–19.
- Kocherlakota, N. R., 1990. "On Tests of Representative Consumer Asset Pricing Models," *Journal of Monetary Economics*, 26, 285–304.
- , 1996. "The Equity Premium: It's Still a Puzzle," *Journal of Economic Literature*, 34, 42–71.
- Koop, G., 2003. *Bayesian Econometrics*, New York: Wiley.
- Koop, G., and S. M. Potter, 2009. "Time Varying VARs with Inequality Restrictions," working paper, Federal Reserve Bank of New York.
- Koopmans, T. C., 1949. "The Econometric Approach to Business Fluctuations," *American Economic Review*, 39, 64–72.
- Kose, A., C. Otrok, and C. H. Whiteman, 2003. "International Business Cycles: World, Region, and Country-Specific Factors," *American Economic Review*, 93, 1216–39.
- Krusell, P., and A. A. Smith Jr., 1999. "On the Welfare Effects of Eliminating Business Cycles," *Review of Economic Dynamics*, 2, 245–72.

- Kydland, F. E., and E. C. Prescott, 1982. "Time to Build and Aggregate Fluctuations," *Econometrica*, 50, 1345–70.
- , 1991a. "The Econometrics of the General Equilibrium Approach to Business Cycles," *Scandinavian Journal of Economics*, 93, 161–78.
- , 1991b. "Hours and Employment Variation in Business Cycle Theory," *Economic Theory*, 1, 63–81.
- , 1996. "The Computational Experiment: An Econometric Tool," *Journal of Economic Perspectives*, 10, 69–85.
- Lagarias, J. C., J. A. Reeds, M. H. Wright, and P. E. Wright, 1998. "Convergence Properties of the Nelder-Mead Simplex Method in Low Dimensions," *SIAM Journal of Optimization*, 9, 112–47.
- Lancaster, T., 2004. *An Introduction to Modern Bayesian Econometrics*, Oxford: Blackwell.
- Leamer, E. E., 1978. *Specification Searches*, New York: Wiley.
- Lee, B. S., and B. F. Ingram, 1991. "Simulation Estimation of Time-Series Models," *Journal of Econometrics*, 47, 197–205.
- Lehmann, E. L., 1986. *Testing Statistical Hypotheses*, New York: John Wiley & Sons.
- Liesenfeld, R. and J.-F. Richard, 2010. "Improving MCMC Using Efficient Importance Sampling," *Journal of Computational Statistics and Data Analysis*, 53, 272–288.
- LeRoy, S. F., and R. D. Porter, 1981. "Stock Price Volatility: Tests Based on Implied Variance Bounds," *Econometrica*, 49, 97–113.
- Liu, J., and R. Chen, 1995. "Blind Deconvolution via Sequential Imputations," *Journal of the Royal Statistical Society (Series B)*, 430, 567–76.
- Ljungqvist, L., and T. Sargent, 2004. *Recursive Macroeconomic Theory*, Cambridge: MIT Press.
- Lucas, R. E., 1976. "Econometric Policy Evaluation: A Critique," in K. Brunner and A. Meltzer, eds., *The Phillips Curve and Labor Markets*, Amsterdam: North-Holland.
- , 1978. "Asset Prices in an Exchange Economy," *Econometrica*, 46, 1429–45.
- , 1980. "Methods and Problems in Business Cycle Theory," *Journal of Money, Credit and Banking*, 12, 696–715.
- , 1987. *Models of Business Cycles*, Oxford: Blackwell.
- , 2003. "Macroeconomic Priorities," *American Economic Review*, 93, 1–14.
- Lucas, R. E., and T. J. Sargent, 1979. "After Keynesian Macroeconomics," *Federal Reserve Bank of Minneapolis Quarterly Review*, vol. 3, no. 2.
- Lundberg, E., 1969. Nobel Prize in Economics Presentation Speech, <http://nobelprize.org/economics/laureates/1969/press.html>.
- Mankiw, N. G., and D. Romer, 1991. *New Keynesian Economics*, Cambridge: MIT Press.
- McFadden, D., 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57, 995–1026.

- McGrattan, E. R., 1994. "The Macroeconomic Effects of Distortionary Taxation," *Journal of Monetary Economics*, 34, 573–601.
- , 1996. "Solving the Stochastic Growth Model with a Finite Element Method," *Journal of Economic Dynamics and Control*, 20, 19–42.
- , 1999. "Application of Weighted Residuals Methods to Dynamic Economic Models," in R. Marimon and A. Scott, eds., *Computational Methods for the Study of Dynamic Economies*, Oxford: Oxford University Press.
- McGrattan, E. R., R. Rogerson, and R. Wright, 1997. "An Equilibrium Model of the Business Cycle with Household Production and Fiscal Policy," *International Economic Review*, 38, 267–90.
- Mehra, R., and E. C. Prescott, 1985. "The Equity Premium: A Puzzle," *Journal of Monetary Economics*, 15, 145–61.
- , 2003. "The Equity Premium in Retrospect," in G. M. Constantinides, M. Harris, and R. Stulz, eds., *Handbook of the Economics of Finance*, Amsterdam: Elsevier.
- Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953. "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087–92.
- Metropolis, N., and Ulam, S. 1949. "The Monte Carlo Method," *Journal of the American Statistical Association*, 44, 335–41.
- Murray, C. J., 2003. "Cyclical Properties of Baxter-King Filtered Time Series," *Review of Economics and Statistics*, 85, 472–76.
- Nelder, J. A., and R. Mead, 1965. "A Simplex Method for Function Minimization," *Computer Journal*, 7, 308–13.
- Nelson, C., and H. Kang, 1981. "Spurious Periodicity in Inappropriately Detrended Time Series," *Econometrica*, 49, 741–51.
- Nelson, C., and C. I. Plosser, 1982. "Trends and Random Walks in Macroeconomic Time Series," *Journal of Monetary Economics*, 10, 139–62.
- Nelson, C., and R. Startz, 1990a. "Some Further Results on the Exact Small Sample Properties of the Instrumental Variables Estimator," *Econometrica*, 58, 967–76.
- , 1990b. "The Distribution of the Instrumental Variables Estimator and Its *t*-ratio when the Instrument Is a Poor One," *Journal of Business*, 63, 5125–40.
- Newey, W., and K. West, 1987a. "A Simple Positive Semi-definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix," *Econometrica*, 55, 703–8.
- , 1987b. "Hypothesis Testing with Efficient Method of Moments Estimation," *International Economic Review*, 28, 777–87.
- Nordhaus, W., 2004. "Retrospective on the 1970s Productivity Slowdown," Working Paper No. 10950, National Bureau of Economic Research, Cambridge, MA.
- Otrok, C., 2001. "On Measuring the Welfare Cost of Business Cycles," *Journal of Monetary Economics*, 47, 61–92.
- Pakes, A., and D. Pollard, 1989. "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57, 1027–57.
- Palka, B., 1991. *An Introduction to Complex Function Theory*, Berlin: Springer.

- Pearson, K., 1894. "Contribution to the Mathematical Theory of Evolution," *Philosophical Transactions of the Royal Society of London*, Series A 185, 71–110.
- Perli, R., and P. Sakellaris, 1998. "Human Capital Formation and Business Cycle Persistence," *Journal of Monetary Economics*, 42, 67–92.
- Perron, P., 1989. "The Great Crash, the Oil Shock, and the Unit Root Hypothesis," *Econometrica*, 57, 1361–1401.
- Pesaran, M. H., and S. M. Potter, 1997. "A Floor and Ceiling Model of US Output," *Journal of Economic Dynamics and Control*, 21, 661–95.
- Phillips, P.C.B., 1991. "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics*, 6, 333–64.
- Pitt, M. K., 2002. "Smooth Particle Filters for Likelihood Evaluation and Maximisation," working paper, University of Warwick.
- Pitt, M. K., and N. Shephard, 1999. "Filtering via Simulation: Auxiliary Particle Filters," *Journal of the American Statistical Association*, 94, 590–99.
- Poirier, D. J., 1995. *Intermediate Statistics and Econometrics: A Comparative Approach*, Cambridge: MIT Press.
- Prescott, E. C., 1986. "Theory Ahead of Business Cycle Measurement," *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 9–22.
- , 2006. "Nobel Lecture: The Transformation of Macroeconomic Policy and Research," *Journal of Political Economy*, 114, 203–35.
- Primiceri, G., 2005. "Time Varying Structural Vector Autoregressions and Monetary Policy," *Review of Economic Studies*, 72, 821–52.
- Ramsey, F. K., 1928. "A Mathematical Theory of Saving," *Economic Journal*, 38, 543–59.
- Richard, J.-F., and W. Zhang, 1997. "Accelerated Monte Carlo Integration: An Application to Dynamic Latent Variable Models," in R. Mariano, M. Weeks, and T. Schuermann, eds., *Simulation Based Inference in Econometrics: Methods and Applications*, Cambridge: Cambridge University Press.
- , 2007. "Efficient Importance Sampling," *Journal of Econometrics*, 141, 1385–1411.
- Ristic, B., S. Arulampalam, and N. Gordon, 2004. *Beyond the Kalman Filter: Particle Filters for Tracking Applications*, Boston: Artech House.
- Robert, C. P., and G. Casella, 1999. *Monte Carlo Statistical Methods*, New York: Springer-Verlag.
- Rogerson, R., 1988. "Indivisible Labor, Lotteries and Equilibrium," *Journal of Monetary Economics*, 21, 3–16.
- Romer, D., 2006. *Advanced Macroeconomics*, 3rd ed., Boston: McGraw-Hill/Irwin.
- Rowenhorst, G., 1991. "Time to Build and Aggregate Fluctuations: A Reconsideration," *Journal of Monetary Economics*, 27, 241–54.
- Rubin, D. B., 1987. "A Noniterative Sampling/Importance Resampling Alternative to the Data Augmentation Algorithm for Creating a Few Imputations When the Fraction of Missing Information Is Modest: The SIR Algorithm (Discussion of Tanner and Wong)," *Journal of the American Statistical Association*, 82, 543–46.
- Rubinstein, R. Y., 1981. *Simulation and the Monte Carlo Method*, New York: Wiley.

## References

- Ruge-Murcia, F. J., 2003. "Methods to Estimate Dynamic Stochastic General Equilibrium Models," manuscript, University of Montreal.
- Santos, M. S., and A. Peralta-Alva, 2005. "Accuracy of Simulations for Stochastic Dynamic Models," *Econometrica*, 73, 1939–76.
- Santos, M. S., and J. Vigo-Aguiar, 1998. "Analysis of a Numerical Dynamic Programming Algorithm Applied to Economic Models," *Econometrica*, 66, 409–26.
- Sargent, T. J., 1987a. *Macroeconomic Theory*, 2nd ed., London: Academic Press.
- , 1987b. *Dynamic Macroeconomic Theory*, Cambridge: Harvard University Press.
- , 1989. "Two Models of Measurements and the Investment Accelerator," *Journal of Political Economy*, 97, 251–87.
- Sargent, T. J., and C. A. Sims, 1997. "Business Cycle Modeling without Pretending to Have Too Much A-Priori Economic Theory," in C. A. Sims et al., eds., *New Methods in Business Cycle Research*, Minneapolis: Federal Reserve Bank of Minneapolis.
- Scarf, H., and T. Hansen, 1973. *The Computation of Economic Equilibria*, New Haven: Yale University Press.
- Schmitt-Grohé S., and M. Uribe, 2004. "Solving Dynamic General Equilibrium Models Using a Second-Order Approximation of the Policy Function," *Journal of Economic Dynamics and Control*, 28, 755–75.
- Schorfheide, F., 2000. "Loss Function-Based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15, 645–70.
- Schotman, P. C., and H. K. van Dijk, 1991. "On Bayesian Routes to Unit Roots," *Journal of Applied Econometrics*, 6, 387–401.
- Schwarz, G., 1978. "Estimating the Dimension of a Model," *Annals of Statistics*, 6, 461–64.
- Sedaghat, H., 2003. *Nonlinear Difference Equations: Theory with Applications to Social Science Models*, Berlin: Springer.
- Shiller, R. J., 1981. "Do Stock Prices Move Too Much to be Justified by Subsequent Changes in Dividends?" *American Economic Review*, 71, 421–36.
- , 1989. *Market Volatility*, Cambridge: MIT Press.
- Shiryayev, A. N., 1995. *Probability*, Berlin: Springer Graduate Texts in Mathematics.
- Shoven, J. B., and J. Whalley, 1972. "A General Equilibrium Calculation of the Effects of Differential Taxation of Income from Capital in the U.S.," *Journal of Public Economics*, 1, 281–322.
- Silverman, B. W., 1986. *Density Estimation*, London: Chapman and Hall.
- Sims, C. A., 1972. "Money, Income, and Causality," *American Economic Review*, 62, 540–52.
- , 1980. "Macroeconomics and Reality," *Econometrica*, 48, 1–48.
- , 1996. "Macroeconomics and Methodology," *Journal of Economic Perspectives*, 10, 105–20.
- , 2001. "Solving Linear Rational Expectations Models," *Computational Economics*, 20, 1–20.
- Sims, C. A., and T. Zha, 1998. "Bayesian Methods for Dynamic Multivariate Models," *International Economic Review*, 39, 949–68.

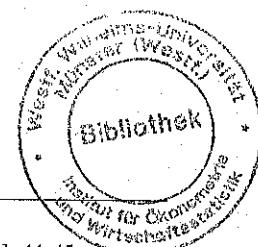
## References

- , 2006. "Vintage Article: Does Monetary Policy Generate Recessions?" *Macroeconomic Dynamics*, 10, 231–72.
- Smets, F., and R. Wouters, 2003. "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1, 1123–75.
- , 2007. "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review* 97, 586–606.
- Smith, A., 1993. "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8, S63–S84.
- Smith, J. Q., and A.A.F. Santos, and 2006. "Second-Order Filter Distribution Approximations for Financial Time Series with Extreme Outliers," *Journal of Business and Economic Statistics*, 24, 329–37.
- Solow, R. M., 1956. "A Contribution to the Theory of Economic Growth," *Quarterly Journal of Economics*, 70, 65–94.
- Stock, J. H., 2004. "Unit Roots, Structural Breaks, and Trends," in R. F. Engle and D. L. McFadden, eds. *Handbook of Econometrics*, vol. 4, Amsterdam: North Holland.
- Stock, J. H., and M. Watson, 2002a. "Forecasting Using Principal Components from a Large Number of Predictors," *Journal of the American Statistical Association*, 97, 1167–79.
- , 2002b. "Macroeconomic Forecasting Using Diffusion Indexes," *Journal of Business and Economic Statistics*, 20, 147–62.
- , 2006. "Forecasting with Many Predictors," in G. Elliott, C.W.J. Granger, and A. Timmermann, eds., *Handbook of Economic Forecasting*, 515–54. Amsterdam: Elsevier.
- , 2007. "Why Has U.S. Inflation Become Harder to Forecast?" *Journal of Money, Credit and Banking*, 39, supplement, 3–33.
- , 2010. "Dynamic Factor Models," in M. P. Clemens and D. F. Hendry, eds, *Oxford Handbook of Economic Forecasting*, Oxford: Oxford University Press.
- Stock, J. H., J. H. Wright, and M. Yogo, 2002. "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business and Economic Statistics*, 20, 518–29.
- Stokey, N. L., and R. E. Lucas, 1989. *Recursive Methods in Economic Dynamics*, Cambridge: Harvard University Press.
- Summers L., 1986. "Some Skeptical Observations on Real Business Cycle Theory," *Federal Reserve Bank of Minneapolis Quarterly Review*, 10, 23–28.
- Sustek, R., 2005. "Essays on Aggregate Fluctuations," doctoral Dissertation, Carnegie-Mellon University.
- Tauchen, G., 1986a. "Finite State Markov-Chain Approximations to Univariate and Vector Autoregressions," *Economics Letters*, 20, 177–81.
- , 1986b. "Statistical Properties of Generalized Method-of-Moments Estimators of Structural Parameters Obtained from Financial Market Data," *Journal of Business and Economic Statistics*, 4, 397–416.
- Taylor, S. J., 1986, *Modeling Financial Time Series*, Chichester: John Wiley and Sons.
- Theil, H., 1971. *Principles of Econometrics*, New York: Wiley.

- Theil, H., and A. S. Goldberger, 1961. "On Pure and Mixed Statistical Estimation in Economics," *International Economic Review*, 2, 65–78.
- Uhlig, H., 1999. "A Toolkit for Analyzing Non-linear Dynamic Stochastic Models Easily," in Ramon Marimon and Andrew Scott, eds., *Computational Methods for the Study of Dynamic Economies*, Oxford University Press, New York.
- Watson, M., 1993. "Measures of Fit for Calibrated Models," *Journal of Political Economy*, 101, 1011–41.
- , 1994. "Vector Autoregressions and Cointegration," in R. F. Engle and D. L. McFadden, eds., *Handbook of Econometrics*, vol. 4, Amsterdam: Elsevier.
- Weil, P., 1989. "The Equity Premium Puzzle and the Risk-Free Rate Puzzle," *Journal of Monetary Economics*, 24, 401–21.
- Wen, Y., 1998. "Can A Real Business Cycle Model Pass the Watson Test?" *Journal of Monetary Economics*, 42, 185–203.
- , 2002. "The Business Cycle Effects of Christmas," *Journal of Monetary Economics*, 49, 1289–1314.
- Whiteman, C. H., 1983. *Linear Rational Expectations Models: A User's Guide*, Minneapolis: University of Minnesota Press.
- Wickens, M., 2008. *Macroeconomic Theory: A Dynamic General Equilibrium Approach*, Princeton: Princeton University Press.
- Woitek, U., 1998. "A Note on the Baxter-King Filter," working paper, University of Glasgow.
- Woodford, M., 2003. *Interest and Prices: Foundations of a Theory of Monetary Policy*, Princeton: Princeton University Press.
- Yu, B., and P. Mykland, 1994. "Looking at Markov Samplers through CUSUM Path Plots: A Simple Diagnostic Idea," Technical Report 413, Department of Statistics, University of California at Berkeley.
- Zellner, A., 1971. *An Introduction to Bayesian Inference in Econometrics*, New York: Wiley.
- Zaritskii, V., V. Svetnik, and L. Shimelevich, 1975. "Monte-Carlo Techniques in Problems of Optimal Data Processing," *Automavion Remote Control*, 12, 2015–2022.

## Index

- Adda, J., 69, 88, 286  
 aggregate uncertainty, 28  
 Akaike Information Criterion (AIC), 158–159  
 Akashi, H., 230  
 algorithms: derivative-based algorithms, 328–330; Howard improvement algorithm, 95; Metropolis-Hastings Algorithm, 218–220, 363; Nelder-Mead Algorithm, 320, 323, 326; optimization algorithms, 318–330; period- $t$  algorithm, 225–226, 228–229, 231, 235; random-walk algorithms, 220; sampling importance resampling algorithm, 222, 226; sequential algorithm, 175–178; sequential importance sampling algorithm, 221–222, 224–228; simplex algorithms, 319–328  
 Altug, S., 318  
 Amemiya, T., 315  
 Andrews, D.W.K., 290, 291, 336  
 An, S., 215  
 AR(1) process: ARMA model, 142; normalized shock and, 35; optimal growth model, 80, 82, 84–86; VAR model, 148  
 arbitrary order, 28  
 Arouba, S. B., 69, 72  
 AR(p) specification: ARMA model, 142, 144; VAR model, 147–148  
 Arulampalam, S., 166, 221, 222  
 asset pricing model: additional resources, 20, 38; alternative versions, 38–47; Bayesian methods, 373–386; CRRA and, 41, 43, 46, 107, 274–275, 374–386; empirical analysis, 373–386; equity premium puzzle, 20, 40, 273–275; expectational errors, 56; GAUSS programming, 47, 287n1; habit/durability, 41–43; households, 38–40; introduction, 19–20; with linear utility, 55–56; multi-asset environment, 39–40; optimality condition, 44, 47; perturbation methods, 105–109; policy functions, 170; preference specifications, alternative, 40–47; risk-free rate puzzle, 20, 38, 40, 274–275; self-control preferences, 44–47; shocks, 41, 45; single-asset environment, 38–39  
 utility function, 40–41, 44–45;  
 volatility, 19, 38, 40, 46; von Neumann-Morgenstern utility function, 44  
 autocovariance-generating function (ACGF), 280, 281  
 autoregressive conditionally heteroskedastic (ARCH) process, 145, 147, 149  
 autoregressive-moving average (ARMA) model: ARCH/GARCH specifications, 145–147; GAUSS programming, 145n1; heteroskedastic innovations, 145–147; OLS regression, 144–145; reduced-form analysis, 139–145; square-summability violation, 141; state-space representations, 167–168; summary statistics, 149–152; time series behavior, 139–145; Wold Decomposition Theorem, 140  
 Auxiliary Particle Filter (APF), 231–232  
 Azariadis, C., 13, 51  
 Bai, J., 187  
 band-pass filters: cycle isolation, 131–134; GAUSS programming, 132n4; spuriousness arising from application of, 135–136; summary statistics/time series, 149–153  
 Barillas, P., 92  
 Barro, R. J., 9, 19, 263  
 Bauwens, L., 195, 217–218, 237, 352  
 Baxter, M., 132–134  
 Bayes factor, 362  
 Bayesian Information Criterion (BIC), 159  
 Bayesian methods: asset pricing models, estimating and comparing, 373–386; calibration and, 264–267, 278; Cholesky decomposition, 357–358; CRRA and, 264–267, 374–386; empirical analysis, 373–386; forecasting application, 364–373; GAUSS programming, 358n1; law of large numbers, 355; likelihood function, 351, 352–353; MCMC method, 355–360, 363; Metropolis-Hastings Algorithm, 218–220, 363; model comparisons, 352, 361–364, 373–386; Monte Carlo methods, 162, 355–360; Normal/Inverted-Gamma distributions, 368, 369; Normal/Inverted-Wishard



Bayesian methods (*cont.*)  
 distribution, 357–358, 367; notation, 351–353; objectives overview, 351–352; OLS regression, 367–369; parameter estimation, 381–386; plausibility of alternatives, 351; posterior distribution, 351, 354, 358–364; posterior odds analysis, 352; preliminaries, 352–355; prior information incorporation, 351, 355–360, 364–373; RBC model, 264–267; reduced-form analysis, 355–360; structural models as information sources, 355–360; structural models implementation, 360–361; trend- vs. difference-stationarity issue, 135–136; VAR model, 384–386

Bayesian vs. classical approaches to likelihood analysis, 314–315

Bayes' Rule, 315, 354, 362, 367

Bayes' Theorem, 194, 224, 225

Becker, G. S., 260, 270

behavioral equations, 254–255

Bell, W. R., 130

Bellman's equation, 88–89, 91, 92, 94

Benassy, J., 19

Benhabib, J., 272

Berger, J. O., 314

Berkowitz, J., 302

Bernake, S. S., 187

BFGS method, 329, 331

Blake, A., 226

Blanchard, O. J., 20, 22, 57–60

Blanchard-Kahn method: expectations errors, 26; explosive/nonexplosive components, decoupling, 59–60; Klein's method compared, 64–65; Sims' method compared, 61–62

Bobrovsky, B. Z., 225

Boivin, J., 187

Boldrin, M., 306

Bollerslev, T., 145, 147

bootstrap particle filter, 222, 226, 227, 320

break date identification, 336–337

Brooks, C. S., 147

Brown, J. W., 120

Burke, Edmund, 18, 147

burn-in phase, 199–200, 217, 246

Burnside, C., 272

business cycle fluctuations: calibration and, 268–273; maximum likelihood and, 337–350; monopolistic competition model, 19;

## Index

RBC model, 19, 21; seasonal adjustment, 114, 129–131; shocks and, 19, 256, 268–273, 337–350; sources of, 19; welfare cost of, 261–268

calibration: Bayesian methods, 264–267, 278; behavioral equations, 254; business cycle fluctuations and, 268–273; Cholesky decomposition, 281, 284; Cobb-Douglas production function, 278; consumption and, 264–267; critiques, 276–279; CRRA and, 264–273; DSGE model, 257–261, 276; equity premium puzzle, 273–275, 277; estimation vs., 260, 278; extensions, 279–284; filters and, 271–272; general-equilibrium models, 278; goal of, 256; habit/durability stocks and, 264–267; Hansen and, 256, 278–279; historical perspective, 253–258; Hodrick-Prescott filter, 271; identities, 253–254; implementation, 258–261; institutional rules, 254; Kydland and, 253–261, 272, 276–278; Lucas and, 254–255, 257, 261–264, 267; matching moments, 261; model-classification, 254; model parameters, 259–260; National Income and Product Accounts, 270; nation income accounting identity, 254; Ostrom and, 264, 266–267, 283; philosophy, 253–258; Prescott and, 253–262, 272–278; production function, 254, 271–272; productivity shocks, 268–273; RBC model, 257, 259, 264–273, 281–282; Sargent and, 254–256; spectra, 281–282; technology constraints, 254; utility function, 264–267; VAR model, 283–284; volatility and, 271–272; welfare cost of business cycles, 261–268

Campbell, J. Y., 20, 38

ANOVA, F., 114, 283

capital, law of motion for, 25, 271

Cappé, O., 221

Carroll, C. C., 92

Casella, C., 195, 216, 352

CES production function, 30–31

Chan, K., 136

change-of-variables formula, 84

Chari, V. V., 19, 306, 338

Chebyshev interpolation theorem, 77, 78

Chebyshev polynomial: extension to the *l*-dimensional case, 78, 80; optimal growth

## Index

model, 84–85; projection methods, 73–74, 76–78, 80

Chen, R., 226

Chib, S., 218

Cholesky decomposition: Bayesian methods, 357–358; calibration, 281, 284; GAUSS programming, 281n2; importance sampling, 209; lower triangular, 209, 357; matching moments, 304; summary statistics, 156–157

Chou, R. Y., 147

Christiano, L. J., 19, 68, 132, 301, 306

Churchill, R. V., 120

Clarida, R., 31

classical approach to estimation, 314–315. *See also* maximum likelihood

Cobb-Douglas production function, 16, 22, 278

Cochrane, J. H., 20, 38

Cogley, T., 136, 182

collocation method, 76–78

competition model, 28–37

complex variables, 121–122

conditionally optimal filters: continuation, 229; filtering, 229; likelihood integral, 230; propagation, 229; resampling, 229; SMC methods, 228–232

conditionally optimal importance sampler, 228–229, 232

conditional probability, defined, 353

conditional probability calculation, 353

constant elasticity of substitution (CES) function, 29–30

Constantinides, G. M., 41

constant relative risk aversion (CRRA): asset pricing model, 41, 43, 46, 107, 274–275, 374–386; Bayesian methods, 374–386; calibration and, 264–273; instantaneous utility function, 22; matching moments, 301; RBC model, 22, 264–273

consumption: asset pricing model, 38–40; band-pass filtered, 134–135; calibration and, 264–267; demand shocks and, 29; H-P filtered, 131–132, 271–272; monopolistic competition model, 28–29; nonseasonally adjusted, 130, 132; RBC model, 20; seasonally adjusted, 130, 132; trend removal, 115, 118–120; volatility and, 271–272; welfare cost, 261–268

consumption/savings decision, 20–21

consumption-smoothing, 23

continuation, unadapted filters, 225

contraction mapping, 89, 91, 95

control variables, 10, 12

convergence: CUSUM statistic, 217–218; derivative-based algorithms, 330; fixed-point, 212–213; guaranteeing, 213; linear approximation methods, 51; nonlinear approximation methods, 92; null hypothesis test, 217; policy-function iterations, 92, 94–95; simplex algorithms, 327; value-function iterations, 92, 94

Cooley, T. F.: calibration exercise, 258; data correspondence, 113; RBC model analysis, 19, 260, 313; time-varying parameters, 182

Cooper, R., 69, 88, 286

cost-push shocks, 30, 32, 35, 337–350

covariance-stationary stochastic processes (CSSPs): mean-zero, 115, 139, 279; trend removal and, 115–117; white noise, 135; Wold Decomposition Theorem, 140–141

Cramér Representations, 124–125

cross-equation restriction, 255–256

curse of dimensionality, 79, 89

CUSUM statistic, 217–218

cycle isolation: band-pass filters, 131–134; complex variables, 121–123; Cramér Representations, 124–125; data preparation, 120–134; data set, 114–115; DeMoivre's Theorem, 122, 126; filters for, 126–134; Fourier transforms, 123; frequency, 114, 126–128; H-P filtered, 128–129; introduction, 113–115; inversion formula, 123, 125; mathematical background, 120–123; Riesz-Fisher Theorem, 123, 125; seasonal adjustment, 130–131; spectra, 125–126; spuriousness, 134–136; time series, 124–125; trend removal and, 114

data: correspondence and, 113; cyclical characteristics of, 114; seasonal adjustment of, 130–131; spuriousness, 134–136

data generation process (DGP), 160

decomposition: Cholesky, 156–157, 209, 281, 284, 304, 357–358; Jordan, 59, 61; OZ, 61, 65; Schur, 65; variance decompositions, 337, 340

decoupling: Blanchard-Kahn method, 59–60; Klein's method, 64; Sims' method, 61

de Freitas, N., 166, 222  
 degeneracy, 221, 225, 226, 235–236  
 Del Moral, P., 226  
 Del Negro, M., 359, 365  
 Delta method, 160  
 demand shocks: business cycle fluctuations and, 337–350; consumption and, 29; monopolistic competition model, 31–39  
 DeMoivre's Theorem, 122, 126  
 denHann, Wouter, 71n2  
 densities, discrete approximations of, 223  
 derivative-based algorithms: BFGS alternative, 329; convergence, 329, 330; DFP alternative, 329; GAUSS programming, 328n2; Newton's Method, 328; overshooting, 329; quasi-Newton methods, 329  
 detrending approach to trend removal, 116–119  
 DFP alternative, 329  
 Diebold, F. X., 302  
 difference-stationary, 136  
 differencing approach to trend removal, 116–119  
 Dirac measures, 223, 225, 232  
 direct Monte Carlo integration: applications, 198; model simulation, 198–201; numerical standard error, 196–197, 203–205; optimal growth model, 200–201; posterior inference via, 201–202; relative numerical efficiency, 205–207  
 discontinuities problem, 330–331  
 Doan, T., 365  
 Doucet, A., 166, 222  
 Duffie, D., 294  
 Durbin, J., 234  
 dynamic factor models (DFMs), 187–190  
 dynamic programming, 87–89, 88n6  
 dynamic stochastic general equilibrium (DSGE) models: asset pricing model, 38–47; background, 3–4; calibration and, 257–261, 276; cross-equation restriction, 3, 255–256; EIS filter implementation, 241–250; empirical analysis, 3–4, 19, 138, 257–258; example equations, 15–16; expectational difference equations, 9–10; first-order conditions, 9; generic representation, 18; iteration techniques, 87–95; linearization, 12–17; log-linear approximations, 14–15; log-linearization, 34–37; log-linear representations, 11; matching

moments, 300–308; model simulation, 198; monopolistic competition model, 28–37; Monte Carlo methods, 241–250; new empirical paradigm for, 3; nonlinear representations, 11–12; nonlinear system, 23–26, 33–34; notation, 9–12, 95–96; overview (text outline), 4–6; perturbation methods, 95–109; RBC model, 20–27; Sims' method, 61–64; simulations, 306–308; state-space representations, 169–171, 241–250; systems-of-equations models vs., 254  
 DYNARE website, 8  
*Econometrica*, 256  
 econometrics: Frisch's vision of, 256; probability approach to, 253, 255, 256–257, 276  
 Econometric Society, 256  
 efficiency: SIR algorithm, 226; SIS algorithm, 226  
 efficient importance sampling (EIS): continuation, 235; degenerate transitions, 235–236; DSGE model implementation, 241–250; example, 239–241; importance sampler initialization, 236–239; likelihood integral, 235; Monte Carlo integration, 211–215; optimization, 235; optimized sampler construction, 234; propagation, 235; RBC model implementation, 246–250; SMC methods, 233–241; unconditional optimality, 222, 233, 234–235  
 efficient importance sampling (EIS) filter, 222, 233–241  
 Eichenbaum, M.: impulse response functions, 302; matching moments, 301, 306; monopolistic competition model, 19; reduced-form analysis, 256; volatility of hours, 272  
 eigenvalues: explosive/nonexplosive, 53, 59–61, 65; generalized, 61–62, 65, 67  
 Einarsson, T., 272  
 elasticities: asset pricing model, 41, 275; log-linear approximations, 15; monopolistic competition model, 29–32; RBC model, 22  
 Eliasz, P., 187  
 empirical analysis: asset pricing model, 373–386; business cycle fluctuations, 268–273, 337–350; derivative-based

algorithms, 328–330; equity premium puzzle, 273–276; Euler equations, 300–301; general-equilibrium models, 278; Generalized Method of Moments, 286–294; identified VARs, 306; ill-behaved surfaces, 330–334; indirect inference methodology, 297–300; linearized DSGE model, 301–307; matching moments, 300–313; matching RBC moments application, 308–313; maximum likelihood, 337–350; model diagnostics, 334–337; monopolistic competition model, 337–350; nonlinear approximation simulations, 307–308; optimization algorithms, 318–330; parameter stability, 334–337; productivity shocks, 268–273; RBC model, 19, 308–313; reduced-form characterizations, 3; simplex algorithms, 319–328; Simulated Method of Moments, 294–297; spuriousness, 134–136; welfare cost of business cycles, 261–268  
 endogenous grid methods (EGMs), 92–94  
 endowment shock, 45–46  
 environment: asset pricing model, 38–40; monopolistic competition model, 28–32; RBC model, 20–23  
 equations: behavioral equations, 254–255; Bellman's equation, 88–89, 91, 92, 94; equilibrium pricing, 41; Euler, 26, 300–301; expectational difference, 9–10; Gibbs sampler, 216–219; log-linear model representations, 11; nonlinear model representations, 11–12; state-transition equation, 167, 170, 187, 235; systems-of-equations models, 3, 253–255, 258.  
*See also* theorems  
 equilibrium: equilibrium pricing equation, 41; general-equilibrium models, 257, 278; RBC model, 272; saddle-path stability, 51, 53, 54–56, 59, 65  
 equilibrium pricing equation, 41  
 equity premium puzzle, 20, 40, 273–275, 277  
 estimation: calibration vs., 260, 278; simulation-based, 307  
 Euler equations, 26, 300–301  
 Evans, C. L., 19, 302  
 exogenous demand shock, 29, 41  
 exogenous state variables, 10  
 expansion/contraction points, 322–325  
 expectational difference equations, 9–10

expectations errors: asset pricing model, 56; Blanchard-Kahn method, 26, 58; introducing, 10, 26; mapping stochastic innovations onto, 56; notation, 10; Ramsey's optimal growth model, 57; RBC model, 26; Sims' method, 60–63; and stochastic innovations, 51, 56; from structural shocks, 10; undetermined coefficients approach, 66  
 expectations notation, 9–10  
 explosive/nonexplosive components, decoupling: Blanchard-Kahn method, 59–60; Klein's method, 64; Sims' method, 61

Fair, R. C., 290  
 Fernandez-Villaverde, J.: Bayesian methods model comparison, 361; CRRA parameter notation, 250; matching moments, 304; MCMC method, 215; Monte Carlo experiment data set, 247; nonlinear solution techniques, 69; optimal growth model, 92; particle filter, 246; projection methods, 72; SMC methods, 221, 227  
 Ferson, W. E., 41  
 Feynman, Richard P., 9  
 filter gain, 128–129  
 filtering density: Auxiliary Particle Filter, 231; discrete approximations use in, 222; EIS filter, 234, 245–246; period-t, 224, 225; state-space representations, 172–175, 178  
 filters: calibration and, 271–272; conditionally optimal, 228–232; for cycle isolation, 126–134; frequency, 126–132; gain function, 128–129; phase effect, 132; seasonal adjustment, 129–131; state-space representations, 171–173, 193–194, 223–224; summary statistics, 149–157; trend removal, 119, 126; unadapted, 224–228. *See also specific filters*  
 filters, specific: Auxiliary Particle Filter, 231–232; band-pass filters, 131–136, 149–153; bootstrap particle filter, 222, 226, 227, 320; EIS filter, 222, 233–241; Hodrick-Prescott filter, 119–120, 128–132, 135–136, 149–157, 271; Kalman filter, 11, 173–182, 195, 337, 340–348  
 finite element method, 72–73, 74, 76  
 Fischer, S., 20, 22  
 Fisher, J., 306

Fitzgerald, T. J., 132  
 forecasting, Bayesian methods for, 364–373  
 Fortran code, projection methods, 71n2  
 Fourier transform: cycle isolation, 123, 125; of a symmetric filter, 132–133  
 Francis, N., 306  
 frequency: calibration and, 281–282; cycle isolation, 114, 126–128; filters, 126–132; spectra, 125–126, 151; trend removal, 114, 119  
 Friedland, B., 166  
 Frisch, R., 256–257  
 full-information analysis, 138, 314. *See also* Bayesian methods; maximum likelihood  
 gain function, 128–129  
 Galerkin method, 76  
 Gali, J., 31, 306  
 Gauss-Hermite weights/nodes, 84–85  
 GAUSS programming: additional resources, 8; ARCH/GARCH-type disturbances, 147; ARMA model, 145n1; asset pricing model, 47, 287n1; band-pass filters, 132n4; Bayesian methods, 358n1; Cholesky decompositions, 281n2; cumulative density function, 226; derivative-based algorithms, 328n2; eigenvalues of a matrix, 59n1; Galerkin method, 76n3; Generalized Method of Moments, 287n1, 289n2, 310; gradient calculation, 340; importance sampling, 209nn1–2; Kalman filter, 174n1; Klein's method, 64n6; maximum likelihood, 319n1, 328n2; measurement errors, 174n1; Monte Carlo integration, 199, 214nn1–2; nonlinear solution techniques, 71n2, 76n3, 80n4, 85; Normal/Inverted-Wishard distribution, 358n1; optimal growth model, 80n4; perturbation methods, 95n7; projection methods, 71n2, 76n3, 80n4, 85; quasi-Newton algorithm, 47; simplex algorithms, 319n1; Time Series Module, 145n1, 147; trend removal, 116n1, 117n2, 119n3; unadapted filters, 226n1; VAR model, 149n3; websites, 8  
 Gaussian quadrature approximations, 83  
 general-equilibrium models, 257, 278  
 generalized autoregressive conditionally heteroskedastic (GARCH) process, 145–147, 149

## Index

generalized eigenvalues/eigenvectors: Klein's method, 65; Sims' method, 61, 62n4; undetermined coefficients approach, 67  
 Generalized Method of Moments (GMM): asset pricing model, 287n1, 293; GAUSS programming, 287n1, 289n2, 310; identification problems, 293; implementation, 286–294; law of large numbers, 289; null hypothesis, 290–291; OLS regression, 291; orthogonality conditions, 287, 289, 291, 293, 300; overidentified case, 291; parameter estimation, 289–294; quadratic minimization, 288; sensitivity analysis, 294; structural stability, 290  
 Gertler, M., 31  
 Geweke, J.: Bayesian methods, 358; dynamic factor models, 187; Gibbs sampler, 217; Markov chains, 216; model comparisons, 361, 363; Monte Carlo integration, 196–197, 203, 205, 209; numerical integration, 352; prior information sources, 355  
 Ghez, G. R., 260, 270  
 Gibbs sampler, 216–219  
 Godsill, S. J., 221  
 Goldberger, A. S., 360  
 Gordon, N., 166, 221, 222, 226  
 Gourieroux, C., 286, 294, 298  
 gradient procedures, 15, 17  
 Granger, C. W. J., 166  
 Greenberg, E., 218  
 Greene, W. H., 160, 286, 315  
 Greenwood, J., 272  
 grid plots, likelihood surface, 331  
 gross returns, 40, 274  
 Guerron-Quintana, P., 69  
 Gui, F., 44, 374  
 Gui, S. M., 95  
 Haavelmo, T., 253, 256  
 habit/durability stocks, 41–43, 264–267  
 Haefke, Christian, 71n2  
 Hamilton, J. D.: ARCH/GARCH and VAR model, 149; asymmetry in mean growth of GDP, 187; decompositions, 156; derivative-based algorithms, 328; Generalized Method of Moments, 291; identified VARS, 304; iteration techniques, 82; likelihood analysis, 315; linear projections, 174; Markov chains, 215;

## Index

matching moments, 286; optimization algorithms, 319; removing trends and isolating cycles, 114, 121; sequential algorithm, 175; state-space representations, 166; time series, 139  
 Handschin, J., 221  
 Hannan, E. J., 303  
 Hansen, B. E., 300, 336  
 Hansen, G. D., 272  
 Hansen, L. P.: calibration critiques, 278–279; contributions to the field, 3, 255–256; expectations, 10; GMM GAUSS code, 287n1; log-linear mappings, 18; matching moments, 291, 297; moment-based estimation, 286–289, 300  
 Hansen, T., 257  
 Harvey, A. C.: derivative-based algorithms, 328, 329; likelihood analysis, 315; optimization algorithms, 319; removing trends and isolating cycles, 114, 121; spuriousness, 135; state-space representations, 166; time series, 139  
 Hastings, W. K., 216  
 Hayya, J. C., 136  
 Heaton, J., 41, 287n1  
 Heckman, J. J., 278–279  
 Heer, B., 69, 88  
 Heiss, 83  
 Hercowitz, Z., 272  
 Hermitian transpose, 65  
 Hessian: derivative-based methods, 328–330; ill-behaved surfaces, 332; likelihood surface, 331; simplex algorithms, 316–317  
 heteroskedastic innovations, 145–147  
 hoarding, labor and, 272  
 Hodrick-Prescott (H-P) filter: calibration and, 271; cycle isolation, 128–132; impulse response functions, 157; spuriousness and, 135–136; summary statistics/time series, 149–157; trend removal, 119–120  
 homogenous systems, 52–54  
 households: asset pricing model, 38–40; habit/durability stocks and, 41–43; labor-leisure trade-off, 18, 20–26, 254, 264–273; monopolistic competition model, 28–29; optimal growth model, 70; RBC model, 20–21, 22; self-control in, 44–47  
 inversion formula, 123, 125  
 inverted-Gamma distribution, 367  
 inverted-Wishart distribution, 357  
 invertibility problem, 304  
 Ireland, P., 4, 28, 33–36, 318, 337–339  
 Isard, M., 226  
 IS curve, 33–36, 338

*Index*

- iteration techniques: dynamic programming, 87–89; policy-function iterations, 94–95; value-function iterations, 89–94
- Jacobian matrices: likelihood evaluation, 236, 242; linearization, 13–15; maximum likelihood, 332, 340; time series behavior, 160
- Jaeger, A., 135
- Jeffreys, H., 364, 386
- Jeffreys prior, 357
- Jordan decomposition, 59, 61
- Judd, K. L.: asset pricing model, 46; collocation method, 76; dynamic programming, 88; gradients and Hessians, 328; likelihood function, 328; nonlinear approximation methods, 69, 71; orthogonal polynomials, 74; perturbation methods, 95, 97; projection methods,  $\sqrt{\ln 2}$ , 72
- Judge, G. G., 157
- Kahn, C. M., 57–60
- Kaiser, R., 114, 121, 128
- Kalman filter: applications, 11; background, 173–175; empirical application, 337, 340–348; forecasting step, 176–177; GAUSS programming, 174n1; likelihood evaluation, 174–182; mean squared error, 174–175, 177–178; prediction step, 176, 177–178; sequential algorithm, 175–178; serially correlated measurement errors, 181–182; smoothing, 178–181, 337, 340–348; state-space representations, 173–182, 195; TVP model, 182; updating step, 176, 177; without measurement errors, 181
- Kang, H., 136
- Kehoe, P. J., 19, 306, 338
- King, R. G., 4, 58, 132–134
- Kitagawa, G., 226
- Klein, P., 64–66
- Klein's method, 64–66
- Kloek, T., 195
- Kocherlakota, N. R., 38, 171, 273–275, 293, 317–318
- Koop, G., 182, 352
- Koopman, S. J., 234
- Koopmans, T. C., 253
- Kose, A., 187
- Kronecker product, 148, 245, 302, 357
- Kroner, K. F., 147
- Krusell, P., 264
- Kumamoto, H., 230
- Kydland, F. E.: calibration exercise, 6, 18–19, 253–261, 272, 276–278; contributions to the field, 3–4; covariance terms, 302
- labor: data set, 114–115; monopolistic competition model and, 28–29; National Income and Product Accounts, 270; RBC model, 18, 20–26, 254, 264–273
- labor hoarding, 272
- labor-leisure trade-off, 18, 20–26, 254, 264–273
- Laertius, Diogenes, 285
- Lagarias, J. C., 319, 323, 327
- lag-length specification: leading approaches to, 157–159; time series, 157–162
- Lancaster, T., 352
- Law of Iterated Expectations, 60, 179–180
- law of large numbers, 196, 199, 289, 355
- law of motion: for capital, RBC model, 25, 271; for dividends, 375; linearized DSGE model, 304; Monte Carlo integration, 194; nonlinear approximation, 70, 307; reduced-form analysis, 184; for state variables, 4, 88, 168
- $\ell$ -dimensional case, 78–79
- Leamer, E. E., 352
- Lee, B. S., 294
- Lehmann, E. L., 213
- leisure, 18, 20–26
- Leontief micro production technologies, 278
- LeRoy, S. F., 19, 38, 39, 373
- Liesenfeld, R., 187, 336, 365
- likelihood density, 172, 174, 181
- likelihood evaluation: Kalman filter, 174–182; particle filter and, 231; sequential Monte Carlo methods, 194, 221–224; state-space representations, 171–173, 193–194, 223–224; via direct Monte Carlo integration, 201
- likelihood function: Auxiliary Particle Filter, 231; Bayesian methods, 352–353; derivation, 352–353; EIS filter, 234; logarithmic transform, 11, 316–317; notation, 11; period- $t$ , 178, 194, 224; state-space representations, 171–173, 178, 224; unadapted filters, 225

*Index*

- likelihood integral, SIS algorithm, 225
- likelihood ratio (LR) test, 157–158, 335
- likelihood surface: discontinuities problem, 330–331; grid plots, 331; identification problem, 331; ill-behaved, 330–334; logistic transformations, 332–333; method-of-moments exercises, 334; multiple local minima problem, 331; multi-step optimization, 334; parameter transformations, 332; quasi-Bayesian methods, 334; simplex algorithms, 319–320
- limited-information analyses: calibration exercises, 138; matching moments, 138, 286
- linearization: example equations, 15–17; log-linear approximations, 14–15; Taylor Series approximations, 12–13
- linearized models simulation, 306–307
- linear/Normal state-space representations, 195
- linear solution techniques: asset pricing with linear utility model, 55–56; Blanchard-Kahn method, 57–60; homogenous systems, 52–54; introduction, 51–52; Klein's method, 64–66; optimal consumption model, 54–55; Ramsey's optimal growth model, 56–57; Sims' method, 61–64; undetermined coefficients approach, 66–68
- linear systems dynamic behavior: asymptotically stable, 51; saddle-path stable, 51; unstable, 51
- Litterman, R., 365
- Liu, J., 226
- Ljungqvist, L. L.: convergence and policy-function iterations, 94; DSGE models, 20; dynamic programming, 88; Law of Iterated Expectations, 60; Markov chains, 215; nonlinear approximation methods, 69
- Lo, A. W., 20, 38
- Lobachevsky, Nicolai, 113
- logistic transformations, likelihood surface, 332–333
- log-likelihood function, 11, 316–317
- log-linear approximations, 14–15
- log-linearization: monopolistic competition model, 34–37; RBC model, 26–27
- log-linear model representations, 11
- Lubrano, M., 195, 217–218, 237, 352
- Lucas, R. E.: asset pricing model, 4, 19–20, 38, 46, 373–374; calibration, 257, 261–264, 267; contributions to the field, 3; dynamic programming, 22, 88–89; systems-of-equations models, 254–255; time-varying parameters, 182
- MA(1) process, 139
- MacKinlay, A. C., 20, 38
- macroeconomic research, empirical: background, 3–4; DSGE models as foundations of, 3–4
- Mankiw, N. G., 19
- mapping: Blanchard-Kahn method, 57–60; Klein's method, 64–66; Law of Iterated Expectations, 60; log-linear, 18; nonlinear DSGE models into state-space representations, 171; Sims' method, 61–64; stochastic innovations, 56; structural innovations to VAR disturbances, 305; undetermined coefficients approach, 66–68
- Maravall, A., 114, 121, 128
- Markov Chain (MCMC) integration: Bayesian methods, 355–360, 363; Gibbs sampler, 216–218; Metropolis-Hastings Algorithms, 218–220; Monte Carlo methods, 215–220
- markets: asset pricing model, 38–47; habit/durability stocks and, 264–267; imperfections and, 19; monopolistic competition model, 28–37
- Markov process: nonlinear approximation methods, 94; optimal growth model, 80–82, 84–86; orthogonal polynomials, 80–86; policy-function iterations, 94; projection methods, 84–86; state-space representations, 186–187; stochastic process and, 199, 215–216; value-function iterations, 91, 92
- Marquis, M. H., 272
- matching moments: calibration and, 261; Cholesky decomposition, 304; CRRA and, 301; DSGE models implementation, 300–308; empirical application, 308–313; estimation stage, 285–286; Euler equations, 300–301; Generalized Method of Moments, 286–294; implementation, 286–300; indirect inference, 297–300; limited-information analyses, 138, 286; linearized model simulation, 301–307; nonlinear approximation simulations,

matching moments (*cont.*)  
 307–308; overidentified case, 286, 287; overview, 285–286; parameter estimation, 289–295, 310–312; RBC model application, 308–313; Simulated Method of Moments, 294–297; structural shocks, 286; testing stage, 285–286; VAR model, 302–306  
 Matlab code: Klein's method, 64n6; perturbation methods, 95n7; projection methods, 71n2; Sims' method, 61n2; undetermined coefficients, 66n9  
 Maussner, A., 69, 88  
 maximum likelihood: BFGS alternative, 329; break date identification, 336–337; business cycle fluctuations application, 337–350; convergence, 327, 329, 330; derivative-based algorithms, 319, 328–330; DFP alternative, 329; discontinuities, 331; empirical application, 337–350; expansion/contraction points, 322–325; flow chart, 326; GAUSS programming, 319n1, 328n2; historical perspective, 316–318; hybrid approach, 318; ill-behaved surfaces, 330–334; introduction, 316–318; IS curve, 338; Kalman filtered, application, 337, 340–348; likelihood ratio test, 335; likelihood surface, 319–320; model diagnostics, 334–337; neighborhoods, 319–320; Newton's Method, 328; optimization algorithms, 318–330; overshooting, 329; overview, 314–315; parameter estimation, 314–315, 319, 334–340; parameter stability, 335–337; Phillips curve, 338–339; purpose, 319; quasi-Newton methods, 329; reflection point, 321; restrictions tests, 334; shocks and, 318; shrink/nonshrink step, 321–327; simplex algorithms, 319–328; structural-stability tests, 336; Taylor Rule, 338; temporal stability tests, 335–336; time data points, 316; triangularization techniques, 319; two-dimensional case, 320–325; VAR model, 318; vertices, 319  
 Mayne, D., 221  
 McFadden, D., 294  
 McGrattan, E. R.: matching moments, 306; maximum likelihood, 318, 338; monopolistic competition model, 19; projection methods, 72; volatility of hours, 272

Mead, R., 319  
 mean squared error (MSE), 174–175, 177–178  
 mean-zero CSSPs, 115, 139, 279  
 mean-zero unforecastable stochastic process, 82  
 measurement (observation) density, 167, 172  
 measurement errors: DSGE models, 11; GAUSS programming, 174n1; Kalman filter without, 181; linearized DSGE model, 304–305; nondegenerate, 170–171; state-space representations, 170–171  
 Mehra, R. M.: asset pricing model, 38, 39–40, 300; equity premium puzzle, 273–275, 277; monopolistic competition model, 19  
 method-of-moments exercises, 334  
 method of undetermined coefficients, 143, 144  
 Metropolis, M., 195, 216  
 Metropolis-Hastings Algorithm, 218–220, 363  
 Metropolis-Hastings MCMC sampler, 361  
 Minnesota prior, 365  
 mixed-estimation procedure, 360  
 model-classification, calibration, 254  
 model impulse response functions, 164–165  
 model likelihood function, 195  
 model variables, 12  
 momentary and temptation utility function, 375–386  
 moment-matching. *See* matching moments  
 monetary authority, 31–32  
 monetary policy, 28–37; shock, 31–39, 337–350  
 monetary utility function, 44  
 Monfort, A., 286, 294, 297, 298  
 monopolistic competition model: additional resources, 19; aggregate uncertainty sources, 28; CES production function, 30–31; consumption and, 28–29; data, 28; elasticities, 29–30; empirical application, 337–350; environment, 29–30; firms, 29–31; households, 28–29; IS curve, 33–36; labor and, 28–29; linearization, 28; log-linearization, 34–37; monetary authority, 31–32; money holdings and, 29; nonlinear system, 33–34; output gap, 31–32, 33, 35–36; Phillips curve, 33–36; purpose of, 19; RBC model

compared, 28, 31; shocks, 28, 29, 30, 31, 32, 36; stationarity, 32; sticky prices, 31; stochastic specifications, 32; Taylor Rule, 31, 33; utility function, 29  
 Monsell, B. C., 130  
 Monte Carlo integration: achieving efficiency, 206–211; Bayesian methods, 162, 355–360; burn-in phase, 199–200, 217; construction of approximations using pseudo-random numbers, 195; direct, 196–202; efficiency, 206–211; efficient importance sampling, 211–215; estimate of standard errors, 160–161; GAUSS programming, 199, 214nn1–2; Gibbs sampler, 216–218; importance sampling, 202–211; introduction, 195–196; Markov Chain method, 215–220; Metropolis-Hastings Algorithms, 218–220; model simulation, 198–201; numerical standard error, 196–197, 203–205; optimal growth model, 200–201; overview, 193–196; posterior inference via direct, 201–202; relative numerical efficiency, 205–207  
 Monte Carlo Markov Chain (MCMC) integration: Bayesian methods, 355–360, 363; Gibbs sampler, 216–218; Metropolis-Hastings Algorithms, 218–220  
 Monte Carlo methods: conditionally optimal filters, 228–232; direct integration, 196–202; efficient importance sampling, 211–215; EIS filter, 233–241; importance sampling, 202–206; likelihood evaluation, 221–224; Markov Chain, 215–220; unadapted filters, 224–228. *See also* sequential Monte Carlo (SMC) methods  
 Moulines, E., 221  
 moving average process, 139–140  
 multi-asset pricing model, 39–40  
 multiple local minima problem, 331  
 multiple-shock method, 318  
 multi-step optimization, 334  
 multivariate- $t$  densities, 208–2209  
 Murray, C. J., 136  
 Mykland, P., 217  
 Nason, J. M., 136  
 national income accounting identity, 254  
 National Income and Product Accounts (NIPA), 270  
 nonstochastic models, 80–81  
 Nordhaus, W., 118  
 Normal equation, 174  
 Normal/Inverted-Gamma distributions, 368, 369

Normal/Inverted-Wishard distribution, 216, 357–358, 367  
 Normal state-space representation, 173–174  
 notation: Bayesian methods, 351–353; Blanchard-Kahn method, 58–59; calibration, 260, 269; DSGE models, 9–12, 95–96, 304; dynamic factor models, 187; Generalized Method of Moments, 295; Howard improvement algorithm, 94; Klein's method, 64; linear solution techniques, 52, 304; log-linear model representations, 11; nonlinear approximations, 11–12, 95–96; perturbation methods, 95–96; reduced-form models, 139; sequential algorithm, 175; Sims' method, 61–62; Simulated Method of Moments, 297; time series behavior, 139  
 null hypothesis: Bayesian perspective, 278, 362; calibration, 277; classical hypothesis testing, 217, 315; of convergence, 217; matching moments, 287–291, 294, 297, 300; of structural stability, 335–336  
 numerical standard error (NSE): direct Monte Carlo integration, 196–197; Gibbs sampler, 217; importance sampling, 203–205, 212; Simulated Method of Moments, 296  
 observation (measurement) density, 167, 172  
 Ogaki, Masao, 287n1  
 Ohanian, L. E., 302  
 optimal consumption model, 54–55  
 optimal growth model: direct Monte Carlo integration, 200–201; EGMs modifications, 92–94; Gauss-Hermite weights/nodes, 84–85; GAUSS programming, 80n4; nonlinear solution techniques, 70, 79–87; notation, 70–71; orthogonal polynomials, 79–86; projection methods application to, 79–87; Ramsey's, 56–57; reprise, 70  
 optimality: Bellman's equation, 88–89; nonlinear approximation, 88; shrink/nonshrink step, 321–327; unconditional, 222, 232, 233–241  
 optimality condition: asset pricing model, 44, 47; intertemporal, 10, 25, 263; in-tratemporal, 23, 254; RBC model, 23, 25–26

optimization algorithms: derivative-based, 319, 328–330; maximum likelihood, 318–330; simplex algorithms, 319–328  
 Ord, K., 136  
 ordinary least squares (OLS) regression: ARMA model, 144–145; Bayesian methods and, 357, 367–369; Generalized Method of Moments, 291; RBC model, 367–369; time series, 144–145, 148–149; trend removal, 116–117; VAR model, 148–149  
 orthogonality conditions, GMM, 287, 289, 291, 293, 300  
 orthogonalized innovations, 156  
 orthogonal polynomials: Chebyshev family, 73–74; defined, 73; Gauss-Hermite weights/nodes, 84–85; implementation, 74–77;  $l$ -dimensional case, 73–74, 78–79; Markov process, 80–86; nonlinear solution techniques, 73–74; optimal growth model, 79–86; projection methods, 73–74  
 Ospina, S., 69  
 Otrrok, C.: calibration, 264, 266–267, 283; dynamic factor models, 187; MCMC method, 215; stochastic singularity, 318  
 output gap, 31–32, 33, 35–36  
 overidentified case, 286, 287, 291  
 overshooting, 329  
 Pakes, A., 294  
 Palka, B., 120  
 parameter estimation: Bayesian methods, 376, 381–386; Generalized Method of Moments, 289–294; matching moments, 289–295, 310–312; maximum likelihood, 314–315, 319, 334–340; Simulated Method of Moments, 295; VAR model, 148–149  
 parameter transformations, likelihood surface, 332  
 parsimony, 157  
 particle filters: Auxiliary Particle Filter, 231–232; bootstrap particle filter, 222, 226, 227, 320  
 particles, 223  
 particle swarm, 223  
 Pearson, K., 286  
 Peralta-Alva, A., 198, 306

period- $t$  algorithm: Auxiliary Particle Filter, 231; conditionally optimal filters, 228–229; efficient importance sampling, 235; unadapted filters, 225–226  
 period- $t$  filtering density, 178, 224, 225  
 period- $t$  likelihood function, 194  
 Perli, R., 272  
 perpetual inventory method, 271  
 Perron, P., 118  
 Persand, P., 147  
 perturbation methods: asset pricing model, 105–109; DSGE models, 99–104; GAUSS programming, 95n7; Matlab code, 95n7; nonlinear approximations, 95–109; notation, 95–96; overview, 97–98  
 Pesaran, M. H., 365  
 Pesendorfer, W., 44, 374  
 phase effect, 132  
 Phillips, P.C.B., 137  
 Phillips curve, 33–36, 338–339  
 physical capital: RBC model, 20–21, 271; state-transition equation, 235  
 Pitt, M. K., 221, 222, 231, 232  
 Plosser, C. I., 4, 117, 365  
 Poincaré, Henri, 3  
 Poirier, D. J., 314, 352  
 policy, models guiding, 255  
 policy-function iterations, 94–95  
 policy functions: approximations, 4; asset pricing model, 105–110, 170; nonlinear approximations, 69–70, 94–95; nonstochastic optimal growth model, 80–81; notation, 11–12; RBC model, 170; stochastic optimal growth model, 81–82  
 Pollard, D., 294  
 Porter, R. D., 19, 38, 39, 373  
 posterior distribution, 315; Bayesian methods, 351, 354, 358–364, 367; random drawings simulation, 354–355  
 posterior inference, 195, 201–202, 204–205  
 posterior odds, 352, 362, 363  
 Potter, S. M., 182, 365  
 power spectrum. *See* spectra  
 predetermined/nonpredetermined variables: Blanchard-Kahn method, 57–60; Klein's method, 64–65; Sims' method, 61  
 predictive density, 172, 174–176, 178, 181, 194  
 Prescott, E. C.: asset pricing model, 19–20, 300; calibration exercise, 6, 38, 39–40, 253–262, 272–278; contributions to the field, 3–4; covariance terms, 302; maximum likelihood, 318; RBC model, 18–19; time-varying parameters, 182  
 Primiceri, G., 182  
 prior odds ratio, 362  
 probability approach to econometrics, 253, 255, 256–257, 276  
 probability distribution functions (pdfs), 167, 195  
 production function: aggregate, 33; calibration and, 254, 271–272; Cobb-Douglas, 22; constant elasticity of substitution, 30; final good, 30; final-good production function, 22; RBC model, 22, 271–272  
 productivity shocks: business cycle fluctuations and, 19, 268–273; calibration, and business cycle fluctuations, 268–273; New Keynesian model, 19–23; RBC model, 19–23, 26, 268–273, 306; total-factor, 164, 318  
 projection methods: approximation error, 86; Chebyshev polynomial, 73–75, 76, 77–78; defined, 71; finite element method, 72–73, 74, 76; Galerkin method, 76; Gauss-Hermite weights/nodes, 84–85; GAUSS programming, 71n2, 76n3, 85; implementation, 74–77;  $l$ -dimensional case, 78–79; optimal growth model, 74–75, 79–87; orthogonal polynomials, 73–74; overview, 71–72; starting values, effective, 77; Weierstrass theorem, 71; weighted least-squares problem, 76  
 propagation: SIR algorithm, 226; SIS algorithm, 225, 226  
 pseudo-random numbers, 195  
 quadratic minimization, 288  
 quadrature approximations, 83, 91, 195  
 quasi-Bayesian methods, 334  
 quasi-Newton algorithm, 47, 86, 329  
 QZ decomposition, 61, 65  
 Ramamurthy, S., 218  
 Ramsey's optimal growth model, 56–57  
 random-walk algorithms, 220  
 Raney, V. A., 306  
 rational expectations models, 56  
 ratio of marginal likelihoods, 362  
 ratio of priors, 362

- real business cycle (RBC) model: additional resources, 19; Bayesian methods, 364–373; business cycle fluctuations and, 19, 21; calibration and, 257, 259, 264–273, 281–282; Cobb-Douglas production function, 22; constraints, 21–22; consumption and, 20–26; consumption-investment trade-off, 254; CRRA and, 22, 264–273; EIS filter implementation, 246–250; empirical analysis, 19, 308–313; environment of, 20–23; Euler equations, 26; framework, 18; functional forms, 22–27; growth and, 19, 21; households in the, 20–21, 22; introduction, 19–20; key features, 18–19; labor-leisure trade-off, 18, 20–26, 254, 264–273; log-linearization, 26–27; market imperfections and, 19; matching moments, 308–313; model impulse response functions, 164–165; monopolistic competition model and, 28, 31; NBER-defined recessions and, 341, 345–347; New Keynesian model extension, 337; nonlinear approximations, 23–26; OLS regression, 367–369; optimality condition, 23, 25–26, 254; prior information source, 364–373; production function, 22, 271–272; productivity shocks, 268–273, 306; purpose, 19; restated, 169; shocks, 19–23, 26; state-space representations, 246–250; state-transition equation, 235; steady state variables, 24–25; structural parameters, 368; systems-of-equations models, 254; total factor productivity, 187; trend removal, 25; uncertainty, 18; utility function, 20, 22–23; volatility of hours, 271–272
- Rebelo, S. T., 4, 272
- reduced-form analysis: ARMA model, 139–145; Bayesian methods, 355–360, 364–373; forecasting application, 352, 364–373; historically, 3, 255; indirect inference methodology, 297; sources of prior information, 355–360, 364–373; time series behavior, 139–149; VAR model, 147–149
- reduced-form state-space representations: dynamic factor models, 187–190; regime switching, 186–187; stochastic volatility, 185–186; time-varying parameters, 182–185
- regime switching model, 186–187
- relative mean square approximation error (RMSAE) statistic, 281–282
- relative numerical efficiency (RNE), 205–207
- Renault, E., 297, 298
- resampling: SIR algorithm, 226; SIS algorithm, 225–226; unadapted filters, 225
- residual function, 77
- restrictions tests, maximum likelihood, 334
- Richard, J.-F.: Bayesian methods, 352; break date identification, 336; efficient importance sampling, 211; EIS filter, 237; Gibbs sampler, 217; Monte Carlo methods, 195; prior information sources, 365; regime switching model, 187
- Riesz-Fisher Theorem, 123, 125
- Ripoll, M.: asset pricing preference specifications, 38, 40–41; Bayesian methods, 361, 374, 378–379; importance sampling, 209; momentary and temptation utility function, 45
- risk aversion, 22
- risk-free rate puzzle, 20, 38, 40, 274–275
- Ristic, B., 166, 221, 222
- Robert, C. P., 195, 216, 352
- Rogerson, R., 272, 318
- Romer, D., 19, 20, 22
- Rowenhorst, G., 259
- Rubenstein, R. Y., 195
- Rubin, D. B., 225
- Rubinstein, R. Y., 161
- Rubio-Ramirez, J.: Bayesian methods, 361; CRRA parameter notation, 250; MCMC method, 215; Monte Carlo experiment data set, 247; nonlinear solution techniques, 69; particle filter, 246; projection methods, 72; SMC methods, 221, 227
- Ruge-Murcia, F. J., 334
- saddle-path stability: asset pricing with linear utility model, 55–56; Blanchard-Kahn method, 59; defined, 51; homogenous systems, 53; Klein's method, 65; optimal consumption model, 54–55
- Sakellaris, P., 272
- Sala-i-Martin, X., 263
- Salmond, D. J., 221, 226
- sample impoverishment, 226, 245–246
- sample pooling, 360

- samplers: blindness problem of unadapted, 228; effective, characteristics of, 360
- sampling: effective, tailoring, 204, 206, 208; multivariate-*t* densities, 209–211
- sampling importance resampling (SIR) algorithm, 222, 226; degeneracy, 226; efficiency, 226; propagation, 226; resampling, 226
- Santos, A.A.F., 232
- Santos, M. S., 91, 198, 306
- Sargent, T. J.: contributions to the field, 3; convergence and policy-function iterations, 94; cross-equation restriction, 56; DSGE models, 20; dynamic factor models, 187; dynamic programming, 87, 88; expectations, 10; Law of Iterated Expectations, 60; linear projections, 174; log-linear mappings, 18; Markov chains, 215; nonlinear approximation methods, 69; removing trends and isolating cycles, 114, 121, 127; serially correlated measurement errors, 181; stochastic singularity, 317–318; systems-of-equations models, 254–256; time-varying parameters, 182
- Savin, N. E., 171, 317–318
- savings/consumption trade-off, 20–21
- Scarf, H., 257
- Schmitt-Grohé, S., 13, 95
- Schorfheide, F., 215, 359, 361, 363, 365
- Schotman, P. C., 365
- Schur decomposition, 65
- Schwarz, G., 159
- seasonal adjustment (SA), 130–132
- Sedaghat, H., 13
- self-control effect, 44–47, 375–386
- sensitivity analysis, GMM, 294
- sequential algorithm, Kalman filter, 175–178
- sequential importance sampling (SIS) algorithm, 221–222, 224–228; continuation, 225; degeneracy, 226; efficiency, 226; initialization, 224–225; likelihood integral, 225; propagation, 225, 226; resampling, 225–226; shortcomings, 222; unadapted filters, 224–228
- sequential Monte Carlo (SMC) methods: conditionally optimal filters, 228–232; degenerate transitions, 235–236; EIS filter, 233–241; likelihood evaluation and filtering, 194, 221–224; sampling importance
- resampling algorithm, 222; sequential importance sampling algorithm, 221; unadapted filters, 224–228
- serially correlated measurement errors, 181–182
- Shaw, George Bernard, 138
- Shephard, N., 221, 222, 231–232
- Shiller, R. J., 19–20, 38–39, 373, 376
- Shimelevich, L., 230
- Shiryayev, A. N., 215
- shocks: asset pricing model, 41, 45; business cycle fluctuations and, 19, 256, 268–273, 337–350; calibration and, 268–273; matching moments, 286; maximum likelihood, 318; monopolistic competition model, 28–39; RBC model, 19–23, 26, 268–273; sticky prices and, 28, 31. *See also specific types of*
- shocks, specific: cost-push shocks, 30, 32, 35, 337–350; demand shocks, 29, 31–39, 41, 337–350; endowment shock, 45–46; monetary policy shock, 31–39, 337–350; productivity shocks, 19–23, 26, 268–273, 306, 318; structural shocks, 9–11, 32–33, 35–36, 66, 171, 305, 318; technology shocks, 20, 28, 31–39, 115, 306, 337–350
- Shoven, J. B., 257
- shrink/nonshrink step, 321–327
- simplex algorithms: convergence, 327; discontinuities, 331; expansion/contraction points, 322–325; flow chart, 326; GAUSS programming, 319n1; likelihood surface, 319–320; neighborhoods, 319–320; parameter space, 319; purpose, 319; reflection point, 321; shrink/nonshrink step, 321–327; triangularization techniques, 319; two-dimensional case, 320–325; vertices, 319
- Sims, C. A.: calibration critiques, 277–279; contributions to the field, 3, 255; dynamic factor models, 187; hybrid optimization algorithm, 331; invertibility problem, 304; prior information sources, 359, 365; Sims' method, 61–64
- Sims' method: expectational errors, 62, 63; expectations errors, 61; explosive/nonexplosive components, decoupling, 61; predetermined/nonpredetermined variables, 61; websites, 61n2

Simulated Method of Moments (SMM), 294–297  
 simulation-based estimation, 307  
 single-asset pricing model, 38–40, 55, 105, 170  
 Singleton, K. J., 256, 287n1, 294, 297, 300  
 sink, 53, 59  
 skill-acquisition activities, 272  
 Slutsky's Theorem, 160  
 Smets, F., 19, 215  
 Smith, A., 297  
 Smith, A. A., 264  
 Smith, A.F.M., 221, 226  
 Smith, J. Q., 232  
 smoothing: Kalman filter, 178–181, 337, 340–348; shocks and, 46; TVP model, 182  
 smoothing density, 173  
 smoothness, 119  
 Solow, R. M., 259, 271  
 Solow residual, 271, 370  
 source, 59  
 special realization of innovations, 144  
 spectral representation, 125  
 spectrum (spectra): ACGFs and, 281; calibration and, 281–282; CSSPs, 135; cycle isolation, 125–126; DSGE models, 302; frequency, 125–126, 151; H-P filtered, 130, 151–153; summary statistics, 150–156; VAR model, 155–156  
 spuriousness, 134–136  
 squared gain of the filter: B-P filtered, 131, 134, 152; H-P filtered, 128–129, 152  
 square-summability, 123, 141, 143  
 square-summability violation, 141  
 stabilization policies, 263  
 Startz, R., 294  
 state-space representations: applications for, 166; ARMA model, 167–168; conditionally optimal filters, 228–233; degenerate transitions, 235–236; distribution of the state, 167; DSGE models, 169–171, 241–252, 301; dynamic factor models, 187–190; in economics, 166; EIS filter, 233–241; filtering, 171–173; filtering density initialization, 245–246; importance sampler initialization, 236–239, 243–245; introduction, 166–167; Kalman filter, 173–182, 195; likelihood evaluation, 171–173; nonlinear approximation methods, 173; observation

(measurement) density, 167; probability distribution functions, 167; RBC model, 246–250; reduced-form examples, 182–190; regime switching, 186–187; sequential algorithm, 175–178; serially correlated measurement errors, 181–182; smoothing, 178–181; state-transition density, 167; stochastic volatility, 185–186; time-varying parameters, 182–185; unadapted filters, 224–228  
 state-transition: density, 167; equation, 167, 170, 187  
 state variables, 10, 11–12  
 stationarity: monopolistic competition model, 28, 32, 37; notation, 9; trend- vs. difference-, 135–136  
 statistics: Akaike Information Criterion, 158–159; Bayesian Information Criterion, 159; CUSUM, 217–218; Delta method, 160; parsimony, 157; RMSAE, 281–282. *See also* data; summary statistics  
 steady state variables: notation, 10; RBC model, 24–25  
 sticky prices, 28, 31  
 stochastic errors/innovations, 51, 56, 96, 171  
 stochastic models: optimal growth, 81–82; volatility, 185–186  
 stochastic process: Markov process and, 199, 215–216; mean-zero covariance-stationary, 278; mean-zero unforecastable, 82; minimum-variance, 281  
 stochastic singularity, 171, 317–318  
 stochastic volatility (SV) model, 185–186  
 Stock, J. H., 117, 182–184, 187, 294  
 stock-price volatility puzzle, 41, 46, 373, 374–386  
 Stokey, N. L., 20, 22, 46, 88–89  
 structural shocks: DSGE model, 9, 10, 11; linearized DSGE model, 305; monopolistic competition model, 32–33, 35–36; multiple-shock approach, 318; notation, 10–11; stochastic behavior and, 171; undetermined coefficients approach, 66  
 structural-stability tests, 336  
 substitution elasticity, 22  
 summary statistics: ARMA model, 149–152; lag-length specification, 157–159; precision of measurements, 159–162; spectra, 150–153; theoretical predictions

of, 162–165; time series, 149–165; VAR model, 153–155, 163  
 Summers, L., 19  
 Sustek, R., 313  
 Svetnik, V., 230  
 system of arbitrary order, 28  
 system reduction, model-specific, 58, 60  
 systems-of-equations models, 3, 253–255, 258  
 tailored samplers: alternatives, 211; effective, 204, 206, 208; multivariate- $t$  densities, 209–211  
 Tauchen, G., 82, 293  
 Taylor, S. J., 185  
 Taylor Rule, 31, 338  
 Taylor Series: linearized DSGE model, 12–14; perturbation methods, 96–97, 99  
 Taylor's Theorem, 97  
 technology constraints, 254  
 technology evolution, 19  
 technology shocks: business cycle fluctuations and, 28, 337–350; labor and, 115, 306; monopolistic competition model, 28, 31–39; RBC model, 20  
 temporal stability tests, 335–336  
 temptation effect, 46  
 temptation utility function, 375–386  
 tensor products, 78–79, 84  
 Theil, Henry, 253, 277, 360  
 theorems: Bayes' Theorem, 194, 224, 225; Chebyshev interpolation theorem, 77, 78; DeMoivre's Theorem, 122, 126; Implicit Function Theorem, 97–98, 100–101; Riesz-Fisher Theorem, 123, 125; Slutsky's Theorem, 160; Taylor's Theorem, 97; Weierstrass theorem, 71; Wold Decomposition Theorem, 140–141, 147.  
*See also* algorithms  
 time series: Akaike Information Criterion, 158–159; ARMA model, 139–145; Bayesian Information Criterion, 159; Cramér Representations, 124–125; cycle isolation, 124–125; Delta method, 160; heteroskedastic innovations, 145–147; H-P filtered, 149–157; Kronecker product, 148; lag-length specification, 157–162; Monte Carlo methods, 160–162; OLS regression, 144–145, 148–149; parsimony, 157; precision of measurements, 159–162; reduced-form models, 139–149; Slutsky's Theorem, 160; summary statistics, 149–165; trend removal, 115–120; VAR model, 147–149, 187; Wold Decomposition Theorem, 140–141, 147  
 time- $t$  data points, 316  
 time-varying parameters (TVP) model, 182–185  
 Tinbergen, Jan, 256  
 Todd, R. M., 306  
 total factor productivity (TFP), 70, 164, 187, 235; shock, 164, 318  
 trend removal: consumption and, 118–120; covariance stationarity and, 116; CSSPs and, 115–117; cycle isolation and, 114; data preparation, 114–115; detrending approach, 116–119; differencing approach, 116–119; filters for, 119, 126; frequency and, 114, 119; GAUSS programming, 116n1, 117n2, 119n3; goal of, 115; H-P filtered, 119–120; introduction, 113–115; lag operators, 157–162; OLS regression, 116–117; RBC model, 25; spuriousness, 134–136; time series, 115–120; volatility and, 119–120  
 triangularization techniques, 319  
 two-dimensional case, simplex algorithms, 320–325  
 Uhlig, H., 66  
 Ulam, S., 195  
 unadapted filters, SMC methods, 224–228  
 uncertainty: monopolistic competition model, 28; RBC model, 21–22  
 unconditional distribution, 167, 224, 227  
 unconditionally optimal sampler, 233  
 undetermined coefficients approach, 66–68  
 uniqueness, 63, 67  
 Uribe, M., 13, 95  
 utility function: asset pricing model, 44, 45; calibration and, 264–267; monopolistic competition model, 29, 45; RBC model, 20, 22–23; von Neumann-Morgenstern, 44; welfare cost, 261–268  
 utility maximization problem, 38  
 utility of wealth tomorrow, 45  
 value-function iterations: EGMs modifications, 92–94; Markov process, 91, 92; nonlinear approximations, 89–94  
 van Dijk, H. K., 195, 365

## Index

- variance decompositions, 337, 340  
 vector autoregressive (VAR) model:  
   ARCH/GARCH specifications, 149;  
   Bayesian methods, 384–386; calibration  
   and, 283–284; cross-equation restriction,  
   255–256; GAUSS programming,  
   149n3; identified VARS, 304, 306;  
   impulse response functions, 156; infinite-  
   order, 305; interpreting forecasts of,  
   352; Kronecker product, 148; matching  
   moments, 302–306; maximum likelihood,  
   318; Minnesota prior, 365; multivariate  
   time series, 82; OLS regression, 148–149;  
   spectra, 155–156; summary statistics,  
   153–155, 163; time series behavior,  
   147–149, 187; Wold Representation, 147  
 vector autoregressive (VAR) specification,  
   64, 66  
 Vigfusson, R., 306  
 Vigo-Aguiar, V. A., 91  
 volatility: asset pricing model, 19, 38,  
   40, 46; calibration and, 271–272; con-  
   sumption and, 271–272; stochastic  
   models, 185–186; trend removal and,  
   119–120  
 volatility puzzle, 19, 38, 46  
 von Neumann-Morgenstern utility function,  
   44  
 Watson, M. W.: calibration extensions,  
   279–282; dynamic factor models, 187;  
   identified VARS, 304; spectra, 302; system  
   reduction, 58; time-varying parameters,  
   182–184  
 wealth, household, 28–29  
 Weierstrass theorem, 71  
 weighted least-squares problem, 76  
 weighting function, 73  
 weighting matrix, 288–290, 295, 298–299  
 Weil, P. W., 19, 38, 40  
 welfare cost of business cycles, 261–268  
 Wen, Y., 114, 282  
 West, K., 288, 290  
 Whalley, J., 257  
 Whiteman, C. H.: asset pricing with linear  
   utility, 56; Bayesian methods, 376; cal-  
   ibration extensions, 283; detrending vs.  
   differencing, 117; dynamic factor models,  
   187; importance sampling, 209; maxi-  
   mum likelihood, 318, 338, 346; prior  
   information sources, 359, 365; trend- vs.  
   difference-stationarity, 137  
 white-noise process, 139, 140, 145  
 Wickens, M., 20  
 Winschel, V., 83  
 Wishart distribution, 357  
 Woitek, U., 132  
 Wold Decomposition Theorem, 140–141  
 Wold Representation, 141, 143–144, 146–  
   147, 152  
 Wolpert, R. L., 314  
 Woodford, M., 19  
 Wouters, R., 19, 215  
 Wright, J. H., 294  
 Wright, R., 272, 318  
 X-11 filter, 130  
 Yogo, M., 294  
 Yu, B., 217  
 Zakai, M., 225  
 Zaritskii, V., 230  
 Zellner, A., 237, 352, 356  
 Zha, T., 304, 359, 365  
 Zhang, W., 211

