Chapter 1

Introduction

> 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 the crest of 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).

The empirical tools outlined in the text share a common foundation: a system of nonlinear expectational difference equations derived as the solution of a DSGE model. The strategies outlined for implementing these models empirically typically involve the derivation of approximations of the systems, and then the establishment of various empirical implications of the systems. The primary focus of this book is on the latter component of these strategies: This text covers a wide range of alternative methodologies that have been used in pursuit of a wide range of empirical applications. Demonstrated applications include: 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).

1.2 Overview

This book is divided into three parts. Part I presents foundational material included to help keep the book self-contained. Following this introduction, chapter 2 outlines two preliminary steps often used in converting a given DSGE model into an empirically implementable system of equations. The first step involves the linear approximation of the model; the second step involves the solution of the resulting linearized system. The solution takes the form of a state-space representation for the observable variables featured in the model.

Chapter 3 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 modelled by the theory. For example, the separation of

trend from cycle is necessary in confronting trending data with models of business cycle activity.

Chapter 4 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. Finally, the Kalman filter is presented as a means for pursuing likelihood-based, or full-information, analyses of state-space representations. Part I concludes in chapter 5 with an introduction of the benchmark models that serve as examples in part II.

Part II, composed of chapters 6 through 9, presents the following empirical methodologies: calibration, limited-information estimation, maximum likelihood estimation, and Bayesian estimation. Each chapter contains a general presentation of the methodology, and then presents applications of the methodology to the benchmark models in pursuit of alternative empirical objectives.

Chapter 6 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 insure 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 limited-information estimation methodologies presented in chapter 7 serve as one way to address problems arising from the statistical informality associated with calibration exercises. Motivation for their im-

plementation 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). Limited-information estimation methodologies 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 over-identified), the estimation stage also yields objective goodness-of-fit measures that can be used to judge the model's empirical performance. Prominent examples of limited-information methodologies include the generalized and simulated methods of moments (GMM and SMM), and indirect-inference methods.

Limited-information estimation procedures 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 8 and 9 present full-information counterparts to these methodologies: likelihood-based analyses. Given a distributional assumption regarding sources of stochastic behavior in a given model, chapter 8 details how the full range of empirical implications of the model may be assessed via maximum-likelihood analysis, facilitated by use of the Kalman filter. 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 9, 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. A drawback associated with the adoption of a Bayesian perspective in this class of models is that posterior analysis must be accomplished via the use of sophisticated numerical techniques; special attention is devoted to this problem in the chapter.

Part III outlines how nonlinear model approximations can be used in place of linear approximations in pursuing the empirical objectives described throughout the book. Chapter 10 presents three leading alternatives to the linearization approach to model solution presented in chapter 2: projection methods, value-function iterations, and policy-function iterations. Chapter 11 then describes how the empirical methodologies presented in chapters 6–9 may be applied to nonlinear approximations of the underlying model produced by these alternative solution methodologies.

The key step in shifting from linear to nonlinear approximations involves the reliance upon simulations from the underlying model for characterizing its statistical implications. In conducting calibration and limited-information estimation analyses, simulations are used to construct numerical estimates of the statistical targets chosen for analysis, because analytical expressions for these targets are no longer available. And in conducting full-information analyses, simulations are used to construct numerical approximations of the likelihood function corresponding with the underlying model, using a numerical tool known as the particle filter.

The organization we have chosen for the book stems from our view that the coverage of empirical applications involving nonlinear model approximations is better understood once a solid understanding of the use of linear approximations has been gained. Moreover, linear approximations usefully serve as complementary inputs into the implementation of nonlinear approximations. However, if one wished to cover linear and nonlinear applications in concert, then we suggest the following approach. Begin exploring model-solution techniques by covering chapters 2 and 10 simultaneously. Then having worked through chapter 3 and sections 4.1 and 4.2 of chapter 4, cover section 4.3 of chapter 4 (the Kalman filter) along with section 11.2 of chapter 11 (the particle filter). Then proceed through chapters 5–9 as organized, coupling section 7.3.4 of chapter 7 with section 11.1 of chapter 11.

In the spirit of reducing barriers to entry into the field, we have developed a textbook Web site 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 Web site address is http://www.pitt.edu/~dejong/text.htm. References to procedures available at 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.cepremap.cnrs.fr/~michel/dynare/

Much of the code provided at our Web site 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.

1.3 Notation

A common set of notation is used throughout the text in presenting models and empirical methodologies. A summary is as follows. Steady state values of levels of variables are denoted with an upper bar. For example, the steady state value of the level of output y_t is denoted as $\bar{y}$. Logged deviations of variables from steady state values are denoted using tildes; e.g.,

$$\tilde{y}_t = \log\left(\frac{y_t}{\bar{y}}\right).$$

The vector x_t denotes the collection of model variables, written (unless indicated otherwise) in terms of logged deviations from steady state values; e.g.,

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

The vector υ_t denotes the collection of structural shocks incorporated in the model, and η_t denotes the collection of expectational errors associated with intertemporal optimality conditions. Finally, the $k \times 1$ vector μ denotes the collection of "deep" parameters associated with the structural model.

Log-linear approximations of structural models are represented as

$$Ax_{t+1} = Bx_t + C\upsilon_{t+1} + D\eta_{t+1}, \tag{1.1}$$

where the elements of the matrices A, B, C, and D are functions of the structural parameters μ. Solutions of (1.1) are expressed as

$$x_{t+1} = F(\mu)x_t + G(\mu)\upsilon_{t+1}. \tag{1.2}$$

In (1.2), 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

$$X_t = H(\mu)'x_t + u_t, \tag{1.3}$$

with

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

The presence of u_t in (1.3) reflects the possibility that the observations of X_t are associated with measurement error. Finally, defining

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

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

$$Q(\mu) = E(e_t e_t'). \tag{1.4}$$

Given assumptions regarding the stochastic nature of measurement errors and the structural shocks, (1.2)–(1.4) yield a log-likelihood function $\log L(X|\Lambda)$, where Λ collects the parameters in $F(\mu)$, $H(\mu)$, Σ_u, and $Q(\mu)$. Often, it will be convenient to take as granted mappings from μ to F, H, Σ_u, and Q. In such cases the likelihood function will be written as $L(X|\mu)$.

Nonlinear approximations of structural models are represented using three equations, written with variables expressed in terms of levels. The first characterizes the evolution of the state variables s_t included in the model:

$$s_t = f(s_{t-1}, \upsilon_t), \tag{1.5}$$

where once again υ_t denotes the collection of structural shocks incorporated in the model. The second equation is known as a policy function,

which represents the optimal specification of the control variables c_t included in the model as a function of the state variables:

$$c_t = c(s_t). \tag{1.6}$$

The third equation maps the full collection of model variables into the observables:

$$X_t = \tilde{g}(s_t, c_t, \upsilon_t, u_t) \tag{1.7}$$

$$\equiv g(s_t, u_t), \tag{1.8}$$

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

The next chapter has two objectives. First, it outlines procedures for mapping nonlinear systems into (1.1). Next, it presents various solution methods for deriving (1.2), given (1.1).

Chapter 2

Approximating and Solving DSGE Models

EMPIRICAL INVESTIGATIONS involving DSGE models invariably require the completion of two preparatory stages. One stage involves preparation of the model to be analyzed, which is the focus of this chapter. The other involves preparation of the data, which is the focus of chapter 3.

Regarding the model-preparation stage, DSGE models typically include three components: a characterization of the environment in which decision makers reside, a set of decision rules that dictate their behavior, and a characterization of the uncertainty they face in making decisions. Collectively, these components take the form of a nonlinear system of expectational difference equations. Such systems are not directly amenable to empirical analysis, but can be converted into empirically implementable systems through the completion of the general two-step process outlined in this chapter.

The first step involves the construction of a linear approximation of the model. Just as nonlinear equations may be approximated linearly via the use of Taylor series expansions, so too may nonlinear systems of expectational difference equations. The second step involves the solution of the resulting linear approximation of the system. The solution is written in terms of variables expressed as deviations from steady state values, and is directly amenable to empirical implementation.

Although this chapter is intended to be self-contained, far more detail is provided in the literature cited below. Here, the goal is to impart an intuitive understanding of the model-preparation stage, and to provide guidance regarding its implementation. But we acknowledge that there are alternatives to the particular approaches to model approximation and solution presented in this chapter. Three such alternatives are provided by projection methods, value-function iterations, and policy-function iterations, which are presented in chapter 10. (For additional textbook discussions, see Judd, 1998; Adda and Cooper, 2003; Ljungqvist and Sargent, 2004; Heer and Maussner, 2005; and Canova, 2006.) These alternatives yield nonlinear approximations of the model under investigation. Details regarding empirical implementations based on nonlinear approximations are provided in chapter 11. In addition, a leading alternative to model approximation is provided by perturbation methods; for a textbook discussion see Judd (1998).

2.1 Linearization

2.1.1 Taylor Series Approximation

Consider the following n-equation system of nonlinear difference equations:

$$\Psi(z_{t+1}, z_t) = 0, \tag{2.1}$$

where the z's and 0 are $n \times 1$ vectors, and the z's represent variables expressed in levels. The parameters of the system are contained in the vector μ. DSGE models are typically represented in terms of such a system, augmented to include sources of stochastic behavior. We abstract from the stochastic component of the model in the linearization stage, because models are typically designed to incorporate stochastic behavior directly into the linearized system (a modest example is provided in section 2.2; detailed examples are provided in chapter 5). Also, whereas expectational terms are typically included among the variables in z (e.g., variables of the form $E_t(z_{t+j})$, where E_t is the conditional expectations operator), these are not singled out at this point, because they receive no special treatment in the linearization stage.

Before proceeding, note that although (2.1) 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^{th}-order equation

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

can be written in first-order form as

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

or more compactly, as

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

Thus (2.1) is sufficiently general to characterize a system of arbitrary order.

The goal of the linearization step is to convert (2.1) into a linear system, which can then be solved using any of the procedures outlined below. The reason for taking this step is that explicit solutions to (2.1) are typically unavailable, rendering quantitative assessments of the system as problematic. For textbook discussions of the analysis of nonlinear systems, see Azariadis (1993) and Sedaghat (2003).

Anticipating the notation that follows in section 2.2, the form for the system we seek is given by

$$Ax_{t+1} = Bx_t, \tag{2.2}$$

where x_t represents a transformation of z_t. Denoting the steady state of the system as $\Psi(\bar{z}) = 0$, where $\bar{z}$ is understood to be a function of μ, linearization is accomplished via a first-order Taylor Series approximation of (2.1) around its steady state, given by

$$0 \approx \Psi(\bar{z}) + \frac{\partial\Psi}{\partial z_t}(\bar{z}) \times (z_t - \bar{z}) + \frac{\partial\Psi}{\partial z_{t+1}}(\bar{z}) \times (z_{t+1} - \bar{z}), \tag{2.3}$$

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^{th} equation in (2.1) with respect to the j^{th} element of z_t. Defining

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

yields (2.2), where variables are expressed as deviations from steady state values. (It is also possible to work with higher-order approximations of (2.2); e.g., see Schmitt-Grohé and Uribe, 2002.)

2.1.2 Logarithmic Approximations

It is often useful to work with log-linear approximations of (2.1), due to their ease of interpretation. For illustration, we begin with a simple example in which the system is 1×1, and can be written as

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

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

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

Then approximating,

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

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

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

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

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

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

Returning to the $n \times 1$ case, rewrite (2.1) as

$$\Psi_1(z_{t+1}, z_t) = \Psi_2(z_{t+1}, z_t), \tag{2.4}$$

because it is not possible to take logs of both sides of (2.1). Again using $z_t = e^{\log z_t}$, taking logs of (2.4) and rearranging yields

$$\log \Psi_1(e^{\log z_{t+1}}, e^{\log z_t}) - \log \Psi_2(e^{\log z_{t+1}}, e^{\log z_t}) = 0. \tag{2.5}$$

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

$$\log \Psi_1(z_{t+1}, z_t) \approx \log[\Psi_1(\bar{z})] + \frac{\partial \log[\Psi_1]}{\partial \log(z_t)}(\bar{z}) \times \left[\log\left(\frac{z_t}{\bar{z}}\right)\right] + \frac{\partial \log[\Psi_1]}{\partial \log(z_{t+1})}(\bar{z}) \times \left[\log\left(\frac{z_{t+1}}{\bar{z}}\right)\right], \tag{2.6}$$

where $\frac{\partial \log[\Psi_1]}{\partial \log(z_t)}(\bar{z})$ and $\frac{\partial \log[\Psi_1]}{\partial \log(z_{t+1})}(\bar{z})$ are $n \times n$ Jacobian matrices, and $\left[\log\left(\frac{z_t}{\bar{z}}\right)\right]$ and $\left[\log\left(\frac{z_{t+1}}{\bar{z}}\right)\right]$ are $n \times 1$ vectors. The approximation of the second term in (2.5) is analogous. Then defining

$$A = \left[\frac{\partial \log[\Psi_1]}{\partial \log(z_{t+1})}(\bar{z}) - \frac{\partial \log[\Psi_2]}{\partial \log(z_{t+1})}(\bar{z})\right],$$

$$B = -\left[\frac{\partial \log[\Psi_1]}{\partial \log(z_t)}(\bar{z}) - \frac{\partial \log[\Psi_2]}{\partial \log(z_t)}(\bar{z})\right],$$

$$x_t = \log\left(\frac{z_t}{\bar{z}}\right),$$

we once again obtain (2.2). 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 II we will discuss several empirical applications that involve the need to approximate (2.1) or (2.5) repeatedly for alternative values of μ. 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 (2.1); the approximation of (2.5) is analogous.

Gradient procedures are designed to construct the Jacobian matrices in (2.3) or (2.6) 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 μ and a corresponding specification of $\bar{z}$. The third component is a procedure designed to return the $n \times 1$ vector of values ς generated by (2.1) 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. Examples follow.

2.1.3 *Examples*

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 (2.1) the equation appears as

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

and in terms of (2.3), with

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

and the equation representing the i^{th} of the system, the i^{th} row of

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

In the notation of (2.5), the equation appears as

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

and in terms of (2.2), the i^{th} row of the right-hand-side matrix B is

$$\left[\frac{\partial \log[\Psi_1]}{\partial \log(z_t)}(\overline{z}) - \frac{\partial \log[\Psi_2]}{\partial \log(z_t)}(\overline{z})\right] = \left[\frac{1}{\overline{y}} \quad \frac{-\overline{c}}{\overline{c}+\overline{i}} \quad \frac{-\overline{i}}{\overline{c}+\overline{i}}\right]. \tag{2.7}$$

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

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

As an additional example consider the Cobb-Douglas production function

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

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, taking logs of the equation and rearranging maps into the notation of (2.5) as

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

With

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

the i^{th} row of the right-hand-side matrix in (2.2) is

$$\left[\frac{\partial \log[\Psi_1]}{\partial \log(z_t)}(\overline{z}) - \frac{\partial \log[\Psi_2]}{\partial \log(z_t)}(\overline{z})\right] = [1 \;\; -1 \;\; -\alpha \;\; -(1-\alpha)]. \tag{2.8}$$

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

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

2.2 Solution Methods

Having approximated the model as in (2.2), we next seek a solution of the form

$$x_{t+1} = Fx_t + G\upsilon_{t+1}. \tag{2.9}$$

This solution represents the time series behavior of $\{x_t\}$ as a function of $\{\upsilon_t\}$, where υ_t is a vector of exogenous innovations, or as frequently referenced, structural shocks.

Here we present four popular approaches to the derivation of (2.9) from (2.2). Each approach involves an alternative way of expressing (2.2), and uses specialized notation. Also, each approach makes intensive use of linear algebra: background for this material is available at the undergraduate level, for example, from Lay (2002); and at the graduate level, for example, from Roman (2005).

Before describing these approaches, we introduce an explicit example of (2.2), which we will map into the notation used under each approach to aid with the exposition. The example is a linearized stochastic version of Ramsey's (1928) optimal growth model. (See, e.g., Romer, 2006, for a detailed textbook exposition.) The model is represented as:

$$\widetilde{y}_{t+1} - \widetilde{a}_{t+1} - \alpha\widetilde{k}_{t+1} = 0 \tag{2.10}$$

$$\widetilde{y}_{t+1} - \gamma_c\widetilde{c}_{t+1} - \gamma_i\widetilde{i}_{t+1} = 0 \tag{2.11}$$

$$\theta_{1c}E_t(\widetilde{c}_{t+1}) + \theta_a E_t(\widetilde{a}_{t+1}) + \theta_k E_t(\widetilde{k}_{t+1}) + \theta_{2c}\widetilde{c}_t = 0 \tag{2.12}$$

$$\widetilde{k}_{t+1} - \delta_k\widetilde{k}_t - \delta_i\widetilde{i}_t = 0 \tag{2.13}$$

$$\widetilde{a}_{t+1} - \rho\widetilde{a}_t = \varepsilon_{t+1}. \tag{2.14}$$

The variables $\{\widetilde{y}_t, \widetilde{c}_t, \widetilde{i}_t, \widetilde{k}_t, \widetilde{a}_t\}$ represent output, consumption, investment, physical capital, and a productivity shock, all expressed as logged deviations from steady state values. The variable ε_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.

Two modifications enable a mapping of the model into a specification resembling (2.2). First, the expectations operator $E_t(\cdot)$ is dropped from (2.12), introducing an expectational error into the modified equation; let

this error be denoted as η_{ct+1}. Next, the innovation term ε_{t+1} in (2.14) must be accommodated. The resulting expression is

$$\underbrace{\begin{bmatrix} 1 & 0 & 0 & -\alpha & -1 \\ 1 & -\gamma_c & -\gamma_i & 0 & 0 \\ 0 & \theta_{1c} & 0 & \theta_k & \theta_a \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{A} \underbrace{\begin{bmatrix} \widetilde{y}_{t+1} \\ \widetilde{c}_{t+1} \\ \widetilde{i}_{t+1} \\ \widetilde{k}_{t+1} \\ \widetilde{a}_{t+1} \end{bmatrix}}_{x_{t+1}}$$

$$= \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & -\theta_{2c} & 0 & 0 & 0 \\ 0 & 0 & \delta_i & \delta_k & 0 \\ 0 & 0 & 0 & 0 & \rho \end{bmatrix}}_{B} \underbrace{\begin{bmatrix} \widetilde{y}_t \\ \widetilde{c}_t \\ \widetilde{i}_t \\ \widetilde{k}_t \\ \widetilde{a}_t \end{bmatrix}}_{x_t}$$

$$+ \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}}_{C} \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \\ \varepsilon_{t+1} \end{bmatrix}}_{v_{t+1}}$$

$$+ \underbrace{\begin{bmatrix} 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}}_{D} \underbrace{\begin{bmatrix} 0 \\ 0 \\ 0 \\ \eta_{ct+1} \\ 0 \end{bmatrix}}_{\eta_{t+1}}. \tag{2.15}$$

2.2.1 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

$$\begin{bmatrix} x_{1t+1} \\ E_t(x_{2t+1}) \end{bmatrix} = \widetilde{A} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} + Ef_t, \tag{2.16}$$

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} (for which $x_{2t+1} = E_t(x_{2t+1}) + \eta_{t+1}$, with η_{t+1} representing an expectational error). The $k \times 1$ vector f_t contains exogenous forcing variables.

In the event that the linearization of the model under investigation does not automatically lend itself to the form given by (2.16), a preliminary step due to King and Watson (2002) may be implemented. 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 $\widetilde{a}_t$ and $\widetilde{k}_t$ are sufficient for determining $\widetilde{y}_t$ using (2.10), and that given $\widetilde{y}_t$, the observation of either $\widetilde{c}_t$ or $\widetilde{i}_t$ is sufficient for determining both variables using (2.11). Thus we proceed in working directly with $\{\widetilde{c}_t, \widetilde{k}_t, \widetilde{a}_t\}$ using (2.12)–(2.14), and recover $\{\widetilde{y}_t, \widetilde{i}_t\}$ as functions of $\{\widetilde{c}_t, \widetilde{k}_t, \widetilde{a}_t\}$ using (2.10) and (2.11). Among $\{\widetilde{c}_t, \widetilde{k}_t, \widetilde{a}_t\}$, $\widetilde{k}_t$ is predetermined (given k_t and i_t, $\widetilde{k}_{t+1}$ is determined as in (2.13)); $\widetilde{c}_t$ is endogenous but not predetermined (as indicated in (2.12), its time-$(t+1)$ realization is associated with an expectations error); and $\widetilde{a}_t$ is an exogenous forcing variable. Thus in the notation of (2.16), we seek a specification of the model in the form

$$\begin{bmatrix} \widetilde{k}_{t+1} \\ E_t(\widetilde{c}_{t+1}) \end{bmatrix} = \widetilde{A} \begin{bmatrix} \widetilde{k}_t \\ \widetilde{c}_t \end{bmatrix} + E\widetilde{a}_t. \tag{2.17}$$

To obtain this expression, let

$$\xi_t = \begin{bmatrix} \widetilde{y}_t & \widetilde{i}_t \end{bmatrix}', \qquad \zeta_t = \begin{bmatrix} \widetilde{k}_t & \widetilde{c}_t \end{bmatrix}',$$

and note that

$$E_t(\widetilde{a}_{t+1}) = \rho \widetilde{a}_t.$$

In terms of these variables, the model may be written as

$$\underbrace{\begin{bmatrix} 1 & 0 \\ 1 & -\gamma_i \end{bmatrix}}_{\Psi_0} \xi_t = \underbrace{\begin{bmatrix} \alpha & 0 \\ 0 & \gamma_c \end{bmatrix}}_{\Psi_1} \zeta_t + \underbrace{\begin{bmatrix} 1 \\ 0 \end{bmatrix}}_{\Psi_2} \widetilde{a}_t \tag{2.18}$$

$$\underbrace{\begin{bmatrix} \theta_k & \theta_{1c} \\ 1 & 0 \end{bmatrix}}_{\Psi_3} E_t(\zeta_{t+1}) = \underbrace{\begin{bmatrix} 0 & -\theta_{2c} \\ \delta_k & 0 \end{bmatrix}}_{\Psi_4} \zeta_t + \underbrace{\begin{bmatrix} 0 & 0 \\ 0 & \delta_i \end{bmatrix}}_{\Psi_5} \xi_t + \underbrace{\begin{bmatrix} -\theta_a \rho \\ 0 \end{bmatrix}}_{\Psi_6} \widetilde{a}_t. \tag{2.19}$$

Next, substituting (2.18) into (2.19), which requires inversion of Ψ_0, we obtain

$$\Psi_3 E_t(\zeta_{t+1}) = [\Psi_4 + \Psi_5\Psi_0^{-1}\Psi_1]\zeta_t + [\Psi_6 + \Psi_5\Psi_0^{-1}\Psi_2]\widetilde{a}_t. \quad (2.20)$$

Finally, premultiplying (2.20) by Ψ_3^{-1} yields a specification in the form of (2.17); Blanchard and Kahn's solution method can now be implemented. Hereafter, we describe its implementation in terms of the notation used in (2.16).

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

$$\widetilde{A} = \Lambda^{-1}J\Lambda, \quad (2.21)$$

where the diagonal elements of J, consisting of the eigenvalues of $\widetilde{A}$, are ordered in increasing absolute value in moving from left to right.[1] Thus J can be written as

$$J = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix}, \quad (2.22)$$

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, because J_2^n diverges as n increases. The matrices Λ and E are partitioned conformably as

$$\Lambda = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix}, \quad E = \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}, \quad (2.23)$$

where Λ_{11} is conformable with J_1, etc. 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 exist (and the system is said to be a sink).

Proceeding under the case of saddle-path stability, substitution for $\widetilde{A}$ in (2.16) yields

$$\begin{bmatrix} x_{1t+1} \\ E_t(x_{2t+1}) \end{bmatrix} = \Lambda^{-1}J\Lambda \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} + \begin{bmatrix} E_1 \\ E_2 \end{bmatrix} f_t. \quad (2.24)$$

[1] Eigenvalues of a matrix Θ are obtained from the solution of equations of the form $\Theta e = \lambda e$, where e is an eigenvector and λ the associated eigenvalue. The GAUSS command `eigv` performs this decomposition.

Next, the system is premultiplied by Λ, yielding

$$\begin{bmatrix} \acute{x}_{1t+1} \\ E_t(\acute{x}_{2t+1}) \end{bmatrix} = \begin{bmatrix} J_1 & 0 \\ 0 & J_2 \end{bmatrix} \begin{bmatrix} \acute{x}_{1t} \\ \acute{x}_{2t} \end{bmatrix} + \begin{bmatrix} D_1 \\ D_2 \end{bmatrix} f_t, \quad (2.25)$$

where

$$\begin{bmatrix} \acute{x}_{1t} \\ \acute{x}_{2t} \end{bmatrix} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \begin{bmatrix} x_{1t} \\ x_{2t} \end{bmatrix} \quad (2.26)$$

$$\begin{bmatrix} D_1 \\ D_2 \end{bmatrix} = \begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ \Lambda_{21} & \Lambda_{22} \end{bmatrix} \begin{bmatrix} E_1 \\ E_2 \end{bmatrix}. \quad (2.27)$$

This transformation effectively "de-couples" the system, so that the nonpredetermined variables depend upon only the unstable eigenvalues of $\widetilde{A}$ contained in J_2, as expressed in the lower part of (2.25).

Having decoupled the system, we derive a solution for the nonpredetermined variables by performing a forward iteration on the lower portion of (2.25). Using f_{2t} to denote the portion of f_t conformable with D_2, this is accomplished as follows. First, re-express the lower portion of (2.25) as

$$\acute{x}_{2t} = J_2^{-1}E_t(\acute{x}_{2t+1}) - J_2^{-1}D_2 f_{2t}. \quad (2.28)$$

This implies an expression for $\acute{x}_{2t+1}$ of the form

$$\acute{x}_{2t+1} = J_2^{-1}E_{t+1}(\acute{x}_{2t+2}) - J_2^{-1}D_2 f_{2t+1}, \quad (2.29)$$

which can be substituted into (2.28) to obtain

$$\acute{x}_{2t} = J_2^{-2}E_t(\acute{x}_{2t+2}) - J_2^{-2}D_2E_t(f_{2t+1}) - J_2^{-1}D_2 f_{2t}. \quad (2.30)$$

In writing (2.30) 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). Because J_2 contains explosive eigenvalues, J_2^{-n} disappears as n approaches infinity, thus continuation of the iteration process yields

$$\acute{x}_{2t} = -\sum_{i=0}^{\infty} J_2^{-(i+1)}D_2E_t(f_{2t+i}). \quad (2.31)$$

Mapping this back into an expression for x_{2t} using (2.26), we obtain

$$x_{2t} = -\Lambda_{22}^{-1}\Lambda_{21}x_{1t} - \Lambda_{22}^{-1}\sum_{i=0}^{\infty} J_2^{-(i+1)}D_2E_t(f_{2t+i}). \quad (2.32)$$

In the case of the example model presented above,

$$E_t(f_{2t+i}) = \rho^i \tilde{a}_t,$$

and thus (2.32) becomes

$$x_{2t} = -\Lambda_{22}^{-1}\Lambda_{21}x_{1t} - \Lambda_{22}^{-1}J_2^{-1}(I - \rho J_2^{-1}D_2)^{-1}\tilde{a}_t. \tag{2.33}$$

Finally, to solve the nonexplosive portion of the system begin by expanding the upper portion of (2.24):

$$x_{1t+1} = \tilde{A}_{11}x_{1t} + \tilde{A}_{22}x_{2t} + E_1 f_t, \tag{2.34}$$

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 (2.32) yields a solution for x_{1t} of the form given by (2.9).

We conclude this subsection by highlighting two requirements of this solution method. First, a model-specific system reduction may be required 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 non-predetermined. Second, invertibility of the lead matrices Ψ_0 and Ψ_3 is required in order to obtain a specification of the model amenable for solution.

Exercise 2.1

Write computer code for mapping the example model expressed in (2.10)–(2.14) into the form of the representation given in (2.16).

2.2.2 *Sims's Method*

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

$$Ax_{t+1} = Bx_t + E + C\upsilon_{t+1} + D\eta_{t+1}, \tag{2.35}$$

where E is a matrix of constants.[2] Relative to the notation we have used 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 η_{t+1}. Also, although 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 the variables υ_{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.

[2] The programs available on Sims's Web site perform all of the steps of this procedure. The Web address is: http://www.princeton.edu/~sims/. The programs are written in Matlab.

Note from (2.15) that the example model has already been cast in the form of (2.35); thus we proceed directly to a characterization of the solution method. The first step uses a "QZ factorization" to decompose A and B into unitary upper triangular matrices:

$$A = Q'\Lambda Z' \tag{2.36}$$

$$B = Q'\Omega Z', \tag{2.37}$$

where (Q, Z) are unitary, and (Λ, Ω) are upper triangular. (A unitary matrix Θ satisfies $\Theta'\Theta = \Theta\Theta' = I$. If Q and/or Z contain complex values, the transpositions reflect complex conjugation; that is, each complex entry is replaced by its conjugate and then transposed.) Next, (Q, Z, Λ, Ω) are ordered such that, in absolute value, the generalized eigenvalues of A and B are organized in Λ and Ω in increasing order moving from left to right, just as in Blanchard and Kahn's Jordan decomposition procedure. (Generalized eigenvalues of Θ are obtained as the solution to $\Theta e = \lambda \Xi e$, where Ξ is a symmetric matrix.) Having obtained the factorization, the original system is then premultiplied by Q, yielding the transformed system expressed in terms of $z_{t+1} = Z'x_{t+1}$:

$$\Lambda z_t = \Omega z_{t-1} + QE + QC\upsilon_t + QD\eta_t, \tag{2.38}$$

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, (2.38) is partitioned into explosive and nonexplosive blocks:

$$\begin{bmatrix} \Lambda_{11} & \Lambda_{12} \\ 0 & \Lambda_{22} \end{bmatrix} \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} = \begin{bmatrix} \Omega_{11} & \Omega_{12} \\ 0 & \Omega_{22} \end{bmatrix} \begin{bmatrix} z_{1t-1} \\ z_{2t-1} \end{bmatrix} + \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} [E + C\upsilon_t + D\eta_t]. \tag{2.39}$$

The explosive block (the lower equations) is solved as follows. Letting

$$w_t = Q(E + C\upsilon_t + D\eta_t),$$

partitioned conformably as w_{1t} and w_{2t}, the lower block of (2.39) is given by

$$\Lambda_{22}z_{2t} = \Omega_{22}z_{2t-1} + w_{2t}. \tag{2.40}$$

Leading (2.40) by one period and solving for z_{2t} yields

$$z_{2t} = Mz_{2t+1} - \Omega_{22}^{-1} w_{2t+1}, \tag{2.41}$$

where

$$M = \Omega_{22}^{-1} \Lambda_{22}.$$

Then recursive substitution for $z_{2t+1}, z_{2t+2}, \ldots$ yields

$$z_{2t} = -\sum_{i=0}^{\infty} M^i \Omega_{22}^{-1} w_{2t+1+i}, \tag{2.42}$$

since

$$\lim_{t \to \infty} M^t z_{2t} = 0.$$

Recalling that w_t is defined as $w_t = Q(E + C\upsilon_t + D\eta_t)$, note that (2.42) expresses z_{2t} as a function of future values of structural and expectational errors. But z_{2t} is known at time t, and

$$E_t(\eta_{t+s}) = E_t(\upsilon_{t+s}) = 0, \qquad s > 0,$$

thus (2.42) may be written as

$$z_{2t} = -\sum_{i=0}^{\infty} M^i \Omega_{22}^{-1} Q_2 E_2, \tag{2.43}$$

where $Q_2 E_2$ are the lower portions of QE conformable with z_2. (Sims also considers the case in which the structural innovations υ_t are serially correlated, which leads to a generalization of (2.43).) Postmultiplying (2.43) by $(\Omega_{22}^{-1} Q_2 E_2)^{-1}$ and noting that

$$-\sum_{i=0}^{\infty} M^i = -(I - M)^{-1},$$

the solution of z_{2t} is obtained as

$$z_{2t} = (\Lambda_{22} - \Omega_{22})^{-1} Q_2 E. \tag{2.44}$$

Having solved for z_{2t}, the final step is to solve for z_{1t} in (2.39). Note that the solution of z_{1t} requires a solution for the expectations errors that appear in (2.39). As Sims notes, when a unique solution for the model exists, a systematic relationship will exist between the expectations 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 Φ that satisfies

$$Q_1 D = \Phi Q_2 D, \tag{2.45}$$

which represents the systematic relationship between the expectations errors associated with z_{1t} and z_{2t} noted above. Given uniqueness, and thus the ability to calculate Φ as in (2.45), the solution of z_{1t} proceeds with the pre-multiplication of (2.38) by $[I - \Phi]$, which yields

$$\begin{aligned} &[\,\Lambda_{11} \quad \Lambda_{12} - \Phi\Lambda_{22}\,] \begin{bmatrix} z_{1t} \\ z_{2t} \end{bmatrix} \\ &\quad = [\,\Omega_{11} \quad \Omega_{12} - \Phi\Omega_{22}\,] \begin{bmatrix} z_{1t-1} \\ z_{2t-1} \end{bmatrix} + [Q_1 - \Phi Q_2][E + C\upsilon_t + D\eta_t]. \end{aligned} \tag{2.46}$$

Then due to (2.45), the loading factor for the expectational errors in (2.46) is zero, and thus the system may be written in the form

$$x_t = \Theta_E + \Theta_0 x_{t-1} + \Theta_1 \upsilon_t, \tag{2.47}$$

where

$$H = Z \begin{bmatrix} \Lambda_{11}^{-1} & -\Lambda_{11}^{-1}(\Lambda_{12} - \Phi\Lambda_{22}) \\ 0 & I \end{bmatrix} \tag{2.48}$$

$$\Theta_E = H \begin{bmatrix} Q_1 - \Phi Q_2 \\ (\Omega_{22} - \Lambda_{22})^{-1} Q_2 \end{bmatrix} E \tag{2.49}$$

$$\Theta_0 = Z\Lambda_{11}^{-1}[\Omega_{11}(\Omega_{12} - \Phi\Omega_{22})]Z' \tag{2.50}$$

$$\Theta_1 = H \begin{bmatrix} Q_1 - \Phi Q_2 \\ 0 \end{bmatrix} D. \tag{2.51}$$

Exercise 2.2

Using the code cited for this method, compute the solution for (2.10)–(2.14) for given values of μ.

2.2.3 *Klein's Method*

Klein (2000) proposes a solution method that is a hybrid of those of Blanchard and Kahn (1980) and Sims (2001).[3] The method is applied to systems written as

$$\widetilde{A} E_t(x_{t+1}) = \widetilde{B} x_t + E f_t, \tag{2.52}$$

[3] GAUSS and Matlab code that implement this solution method are available at http://www.ssc.uwo.ca/economics/faculty/klein/

where the vector f_t (of length n_z) has a zero-mean vector autoregressive (VAR) specification with autocorrelation matrix Φ; additionally $\widetilde{A}$ may be singular. (See chapter 4, section 4.1.2, for a description of VAR models.)

Like Blanchard and Kahn, Klein distinguishes between the predetermined and nonpredetermined variables of the model. The former are contained in x_{1t+1}, the latter in x_{2t+1}:

$$E_t(x_{t+1}) = [x_{1t+1} \quad E_t(x_{2t+1})]'.$$

The solution approach once again involves de-coupling the system into nonexplosive and explosive components, and solving the two components in turn.

Returning to the example model expressed in (2.10)–(2.14), the form of the model amenable to the implementation of Klein's method is given by (2.20), repeated here for convenience:

$$\Psi_3 E_t(\zeta_{t+1}) = [\Psi_4 + \Psi_5\Psi_0^{-1}\Psi_1]\zeta_t + [\Psi_6 + \Psi_5\Psi_0^{-1}\Psi_2]\widetilde{a}_t. \tag{2.53}$$

An advantage of Klein's approach relative to Blanchard and Kahn's is that Ψ_3 may be singular. In addition, it is also faster to implement computationally. To proceed with the description of Klein's approach, we revert to the notation used in (2.52).

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

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

$$Q\widetilde{A}Z = S \tag{2.54}$$

$$Q\widetilde{B}Z = T, \tag{2.55}$$

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

$$Z = \begin{bmatrix} Z_{11} & Z_{12} \\ Z_{21} & Z_{22} \end{bmatrix}, \tag{2.56}$$

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 (2.52) as was done in working with the QZ decomposition. Begin by defining

$$z_t = Z^H x_t, \tag{2.57}$$

where Z^H refers to a Hermitian transpose. (Given a matrix Θ, if the lower triangular portion of Θ is the complex conjugate transpose of the upper triangle portion of Θ, then Θ is denoted as Hermitian.) This transformed vector is divided into $n_1 \times 1$ stable (s_t) and $n_2 \times 1$ unstable (u_t) components. Then since

$$\widetilde{A} = Q'SZ^H$$

and

$$\widetilde{B} = Q'TZ^H,$$

(2.52) may be written as

$$\begin{bmatrix} S_{11} & S_{12} \\ 0 & S_{22} \end{bmatrix} E_t \begin{bmatrix} s_{t+1} \\ u_{t+1} \end{bmatrix} = \begin{bmatrix} T_{11} & T_{12} \\ 0 & T_{22} \end{bmatrix} \begin{bmatrix} s_t \\ u_t \end{bmatrix} + \begin{bmatrix} Q_1 \\ Q_2 \end{bmatrix} Ef_t; \tag{2.58}$$

once again, the lower portion of (2.58) contains the unstable components of the system. Solving this component via forward iteration, we obtain

$$u_t = Mf_t \tag{2.59}$$

$$vec(M) = [(\Phi^T \otimes S_{22}) - I_{n_z} \otimes T_{22}]^{-1} vec(Q_2 E). \tag{2.60}$$

The appearance of the *vec* operator accommodates the VAR specification for f_t. In the context of the example model, Φ^T is replaced by the scalar ρ^T, and (2.60) 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

$$s_{t+1} = S_{11}^{-1} T_{11} s_t + S_{11}^{-1}\{T_{12}M - S_{12}M\Phi + Q_1 E\}f_t - Z_{11}^{-1} Z_{12} M \upsilon_{t+1}, \tag{2.61}$$

where υ_{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 $\widetilde{a}_t$, the innovation to which is ε_t. In terms of the original variables the solution is expressed as

$$x_{2t} = Z_{21}Z_{11}^{-1}x_{1t} + Nf_t \tag{2.62}$$

$$x_{1t+1} = Z_{11}S_{11}^{-1}T_{11}Z_{11}^{-1}x_{2t} + Lf_t \tag{2.63}$$

$$N = (Z_{22} - Z_{21}Z_{11}^{-1}Z_{12})M \tag{2.64}$$

$$L = -Z_{11}S_{11}^{-1}T_{11}Z_{11}^{-1}Z_{12}M + Z_{11}S_{11}^{-1} \times [T_{12}M - S_{12}M\Phi + Q_1E] + Z_{12}M\Phi. \tag{2.65}$$

This solution can be cast into the form of (2.9) as

$$x_{1t+1} = [Z_{11}S_{11}^{-1}T_{11}Z_{11}^{-1}Z_{21}Z_{11}^{-1}]x_{1t} + [Z_{11}S_{11}^{-1}T_{11}Z_{11}^{-1}N + L]f_t. \tag{2.66}$$

Exercise 2.3

Apply Klein's code to the example model presented in (2.10)–(2.14).

2.2.4 *An Undetermined Coefficients Approach*

Uhlig (1999) proposes a solution method based on the method of undetermined coefficients.[4] The method is applied to systems written as

$$0 = E_t[Fx_{t+1} + Gx_t + Hx_{t-1} + Lf_{t+1} + Mf_t] \tag{2.67}$$

$$f_{t+1} = Nf_t + \upsilon_{t+1}, \qquad E_t(\upsilon_{t+1}) = 0. \tag{2.68}$$

With respect to the example model in (2.10)–(2.14), let

$$x_t = [\widetilde{y}_t \ \widetilde{c}_t \ \widetilde{i}_t \ \widetilde{k}_t]'.$$

Then lagging the first two equations, which are subject neither to structural shocks nor expectations errors, the matrices in (2.67) and (2.68) are given by

$$F = \begin{bmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 \\ 0 & \theta_{1c} & 0 & \theta_k \\ 0 & 0 & 0 & 1 \end{bmatrix}, \quad G = \begin{bmatrix} 1 & 0 & 0 & -\alpha \\ 1 & -\gamma_c & -\gamma_i & 0 \\ 0 & \theta_{2c} & 0 & 0 \\ 0 & 0 & -\delta_i & -\delta_k \end{bmatrix}, \tag{2.69}$$

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

[4] Matlab code available for implementing this solution method is available at: http://www.wiwi.hu-berlin.de/wpol/html/toolkit.htm

Solutions to (2.67)–(2.68) take the form

$$x_t = Px_{t-1} + Qf_t. \tag{2.70}$$

In deriving (2.70), we will confront the problem of solving matrix quadratic equations of the form

$$\Psi P^2 - \Gamma P - \Theta = 0 \tag{2.71}$$

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

To begin, define

$$\underset{2m\times 2m}{\Xi} = \begin{bmatrix} \Gamma & \Theta \\ I_m & 0_{m\times m} \end{bmatrix}, \quad \underset{2m\times 2m}{\Delta} = \begin{bmatrix} \Psi & 0_{m\times m} \\ 0_{m\times m} & I_m \end{bmatrix}. \tag{2.72}$$

Given these matrices, let s and λ denote the generalized eigenvector and eigenvalue of Ξ with respect to Δ, and note that $s' = [\lambda x', x']$ for some $x \in \Re^m$. Then the solution to the matrix quadratic is given by

$$\begin{aligned} P &= \Omega\Lambda\Omega^{-1}, \\ \Omega &= [x_1, \ldots, x_m], \\ \Lambda &= diag(\lambda_1, \ldots, \lambda_m), \end{aligned} \tag{2.73}$$

so long as the m eigenvalues contained in Λ and $(x_1, \ldots, x_m)$ 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 (2.67)–(2.68), the first step towards obtaining (2.70) is to combine these three equations into a single equation. This is accomplished in two steps. First, write x_t in (2.67) in terms of its relationship with x_{t-1} given by (2.70), and do the same for x_{t+1}, where the relationship is given by

$$x_{t+1} = P^2x_{t-1} + PQf_t + Qf_{t+1}. \tag{2.74}$$

Next, write f_{t+1} in terms of its relationship with f_t given by (2.68). Taking expectations of the resulting equation yields

$$0 = [FP^2 + GP + H]x_{t-1} + [(FP + G)Q + M + (FQ + L)N]f_t. \tag{2.75}$$

Note that in order for (2.75) 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

$$0 = FP^2 + GP + H, \tag{2.76}$$

the solution of which is obtained as indicated in (2.72) and (2.73). The second restriction requires the derivation of Q which satisfies

$$(FP + G)Q + M + (FQ + L)N = 0. \tag{2.77}$$

The required Q can be shown to be given by

$$Q = V^{-1}\left[-vec(LN + M)\right], \tag{2.78}$$

where V is defined as

$$V = N' \otimes F + I_k \otimes (FP + G). \tag{2.79}$$

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 (2.67)–(2.68), with the expectations operator $\mathring{E}_t$ replacing E_t. In terms of calculating the expectation of an $n \times 1$ vector X_t, $\mathring{E}_t$ is defined as

$$\mathring{E}_t(X_t) = \begin{bmatrix} E(X_{1t}|\Xi_{1t}) \\ \vdots \\ E(X_{nt}|\Xi_{nt}) \end{bmatrix}; \tag{2.80}$$

where Ξ_{it} represents the information set available for formulating expectations over the i^{th} element of X_t. Thus systems involving this form of heterogeneity may be accommodated using an expansion of the system (2.67)–(2.68) specified for a representative agent. The solution of the expanded system proceeds as indicated above; for details and extensions, see Christiano (2002).

Exercise 2.4

Apply Uhlig's code to the example model presented in (2.10)–(2.14).

We conclude by repeating our acknowledgement that there are alternatives to the approaches to model approximation and solution presented in this chapter. Those interested in exploring alternatives at this point may wish to jump to chapter 10. There, we present three leading alternatives: projection methods, value-function iterations, and policy-function iterations. Each yields a nonlinear approximation of the model under investigation.

Chapter 3

Removing Trends and Isolating Cycles

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 collection of 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. Having satisfied the constraint, 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) and Kaiser and Maravall (2001).

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 5. The data are contained in the text file `rbcdata.txt`, available for downloading at the textbook Web site. 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 2004:IV. 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 3.1.

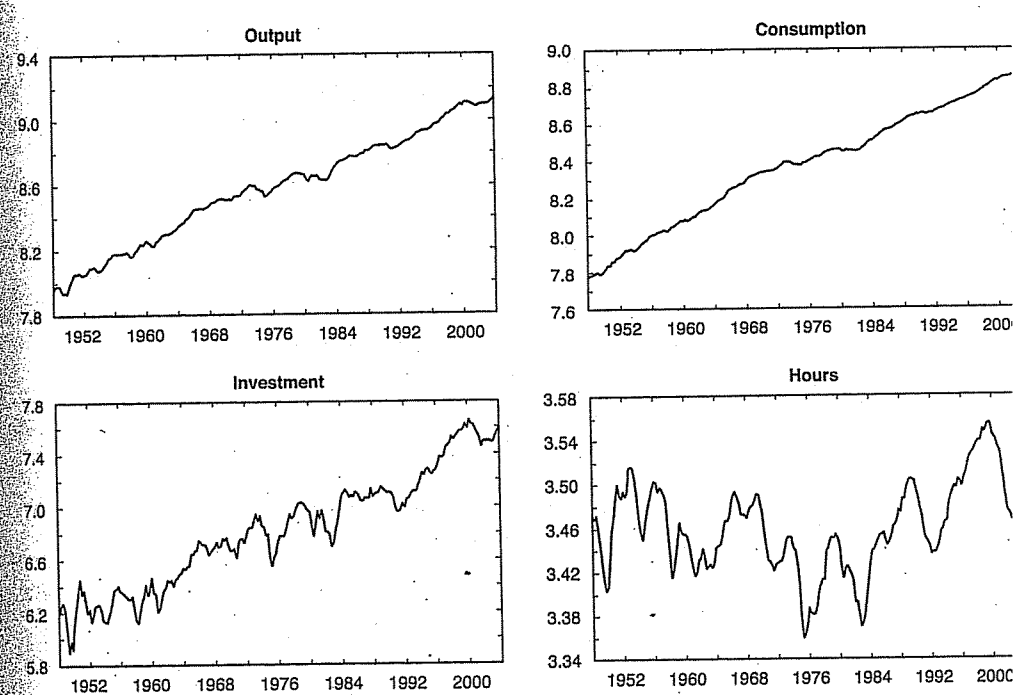

Figure 3.1 Business cycle data.

3.1 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 3.1). This is because changes in the log of a variable y_t over time represent the growth rate of the variable:

$$\frac{\partial}{\partial t}\log y_t = \frac{\frac{\partial}{\partial t}y_t}{y_t} \equiv \frac{\dot{y}_t}{y_t} \equiv g_{y_t}, \tag{3.1}$$

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

$$y_t = y_0(1+g_y)^t e^{u_t}, \qquad u_t \sim CSSP. \tag{3.2}$$

Then taking logs,

$$\log y_t = \log y_0 + g_y t + u_t, \tag{3.3}$$

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:

$$\widetilde{y}_t = \log y_t - \widehat{\alpha}_0 - \widehat{\alpha}_1 t = \widehat{u}_t, \tag{3.4}$$

where the $\widehat{\alpha}'s$ 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 α_1^j as the trend coefficient associated with variable j, this is accomplished via the imposition of the linear restrictions

$$\alpha_1^1 - \alpha_1^j = 0, \qquad j = 2, \ldots, m,$$

easily imposed in an OLS estimation framework.[1]

Differencing proceeds under the assumption that y_t obeys

$$y_t = y_0 e^{\varepsilon_t}, \tag{3.5}$$

$$\varepsilon_t = \gamma + \varepsilon_{t-1} + u_t, \qquad u_t \sim CSSP. \tag{3.6}$$

Note from (3.6) that iterative substitution for $\varepsilon_{t-1}, \varepsilon_{t-2}, \ldots$, yields an expression for ε_t of the form

$$\varepsilon_t = \gamma t + \sum_{j=0}^{t-1} u_{t-j} + \varepsilon_0, \tag{3.7}$$

and thus the growth rate of y_t is given by γ. From (3.5),

$$\log y_t = \log y_0 + \varepsilon_t. \tag{3.8}$$

Thus the first difference of $\log y_t$, given by

$$\log y_t - \log y_{t-1} \equiv (1-L)\log y_t,$$

[1] The GAUSS procedure `ct.prc`, available at the textbook Web site, serves this purpose.

where the lag operator L is defined such that $L^p y_t = y_{t-p}$, is stationary:

$$\begin{aligned} \log y_t - \log y_{t-1} &= \varepsilon_t - \varepsilon_{t-1} \\ &= \gamma + u_t. \end{aligned} \tag{3.9}$$

In this case, $\log y_t$ is said to be difference stationary. Estimating γ using the sample average of $\log y_t - \log y_{t-1}$ yields the desired transformation of y_t:

$$\widetilde{y}_t = \log y_t - \log y_{t-1} - \widehat{\gamma} = \widehat{u}_t. \tag{3.10}$$

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[2]

$$\widehat{\gamma}^1 - \widehat{\gamma}^j = 0, \qquad j = 2, \ldots, m. \tag{3.11}$$

The choice between detrending versus differencing hinges on assumptions regarding whether (3.3) or (3.9) 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, 1994.) A remedy for this difficulty is to work with both specifications in turn, and evaluate the sensitivity of results to the chosen specification.

As figure 3.2 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.

As figure 3.2 indicates, 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 recent 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 were 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 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 cyclicality.

[2] The GAUSS procedure `ct.prc` is also available for this purpose.

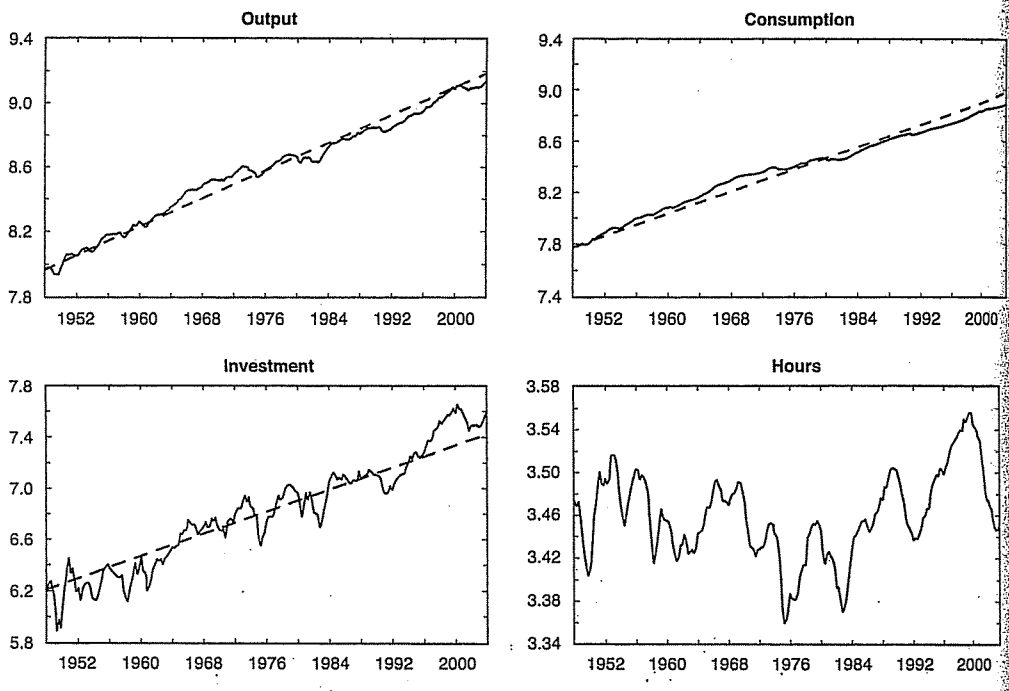

Figure 3.2 Logged trajectories and fitted trends.

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

$$\log y_t = g_t + c_t, \tag{3.12}$$

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

$$\sum_{t=1}^{T} c_t^2 + \lambda \sum_{t=3}^{T} [(1-L)^2 g_t]^2, \tag{3.13}$$

taking λ as given.[3] Trend removal is accomplished simply as

$$\widetilde{y}_t = \log y_t - \widehat{g}_t = \widehat{c}_t. \tag{3.14}$$

[3] The GAUSS procedure `hpfilter.prc` is available for this purpose.

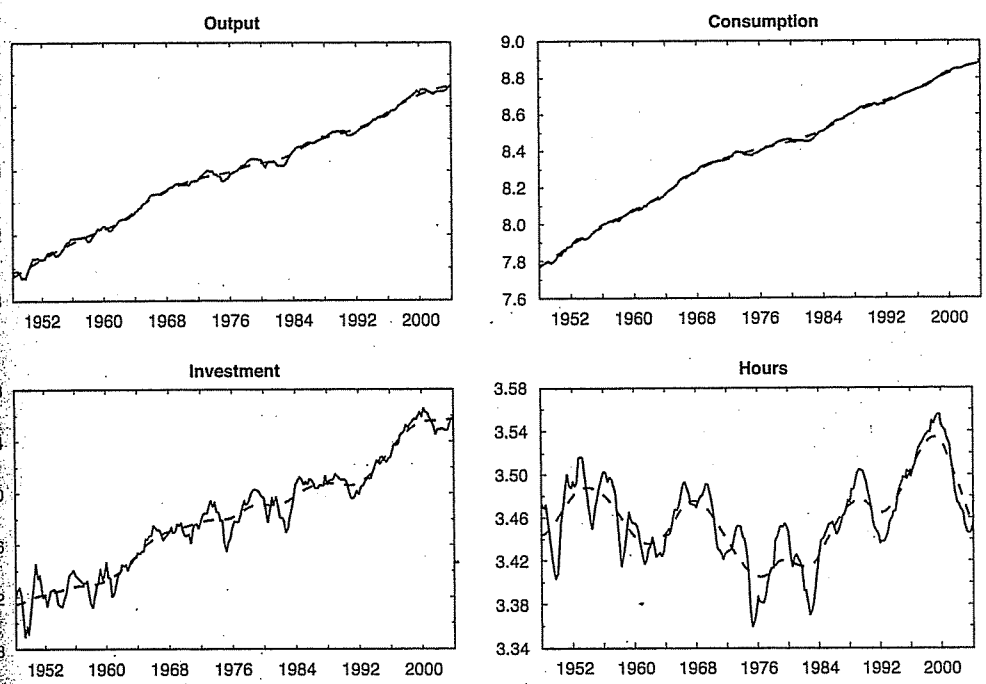

Figure 3.3 Logged trajectories and H-P trends.

The parameter λ in (3.13) 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 $\lambda \to \infty$, the trend is assigned to be maximally smooth, that is, linear.

In general, λ 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 $\widehat{g}_t$ 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 3.3. The evolution of the estimated $\widehat{g}_t$'s serves to underscore the mid-1970s reduction in the growth rates of consumption, investment, and output discussed above.

The versions of detrended output $\widetilde{y}_t$ generated by these three trend-removal procedures are illustrated in figure 3.4. Most striking is the difference in volatility observed across the three measures: the standard deviation of the linearly detrended series is 0.046, compared with 0.010 for the

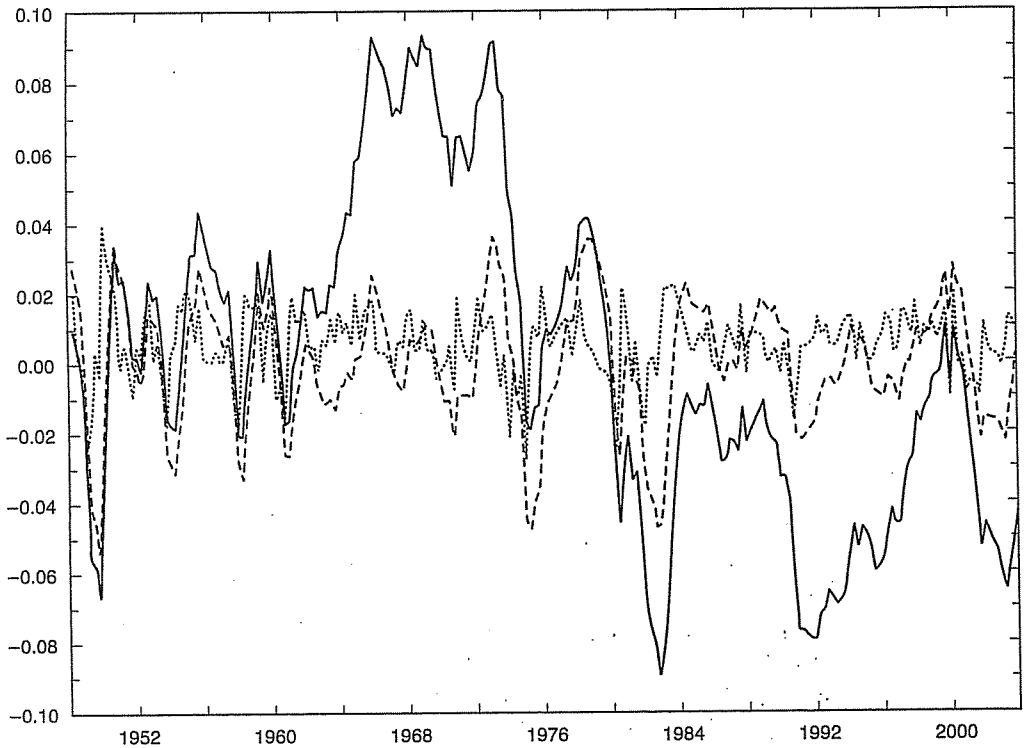

Figure 3.4 Detrended output. (Lin. Det.: solid; Diff'ed: Dots; H-P Filtered: Dashes)

differenced series and 0.018 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 mid-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 0.12; the correlation between the linearly detrended and H-P filtered series is 0.49; and the correlation between the H-P filtered and differenced series is 0.27.

3.2 Isolating Cycles

3.2.1 *Mathematical Background*

In venturing into the frequency domain, background information on complex variables is useful. Brown and Churchill (2003) and Palka (1991) 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

$$z = x + iy, \tag{3.15}$$

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

$$\begin{aligned}\sqrt{x^2 + y^2} &= \sqrt{(x + iy)(x - iy)} \\ &\equiv |z|. \end{aligned} \tag{3.16}$$

In (3.16), $(x - iy)$ 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 3.5 for an illustration.

An additional representation of z is in polar coordinates. Let ω denote the radian angle of z in (x, y) space: that is, ω is the distance traveled counterclockwise along a circle starting on the x axis before reaching z (again, see figure 3.5 for an illustration). In terms of polar coordinates, z may be represented as

$$\begin{aligned} z &= |z|(\cos\omega + i\sin\omega) \\ &= |z|e^{i\omega}, \end{aligned} \tag{3.17}$$

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 (3.17), we obtain DeMoivre's Theorem:

$$\begin{aligned} z^j &= |z|^j e^{i\omega j} \\ &= |z|^j(\cos\omega j + i\sin\omega j). \end{aligned} \tag{3.18}$$

In addition to DeMoivre's Theorem, another important result from complex analysis we shall reference below comes from the Riesz-Fischer

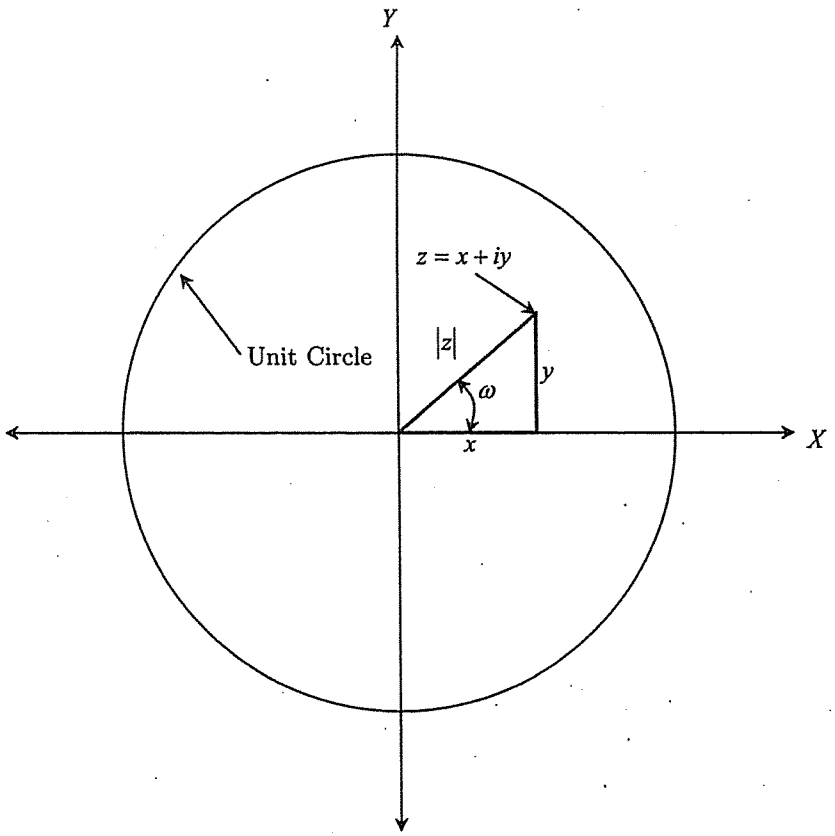

Figure 3.5 The unit circle.

Theorem, which we now sketch. For a sequence of complex numbers $\{a_j\}_{j=-\infty}^{\infty}$ that satisfy

$$\sum_{j=-\infty}^{\infty} |a_j|^2 < \infty, \tag{3.19}$$

known as a square-summability condition, there exists a complex function $f(\omega)$ such that

$$f(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j}, \tag{3.20}$$

where $\omega \in [-\pi, \pi]$. Note from (3.17) 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

$$a_j = \frac{1}{2\pi}\int_{-\pi}^{\pi} f(\omega)e^{i\omega j}\,d\omega. \tag{3.21}$$

Finally, note that for any two functions $f(\omega)$ and $g(\omega)$, where

$$f(\omega) = \sum_{j=-\infty}^{\infty} a_j e^{-i\omega j}, \qquad g(\omega) = \sum_{j=-\infty}^{\infty} b_j e^{-i\omega j},$$

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 α times the Fourier transform of $\{a_j\}_{j=-\infty}^{\infty}$.

3.2.2 *Cramér Representations*

Consider the behavior of a time series y_t^{ω} given by

$$y_t^{\omega} = \alpha(\omega)\cos(\omega t) + \beta(\omega)\sin(\omega t), \tag{3.22}$$

where $\alpha(\omega)$ and $\beta(\omega)$ are uncorrelated zero-mean random variables with equal variances. As above, ω 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 2π, 2π to 4π, and so on (the frequency for $\cos(t)$ being 1). The upper panels of figure 3.6 depict $\cos(\omega t)$ and $\sin(\omega t)$ as t evolves from 0 to 2π for $\omega = 1$ and $\omega = 2$. Accordingly, given realizations for $\alpha(\omega)$ and $\beta(\omega)$, y_t^{ω} follows a deterministic cycle that is completed ω times as t ranges from 0 to 2π, and so on. This is depicted in the lower panels of figure 3.6, 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^{ω}'s, differentiated by infinitessimal variations in ω over the interval $[0, \pi]$:

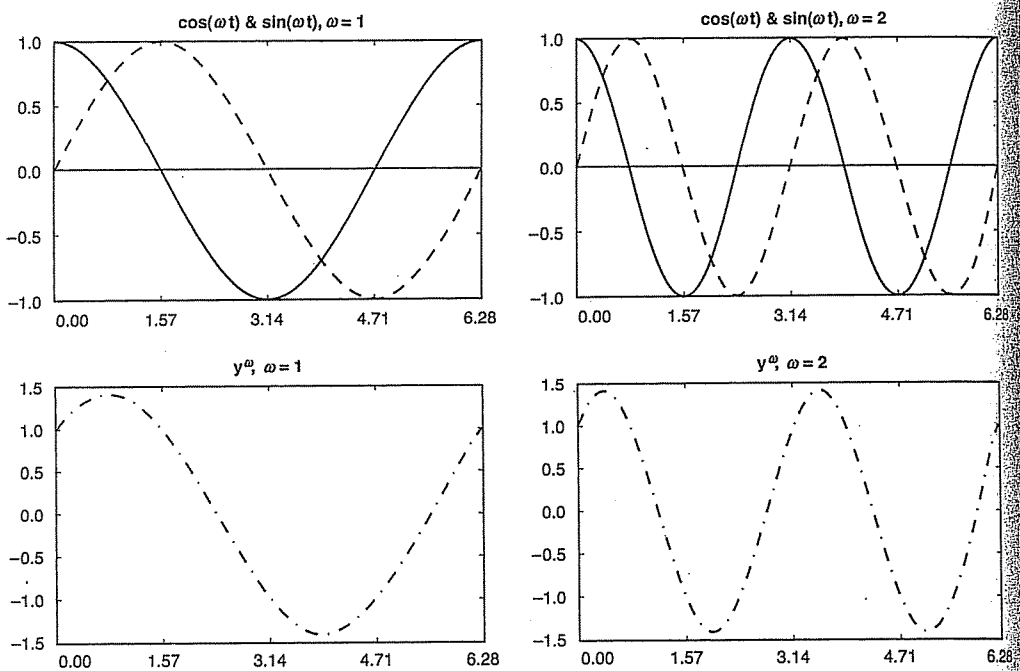

Figure 3.6 Evolution of $\cos(\omega t)$ (solid), $\sin(\omega t)$ (dashes), and y_t^{ω} (bottom panels).

$$y_t = \int_0^{\pi} \alpha(\omega)\cos(\omega t)d\omega + \int_0^{\pi} \beta(\omega)\sin(\omega t)d\omega. \tag{3.23}$$

(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.

3.2.3 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^{ω} 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 $y_{t-\tau}$):

$$\gamma(\tau) = E(y_t - \mu_t)(y_{t+\tau} - \mu_{t+\tau}), \qquad E(y_t) = \mu_t.$$

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:

$$f_y(\omega) = \sum_{\tau=-\infty}^{\infty} \gamma(\tau)e^{-i\omega\tau}. \tag{3.24}$$

Moreover, by the inversion formula,

$$\gamma(\tau) = \frac{1}{2\pi}\int_{-\pi}^{\pi} f_y(\omega)e^{i\omega\tau}d\omega. \tag{3.25}$$

The power spectrum of y_t (spectrum hereafter) is defined as

$$s_y(\omega) = \frac{1}{2\pi} f_y(\omega). \tag{3.26}$$

From (3.25) and (3.26), 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 ω 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 (3.24) and (3.26), we obtain

$$s_y(\omega) = \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} \gamma(\tau)(\cos\omega\tau - i\sin\omega\tau). \tag{3.27}$$

Moreover, since $\gamma(\tau) = \gamma(-\tau)$, $\cos(0) = 1$, $\sin(0) = 0$, $\sin(-\omega) = -\sin\omega$, and $\cos(-\omega) = \cos\omega$, (3.27) simplifies to

$$s_y(\omega) = \left(\frac{1}{2\pi}\right)\left[\gamma(0) + 2\sum_{\tau=1}^{\infty}\gamma(\tau)\cos(\omega\tau)\right]. \tag{3.28}$$

Because $\cos(\omega\tau)$ is symmetric over $[-\pi,0]$ and $[0,\pi]$, so too is $s(\omega)$; 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 ω to its associated period p, defined as

the number of units of time necessary for y_t^ω in (3.22) to complete a cycle: $p = 2\pi/\omega$. In turn, $1/p = \omega/2\pi$ indicates the number of cycles completed by y_t^ω per period. For example, with a period representing a quarter, a 10-year or 40-quarter cycle has an associated value of ω of $2\pi/40 = 0.157$. For a 6-quarter cycle, $\omega = 2\pi/6 = 1.047$. Thus values for ω in the range $[0.157, 1.047]$ are of central interest in analyzing business cycle behavior.

3.2.4 *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; etc.

The general form of a linear filter applied to y_t, producing y_t^f, is given by

$$y_t^f = \sum_{j=-r}^{s} c_j y_{t-j} \equiv C(L) y_t. \tag{3.29}$$

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: $C(e^{-i\omega})$.

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

$$\begin{aligned} E\left(y_t^f y_{t-\tau}^f\right) &= E\left(\sum_{j=-r}^{s} c_j y_{t-j}\right)\left(\sum_{k=-r}^{s} c_k y_{t-k-\tau}\right) \\ &= E\sum_{j=-r}^{s}\sum_{k=-r}^{s} c_j c_k y_{t-j} y_{t-k-\tau} \\ &= \sum_{j=-r}^{s}\sum_{k=-r}^{s} c_j c_k \gamma(\tau + k - j) \\ &\equiv \gamma_{yf}(\tau). \end{aligned} \tag{3.30}$$

Taking the Fourier transform of $\gamma_{yf}(\tau)$, the spectrum of y_t^f is given by

$$\begin{aligned} s_{yf}(\omega) &= \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty} \gamma_{yf}(\tau) e^{-i\omega\tau} \\ &= \frac{1}{2\pi}\sum_{\tau=-\infty}^{\infty}\sum_{j=-r}^{s}\sum_{k=-r}^{s} c_j c_k \gamma(\tau + k - j) e^{-i\omega\tau}. \end{aligned} \tag{3.31}$$

Let $h = \tau + k - j$, and re-write $e^{-i\omega\tau}$ in (3.31) as

$$\begin{aligned} e^{-i\omega\tau} &= e^{-i\omega(h+j-k)} \\ &= e^{-i\omega h} e^{-i\omega j} e^{i\omega k}. \end{aligned} \tag{3.32}$$

Finally, substituting for $e^{-i\omega\tau}$ in (3.31) using (3.32), we obtain

$$\begin{aligned} s_{yf}(\omega) &= \frac{1}{2\pi}\sum_{j=-r}^{s} c_j e^{-i\omega j}\sum_{k=-r}^{s} c_k e^{i\omega k}\sum_{h=-\infty}^{\infty}\gamma(h) e^{-i\omega h} \\ &= \sum_{j=-r}^{s} c_j e^{-i\omega j}\sum_{k=-r}^{s} c_k e^{i\omega k} s_y(\omega) \\ &= C(e^{-i\omega}) C(e^{i\omega}) s_y(\omega), \end{aligned} \tag{3.33}$$

where the second equality stems from the definition of $s_y(\omega)$.

Before interpreting this expression, we introduce the gain function:

$$G(\omega) = |C(e^{-i\omega})|, \tag{3.34}$$

where $|C(e^{-i\omega})|$ denotes the modulus of $C(e^{-i\omega})$:

$$|C(e^{-i\omega})| = \sqrt{C(e^{-i\omega}) C(e^{i\omega})}. \tag{3.35}$$

For example, for the first-difference filter $(1 - L)$, the gain function is given by

$$\begin{aligned} G(\omega) &= \sqrt{(1 - e^{-i\omega})(1 - e^{i\omega})} \\ &= \sqrt{2}\sqrt{1 - \cos(\omega)}, \end{aligned} \tag{3.36}$$

where the second equality follows from the identity

$$e^{-i\omega} + e^{i\omega} = 2\cos(\omega). \tag{3.37}$$

Given the definition of the gain function, the relationship between $s_{yf}(\omega)$ and $s_y(\omega)$ in (3.33) can be written as

$$\begin{aligned} s_{yf}(\omega) &= |C(e^{-i\omega})|^2 s_y(\omega) \\ &\equiv G(\omega)^2 s_y(\omega), \end{aligned} \tag{3.38}$$

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 (3.36) that the first-difference filter $(1 - L)$ shuts down cycles of frequency zero.

3.2.5 *The Hodrick-Prescott Filter*

Regarding the H-P filter, the specification of λ determines the division of the influence of y_t^ω on y_t between g_t and c_t in (3.12). Following Kaiser and Maravall (2001), its gain function is given by

$$G(\omega) = \left[1 + \left(\frac{\sin(\omega/2)}{\sin(\omega_0/2)}\right)^4\right]^{-1}, \tag{3.39}$$

where

$$\omega_0 = 2 \arcsin\left(\frac{1}{2\lambda^{1/4}}\right). \tag{3.40}$$

The parameter ω_0, selected through the specification of λ, 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 λ highlighted above) are illustrated in figure 3.7. 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 1/4 and 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

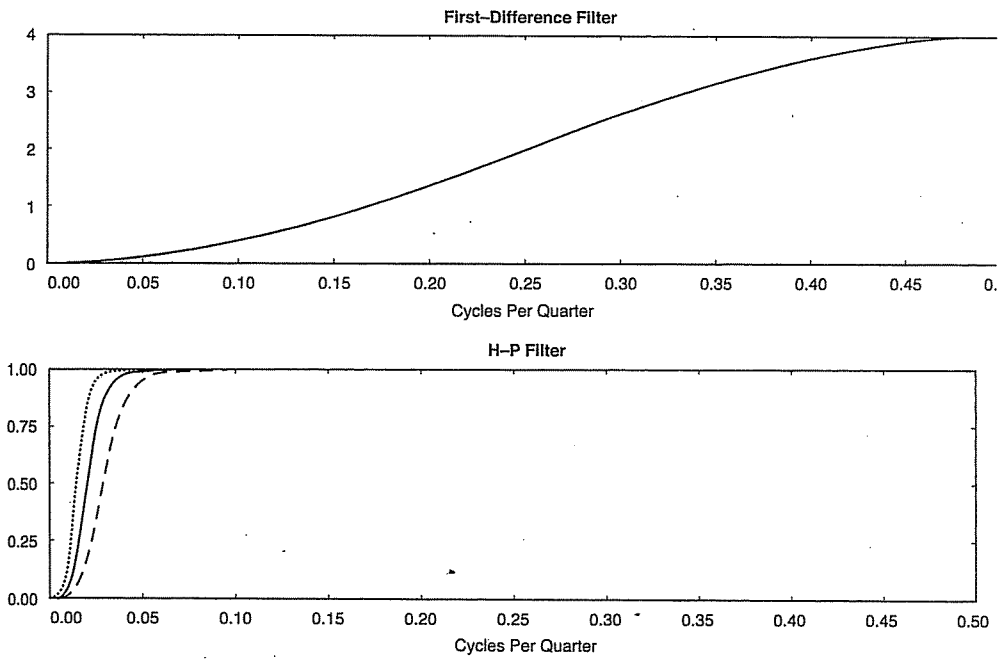

Figure 3.7 Squared gains of the first-difference and H-P filters. (Bottom Panel: $\lambda = 6,400$: Dots; $\lambda = 1,600$: Solid; $\lambda = 400$: Dashes)

designed to explain seasonal variation, thus it is desirable to work with variables that have had seasonal variations eliminated.

3.2.6 *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 3.8 presents the consumption series discussed above, along with its nonseasonally adjusted (NSA) counterpart (including H-P trends for both). Trend behavior dominates both series, but the recurrent seasonal spikes associated

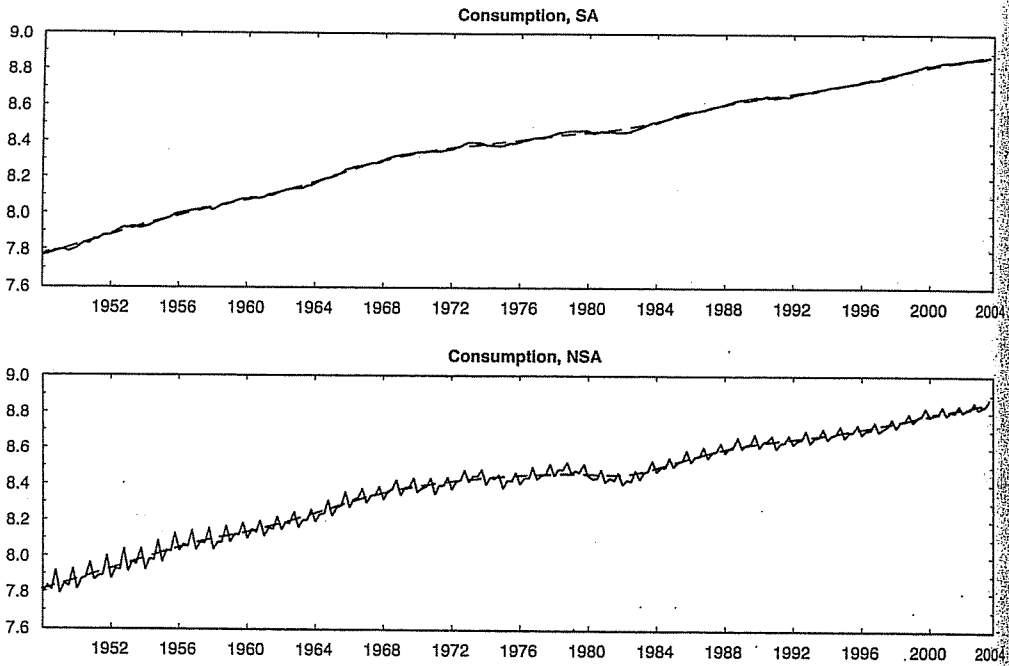

Figure 3.8 Seasonally adjusted and non–seasonally adjusted consumption.

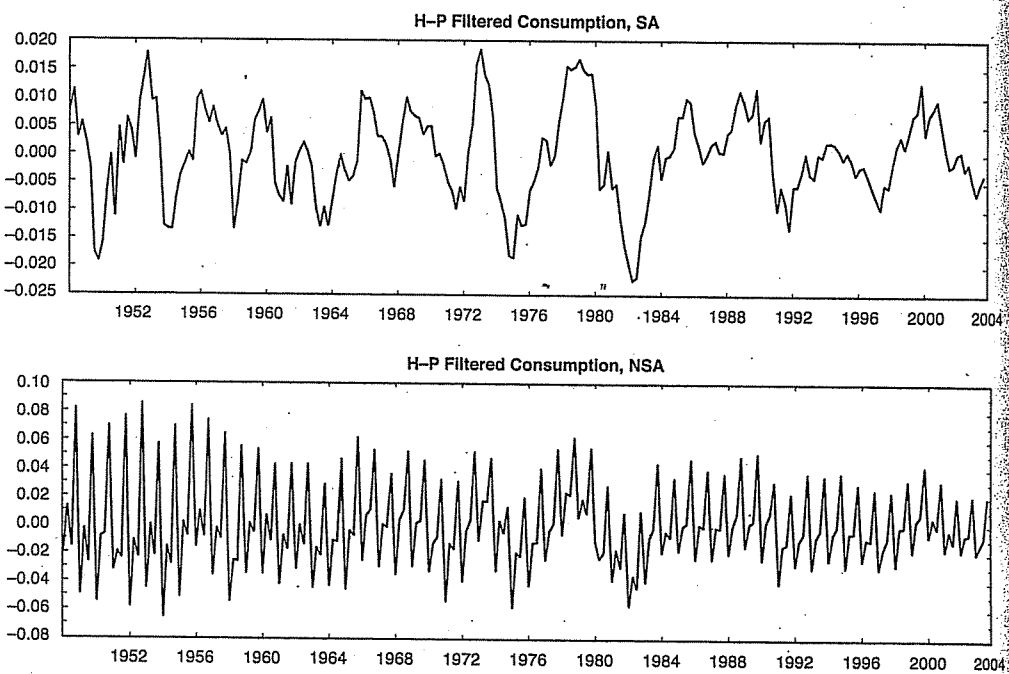

Figure 3.9 H-P filtered consumption.

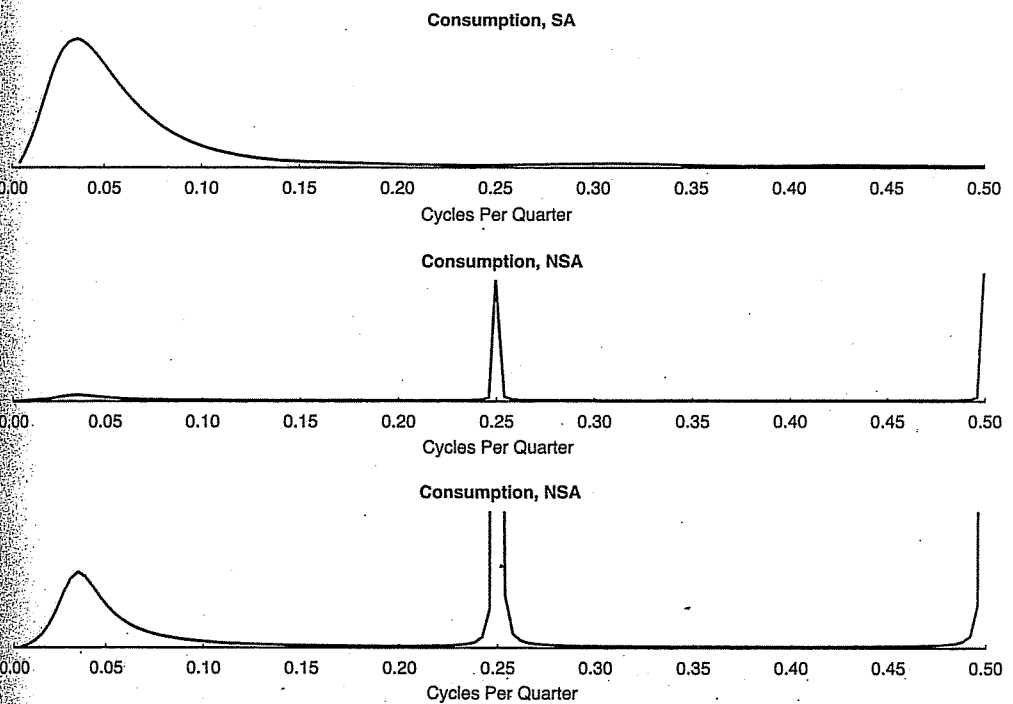

Figure 3.10 Spectra of H-P filtered consumption. (Note: The third panel zooms in on the middle panel.)

with the NSA series are distinctly apparent. The spikes are even more apparent in figure 3.9, which presents the H-P filtered series.

Figure 3.10 presents spectra estimated for both versions of the H-P filtered data (chapter 4, section 4.2 presents methods for estimating spectra). 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).

3.2.7 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

$$G(\omega)^2 = \left\{ \begin{array}{ll} 1, & \omega \in [2\pi/p_u, 2\pi/p_l] \\ 0, & \textit{otherwise.} \end{array} \right\} \tag{3.41}$$

As (3.42) 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 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).[4]

Let the ideal symmetric B-P filter for a chosen frequency range be given by

$$\alpha(L) = \sum_{j=-\infty}^{\infty} \alpha_j L^j, \tag{3.42}$$

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,

$$\begin{aligned} \alpha(e^{-i\omega}) \equiv \alpha(\omega) &= \sum_{j=-\infty}^{\infty} \alpha_j e^{-i\omega j} \\ &= \alpha_0 + \sum_{j=1}^{\infty} \alpha_j (e^{-i\omega j} + e^{i\omega j}) \\ &= \alpha_0 + 2\sum_{j=1}^{\infty} \alpha_j \cos(\omega), \end{aligned} \tag{3.43}$$

where the second equality follows from symmetry and the last equality results from (3.37).

Baxter and King's approximation to $\alpha(\omega)$ is given by the symmetric, finite-ordered filter

$$A(\omega) = a_0 + 2\sum_{j=1}^{K} a_j \cos(\omega), \tag{3.44}$$

[4] The collection of GAUSS procedures contained in `bp.src` are available for constructing Baxter and King's B-P filter.

where

$$A(0) = \sum_{j=-K}^{K} a_j = 0,$$

insuring 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

$$\min_{a_j} \int_{-\pi}^{\pi} |\alpha(\omega) - A(\omega)|^2 \, d\omega \qquad \text{subject to } A(0) = 0; \tag{3.45}$$

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

$$\begin{aligned} a_j &= \alpha_j + \theta, \qquad j = -K, \ldots, K; \\ \alpha_j &= \left\{ \begin{array}{ll} \frac{\omega_u - \omega_l}{\pi}, & j = 0 \\ \frac{\sin(\omega_2 j) - \sin(\omega_1 j)}{\pi j}, & j = \pm 1, \ldots, K; \end{array} \right\} \\ \theta &= \frac{-\sum_{j=-K}^{K} \alpha_j}{2K+1}, \end{aligned} \tag{3.46}$$

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 and end of the sample period. Figure 3.11 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 3.12. The series exhibit close correspondence, with no discernible trends or seasonal variations. The spectra estimated for these series, illustrated in figure 3.13, confirm the absence of both influences on the variations in the series.

3.3 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

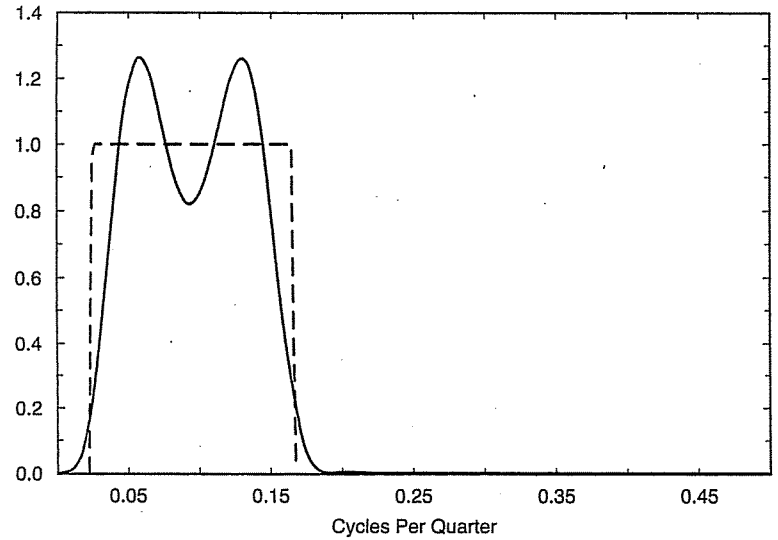

Figure 3.11 Squared gains of band pass filter and optimal filter. (fB-P filter: solid; Optimal filter: Dashes)

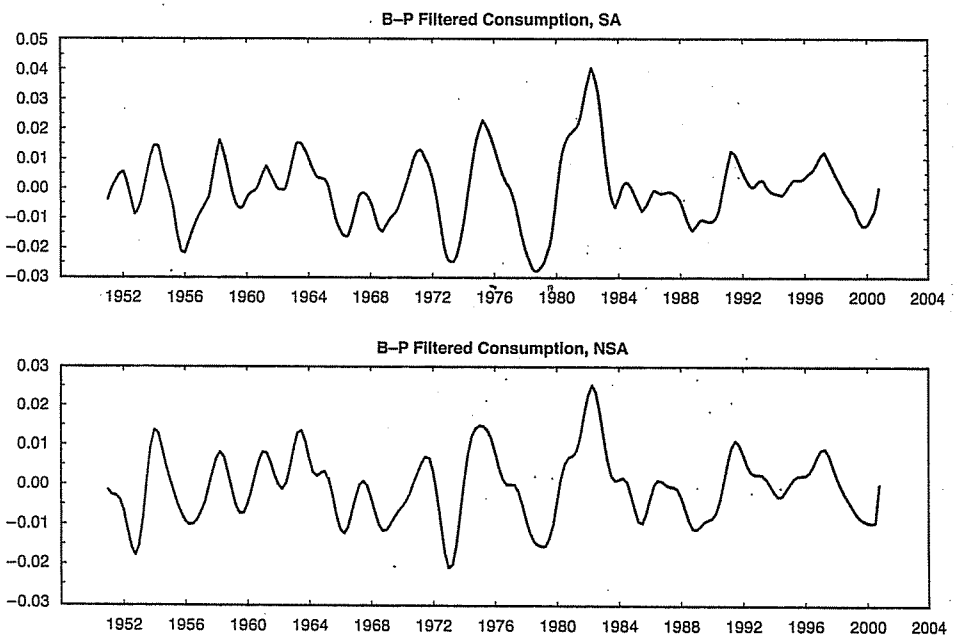

Figure 3.12 B-P filtered consumption.

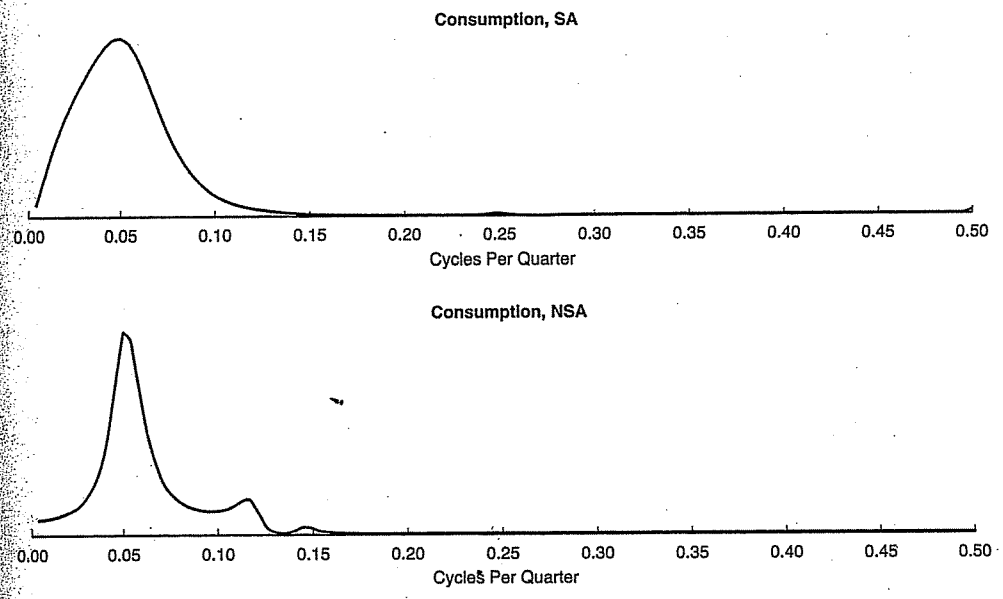

Figure 3.13 Spectra of B-P filtered consumption.

can also pose challenges, it is difficult to insure that appropriate steps have 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

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 inferences can be informative in many instances. (See DeJong and Whiteman, 1991a,b, for examples in macroeconomic applications; and Phillips, 1991, for skepticism. Tools for implementing Bayesian methods in general are presented in chapter 9.) 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 4

Summarizing Time Series Behavior

> The sign of a truly educated man is to be deeply moved by statistics.
> —George Bernard Shaw

THIS CHAPTER OPENS WITH the assumption that both the model- and data-preparation stages characterized in chapters 2 and 3 have been completed successfully. Completion of the model-preparation stage implies that the structural model under investigation is written as

$$x_t = F(\mu)x_{t-1} + e_t,$$

$$e_t = G(\mu)\upsilon_t,$$

$$E(e_t e_t') = G(\mu)E(\upsilon_t \upsilon_t')G(\mu)' = Q(\mu).$$

These equations are collectively referred to as the state system, describing the evolution of the $n \times 1$ vector x_t of model variables. Certain variables contained in x_t are unobservable, whereas others (or linear combinations thereof) are observable. Observable variables are contained in the $m \times 1$ vector X_t, and are related to x_t via either

$$X_t = H(\mu)'x_t$$

or

$$X_t = H(\mu)'x_t + u_t, \qquad E(u_t u_t') = \Sigma_u,$$

either of which is known as a measurement equation. Hereafter, the dependence of $[F, G, Q, H]$ upon the structural parameters contained in μ will often be taken as granted for ease of notation.

Completion of the data-preparation stage implies that the variables contained in X_t are mean-zero covariance stationary stochastic processes (CSSPs) exhibiting fluctuations isolated to desired frequencies. The data represent deviations from steady state values.

This chapter has three purposes. First, it presents two important reduced-form models that provide flexible characterizations of the time-series behavior of X_t. The autoregressive-moving average (ARMA) model

is presented for representing a single element of X_t, and the vector autoregressive (VAR) model is presented for representing the elements of X_t collectively. Second, the chapter presents 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. Empirical analyses involving collections of summary statistics are broadly categorized as limited-information analyses. The statistics are all calculable as functions of the parameters of either an ARMA or VAR model. The chapter concludes with a presentation of the means by which full-information analyses of structural models are conducted: evaluation of the likelihood functions corresponding with their associated state-space representations.

The chapter uses the following notation. A univariate time series considered in isolation is denoted as y_t. The variable X_{it} denotes the i^{th} element of X_t. The parameters of a given reduced-form model are collected in the vector ϑ, 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 3, 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 3 is used to illustrate the material presented in this chapter.

4.1 Two Useful Reduced-Form Models

4.1.1 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 ε_t is denoted as σ^2. Recall from chapter 3 that such a process is referred to as white noise, due to the shape of its spectrum. The variable y_t constructed as

$$\begin{aligned} y_t &= \varepsilon_t + \theta\varepsilon_{t-1} \\ &\equiv (1 + \theta L)\varepsilon_t \end{aligned} \tag{4.1}$$

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 ε_t. The variance of y_t, denoted as σ_y^2 or $\gamma(0)$, is given by

$$\begin{aligned} \sigma_y^2 &= E(y_t^2) \\ &= E(\varepsilon_t^2 + 2\theta\varepsilon_t\varepsilon_{t-1} + \theta^2\varepsilon_{t-1}^2) \\ &= (1 + \theta^2)\sigma^2, \end{aligned} \tag{4.2}$$

and its autocovariance pattern is simply

$$\begin{aligned} \gamma(1) &= E(y_t y_{t-1}) \\ &= E(\varepsilon_t + \theta\varepsilon_{t-1})(\varepsilon_{t-1} + \theta\varepsilon_{t-2}) \\ &= \theta\sigma^2; \\ \gamma(s) &= E(y_t y_{t-s}) = 0, \qquad s > 1. \end{aligned} \tag{4.3}$$

Denoting the s^{th}-order autocorrelation of y_t as $\varphi(s) = \gamma(s)/\gamma(0)$, the corresponding autocorrelation pattern is

$$\begin{aligned} \varphi(1) &= \frac{\theta}{(1 + \theta^2)}; \\ \varphi(s) &= 0, \qquad s > 1. \end{aligned} \tag{4.4}$$

Thus the impact of ε_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

$$\begin{aligned} y_t &= \varepsilon_t + \theta_1\varepsilon_{t-1} + \cdots + \theta_q\varepsilon_{t-q} \\ &= \sum_{j=0}^{q} \theta_j\varepsilon_{t-j} \\ &= \sum_{j=0}^{q} (\theta_j L^j)\varepsilon_t \\ &\equiv \theta(L)\varepsilon_t, \end{aligned} \tag{4.5}$$

where $\theta_0 = 1$. The variance of y_t in this case is given by

$$\sigma_y^2 = \sigma^2 \sum_{j=0}^{q} \theta_j^2, \tag{4.6}$$

and its covariance pattern is

$$\gamma(s) = \begin{cases} \sigma^2\left[\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \cdots + \theta_q\theta_{q-s}\right], & s = 1, \ldots, q \\ 0 & s > q. \end{cases} \tag{4.7}$$

Note that the persistence of ε_t is limited to the horizon corresponding with q.

Finally, an infinite-order MA process specified for y_t is given by

$$y_t = \sum_{j=0}^{\infty} \psi_j \varepsilon_{t-j}$$
$$\equiv \psi(L)\varepsilon_t. \tag{4.8}$$

The Wold Decomposition Theorem holds that any CSSP may be represented as in (4.8), where

$$\sum_{j=0}^{\infty} \psi_j^2 < \infty, \tag{4.9}$$

with the white noise process $\{\varepsilon_t\}$ representing a sequence of one-step-ahead forecast errors:

$$\varepsilon_t = y_t - E_{t-1}(y_t | y_{t-1}, y_{t-2}, \ldots). \tag{4.10}$$

The expectations operator $E_{t-1}(\cdot|\Omega)$ is conditional on information contained in Ω available at time $t-1$. The condition (4.9) is referred to as square summability. It is necessary to insure a finite variance and covariances for y_t:

$$\sigma_y^2 = \sigma^2 \sum_{j=0}^{\infty} \psi_j^2; \tag{4.11}$$

$$\gamma(s) = \sigma^2 [\psi_s \psi_0 + \psi_{s+1}\psi_1 + \psi_{s+2}\psi_2 + \cdots]. \tag{4.12}$$

Consider now a specification for y_t of the form

$$y_t = \rho y_{t-1} + \varepsilon_t, \qquad |\rho| < 1. \tag{4.13}$$

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 (4.13), 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 (4.13), yielding

$$y_t = \rho(\rho y_{t-2} + \varepsilon_{t-1}) + \varepsilon_t$$
$$= \varepsilon_t + \rho\varepsilon_{t-1} + \rho^2 y_{t-2}.$$

Repeated substitution for y_{t-2}, y_{t-3}, and so on yields

$$y_t = \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}. \tag{4.14}$$

Notice the role played by the restriction $|\rho| < 1$: this insures satisfaction of the square-summability condition, thus (4.14) 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 (4.14) 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 σ_y^2 and $\gamma(s)$ in (4.11) and (4.12) specialize to

$$\sigma_y^2 = \frac{\sigma^2}{1-\rho^2}; \tag{4.15}$$

$$\gamma(s) = \rho^s \frac{\sigma^2}{1-\rho^2}, \qquad s = 1, 2, \ldots. \tag{4.16}$$

Correspondingly, $\varphi(s) = \rho^s$. From both (4.14) and (4.16), it is evident that ρ determines the persistence of ε_t: the nearer is ρ to 1, the stronger is the influence of ε_{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 (4.14). Rewriting (4.13) as

$$y_t = \rho L y_t + \varepsilon_t,$$

(4.14) can be derived as follows:

$$(1-\rho L)y_t = \varepsilon_t;$$
$$y_t = \frac{1}{(1-\rho L)}\varepsilon_t$$
$$= \sum_{j=0}^{\infty} (\rho^j L^j)\varepsilon_t$$
$$= \sum_{j=0}^{\infty} \rho^j \varepsilon_{t-j}.$$

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. Again returning 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

$$y_t = \rho_1 y_{t-1} + \rho_2 y_{t-2} + \cdots + \rho_p y_{t-p} + \varepsilon_t, \tag{4.17}$$

or

$$(1 - \rho_1 L - \cdots - \rho_p L^p) y_t \equiv \rho(L) y_t = \varepsilon_t. \tag{4.18}$$

Factoring $(1 - \rho_1 L - \cdots - \rho_p L^p)$ as $(1 - \lambda_1 L)(1 - \lambda_2 L) \ldots (1 - \lambda_p L)$, y_t can be expressed as

$$\begin{aligned} y_t &= \left(\frac{1}{\rho(L)}\right)\varepsilon_t \\ &= \left(\frac{1}{1-\lambda_1 L}\right)\left(\frac{1}{1-\lambda_2 L}\right)\cdots\left(\frac{1}{1-\lambda_p L}\right)\varepsilon_t \\ &= \psi(L)\varepsilon_t. \end{aligned} \tag{4.19}$$

So long as the roots $1/\lambda_j, j = 1, \ldots, p$ of the polynomial $\rho(L)$ lie outside the unit circle, $\psi(L)$ is square-summable and (4.19) constitutes a Wold Representation.

The persistence of ε_t is determined by the proximity of the smallest root to unity: the closer to unity, the greater is the associated value of λ, 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 (4.19), 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 σ_y^2 and $\gamma(s)$ using (4.11) and (4.12)) as functions of the coefficients of $\rho(L)$ is provided by the method of undetermined coefficients. This begins by combining (4.8) and (4.18) to obtain

$$\rho(L)^{-1}\varepsilon_t = \psi(L)\varepsilon_t,$$

implying

$$\begin{aligned} 1 &= \rho(L)\psi(L) \\ &= (1 - \rho_1 L - \cdots - \rho_p L^p)(\psi_0 + \psi_1 L + \psi_2 L^2 + \cdots). \end{aligned} \tag{4.20}$$

Both sides of (4.20) may be thought of as infinite-order polynomials in L. Equating the coefficients associated with $L^0, L^1, \ldots$ on both sides of (4.20) yields the following system of equations:

$$\begin{aligned} 1 &= \psi_0 \\ 0 &= \psi_1 - \rho_1 \Rightarrow \psi_1 = \rho_1 \\ 0 &= \psi_2 - \rho_1\psi_1 - \rho_2 \Rightarrow \psi_2 = \rho_1^2 + \rho_2 \\ &\cdots \\ 0 &= \psi_p - \rho_1\psi_{p-1} - \rho_2\psi_{p-2} - \cdots - \rho_p\psi_0 \\ 0 &= \psi_{p+j} - \rho_1\psi_{p+j-1} - \rho_2\psi_{p+j-2} - \cdots - \rho_p\psi_j, \qquad j = 1, 2, \ldots. \end{aligned} \tag{4.21}$$

Exercise 4.1

Derive expressions for $\psi_j, j = 0, 1, 2, \ldots$ as functions of the coefficients of $\rho(L)$ for $p = 1, 2$, and 3.

The specifications for y_t as outlined are encompassed by an ARMA (p, q) model, expressed as

$$(1 - \rho_1 L - \cdots - \rho_p L^p) y_t = (1 + \theta_1 L + \cdots + \theta_q L^q)\varepsilon_t, \quad \text{or}$$
$$\rho(L) y_t = \theta(L)\varepsilon_t. \tag{4.22}$$

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

$$y_t = \left(\frac{\theta(L)}{\rho(L)}\right)\varepsilon_t. \tag{4.23}$$

Combining (4.8) and (4.23) yields the relationship

$$\theta(L) = \rho(L)\psi(L), \tag{4.24}$$

which generalizes (4.20). Accordingly, the method of undetermined coefficients may be applied to (4.24) to obtain expressions for $\psi_j, j = 0, 1, 2, \ldots$ as functions of the coefficients of $\theta(L)$ and $\rho(L)$, again yielding expressions for σ_y^2 and $\gamma(s)$ via (4.11) and (4.12).

Exercise 4.2

Derive expressions for $\psi_j, j = 0, 1, 2, \ldots$ 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

$\{y_{-(p+1)}, y_{-(p)}, \ldots y_{-(1)}\} = 0$, $\varepsilon_0 = 1$, and $\{\varepsilon_t\}_{t=1}^{\infty} = 0$. Then from (4.8), the resulting sequence $\{y_t\}_{t=0}^{\infty}$ is identically $\{\psi_j\}_{j=0}^{\infty}$. From (4.22), the sequence $\{\psi_j\}_{j=0}^{\infty}$ may thus be constructed iteratively as follows:

$$\begin{aligned}
\psi_0 &= 1 \\
\psi_1 &= \rho_1 + \theta_1 \\
\psi_2 &= \rho_1\psi_1 + \rho_2 + \theta_2 \\
&\ldots \\
\psi_p &= \rho_1\psi_{p-1} + \rho_2\psi_{p-2} + \cdots + \rho_p + \theta_p \\
&\ldots \\
\psi_j &= \rho_1\psi_{j-1} + \rho_2\psi_{j-2} + \cdots + \rho_p\psi_{j-p} + \theta_j, \qquad (4.25)
\end{aligned}$$

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, because 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 (4.17) that the coefficients of an AR(p) model can be estimated in straightforward fashion using an ordinary least squares (OLS) regression. In the regression, the $(T-p) \times 1$ vector $y = [y_{p+1}, y_{p+2}, \ldots y_T]'$ serves as the dependent variable, and Ly, $L^2y, \ldots, L^Py$ serve as independent variables, where $L^jy = [y_{p+1-j}, y_{p+2-j}, \ldots y_{T-j}]'$, 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, as you will be asked to demonstrate in exercise 4.5 below (where conditioning is on the initial observations $[y_1, y_2, \ldots, y_p]'$).

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 ε_t. But as demonstrated in section 4.3, this complication may be overcome via use of the Kalman filter; see in particular exercise 4.6.[1]

4.1.2 *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

[1] 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

y_t. The multivariate analogue to the AR(p) specification (4.18) for X_t is given by

$$\begin{bmatrix} \rho_{11}(L) & \rho_{12}(L) & \ldots & \rho_{1m}(L) \\ \rho_{21}(L) & \rho_{22}(L) & \ldots & \rho_{2m}(L) \\ \ldots & \ldots & \ldots & \ldots \\ \rho_{m1}(L) & \rho_{m2}(L) & \ldots & \rho_{mm}(L) \end{bmatrix} \begin{bmatrix} X_{1t} \\ X_{2t} \\ \ldots \\ X_{mt} \end{bmatrix} = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \ldots \\ \varepsilon_{mt} \end{bmatrix}, \quad E(\varepsilon_t\varepsilon_t') = \Sigma, \qquad (4.26)$$

where

$$\rho_{ij}(L) = (1 + \rho_{ij1}L + \cdots + \rho_{ijp}L^p)$$

is a p^{th}-order polynomial in the lag operator, expressing the influence of X_j on X_i. Defining the lead matrix in (4.26) 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 (4.8):

$$X_t = \sum_{j=0}^{\infty} \Psi_j \varepsilon_{t-j}, \qquad (4.27)$$

where Ψ_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

$$\begin{bmatrix} X_{it} \\ X_{it-1} \\ X_{it-2} \\ \ldots \\ X_{it-(p+1)} \end{bmatrix} = \begin{bmatrix} \rho_{ii1} & \rho_{ii2} & \rho_{ii3} & \ldots & \rho_{iip} \\ 1 & 0 & 0 & \ldots & 0 \\ 0 & 1 & 0 & \ldots & 0 \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ 0 & 0 & 0 & \ldots 1 & 0 \end{bmatrix} \begin{bmatrix} X_{it-1} \\ X_{it-2} \\ X_{it-3} \\ \ldots \\ X_{it-p} \end{bmatrix} + \begin{bmatrix} \varepsilon_{it} \\ 0 \\ 0 \\ \ldots \\ 0 \end{bmatrix}. \qquad (4.28)$$

Defining the $(mp) \times 1$ vector

$$z_t = [X_{1t}\ X_{1t-1} \ldots X_{1t-(p+1)}\ X_{2t}\ X_{2t-1} \ldots X_{2t-(p+1)} \ldots X_{mt}\ X_{mt-1} \ldots X_{mt-(p+1)}]',$$

and the $(mp) \times 1$ vector

$$e_t = [\varepsilon_{1t}\ 0 \ldots 0\ \varepsilon_{2t}\ 0 \ldots 0\ \varepsilon_{mt}\ 0 \ldots 0]',$$

the companion form for the VAR is given by

$$z_t = Az_{t-1} + e_t, \tag{4.29}$$

where the $(mp) \times (mp)$ companion matrix A contains VAR equations in the 1^{st}, $(p+1)^{th}, \ldots,$ and $[(m-1)p+1]^{th}$ rows, and maintains identities between elements of z_t and z_{t-1} in the remaining rows. Note the correspondence between (4.29) and the AR(1) expression for y_t in (4.13).

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^{th}-order covariance matrix. From (4.29),

$$\begin{aligned} \Gamma(0) &= E[(Az_{t-1} + e_t)(Az_{t-1} + e_t)'] \\ &= A\Gamma(0)A' + \Sigma, \end{aligned} \tag{4.30}$$

the solution to which is given by

$$vec\,[\Gamma(0)] = [I - A \otimes A]^{-1}\, vec[\Sigma], \tag{4.31}$$

where $\otimes$ denotes the Kronecker product. Further,

$$\begin{aligned} \Gamma(1) &= E(z_t z_{t-1}') \\ &= E((Az_{t-1} + e_t)z_{t-1}') \\ &= A\Gamma(0), \end{aligned}$$

and in general,

$$\begin{aligned} \Gamma(s) &= A\Gamma(s-1) \\ &= A^s\Gamma(0). \end{aligned} \tag{4.32}$$

Note the symmetry between (4.31) and (4.15), and between (4.32) and (4.16).

As with the AR(p) specification, the parameters of the VAR model may be estimated using OLS by rewriting (4.26) in standard regression notation:

$$\Upsilon = \Xi B + u, \tag{4.33}$$

where the $(T-p) \times m$ matrix

$$\Upsilon = [X_1 X_2 \ \ldots \ X_m],$$

with

$$X_i = [X_{ip+1} X_{ip+2} \ \ldots \ X_{iT}]',$$

the $(T-p) \times (mp)$ matrix Ξ contains in its t^{th} row

$$\begin{aligned} &[X_{1t-1}\ X_{1t-2}\ \ldots\ X_{1t-p}\ X_{2t-1}\ X_{2t-2}\ \ldots\ X_{2t-p} \\ &\ldots\ X_{mt-1}\ X_{mt-2}\ \ldots\ X_{mt-p}], \end{aligned}$$

the $(mp) \times m$ matrix B contains VAR parameters for X_i in its i^{th} column, and the $(T-p) \times m$ matrix u contains the $(T-p) \times 1$ vector of innovations $[\varepsilon_{ip+1}\ \varepsilon_{ip+2} \ldots \varepsilon_{iT}]'$ 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 ε_t.[2]

4.2 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, because as indicated in (4.12), the greater the persistence (i.e., the greater the horizon over which ψ_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.

[2] The collection of procedures contained in `var.src` is available for estimating VAR specifications, constructing the companion matrix A, and calculating $\Gamma(s)$, $s = 0, 1, \ldots$.

Estimates of the elements of $\varphi(s)$ may be obtained using the following collection of sample averages:[3]

$$\bar{y} = \left(\frac{1}{T}\right)\sum_{t=1}^{T} y_t$$

$$\hat{\gamma}(0) = \left(\frac{1}{T}\right)\sum_{t=1}^{T}(y_t - \bar{y})^2$$

$$\hat{\gamma}(s) = \left(\frac{1}{T}\right)\sum_{t=s+1}^{T}(y_t - \bar{y})(y_{t-s} - \bar{y})$$

$$\hat{\varphi}(s) = \hat{\gamma}(s)/\hat{\gamma}(0). \tag{4.34}$$

Alternatively, given estimates of an ARMA specification for y_t, the corresponding Wold Representation of y_t may be constructed using (4.25), which can then be mapped into $\hat{\varphi}(s)$ using (4.12).

Plots of $\hat{\varphi}(s)$ (estimated using sample averages) for the four filtered versions of output described in chapter 3 are illustrated in figure 4.1. The four 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 4.1 is the high degree of persistence exhibited by the detrended series. Recall from figure 3.2 that the level of the series follows 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 $\hat{\varphi}(s)$ depicted in figure 4.1, which decays very slowly, and remains at approximately 0.25 even at the 40-quarter horizon. Plots of $\hat{\varphi}(s)$ 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 cyclicality of y_t is through the construction of its spectrum. The spectrum was described in detail in chapter 3, section 3.2; here we provide a brief summary. The spectrum $s_y(\omega)$ is a decomposition of the variance of y_t by frequency ω. Frequency, measured in radians, is usefully interpreted through its relationship

[3] The GAUSS command `autocor` performs these calculations.

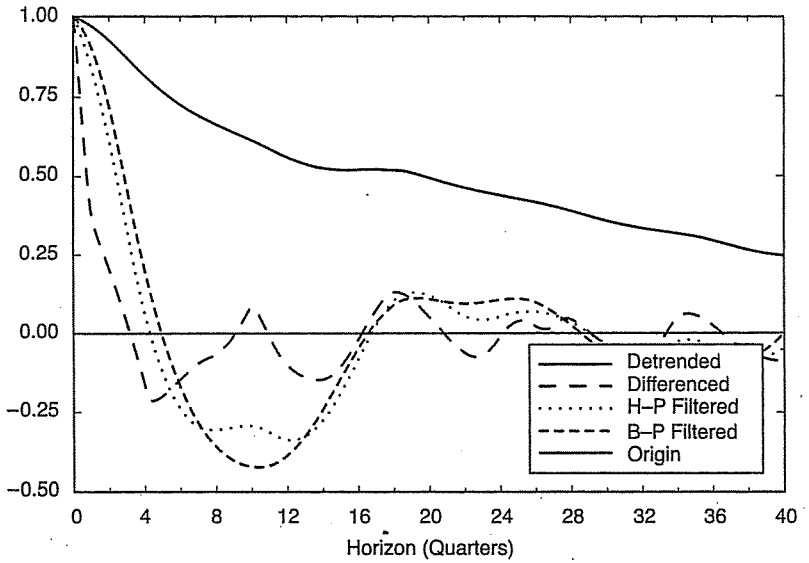

Figure 4.1 Sample autocorrelations of output.

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:

$$s_y(\omega) = \left(\frac{1}{2\pi}\right)\left[\gamma(0) + 2\sum_{\tau=1}^{\infty}\gamma(\tau)\cos(\omega\tau)\right]. \tag{4.35}$$

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 ω indicate the relative importance of fluctuations at the chosen frequencies in influencing variations in y_t. Recall that because $\cos(\omega\tau)$ 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 (4.35), the construction of an estimate of $s_y(\omega)$ for y_t is straightforward given estimates $\hat{\gamma}(s)$. Alternatively, an estimate of $s_y(\omega)$ can be obtained through its relationship with the parameters of an ARMA specification estimated for y_t; the relationship is given by:

$$s_y(\omega) = \left(\frac{\sigma^2}{2\pi}\right)\frac{|\theta(e^{-i\omega})|^2}{|\rho(e^{-i\omega})|^2}, \tag{4.36}$$

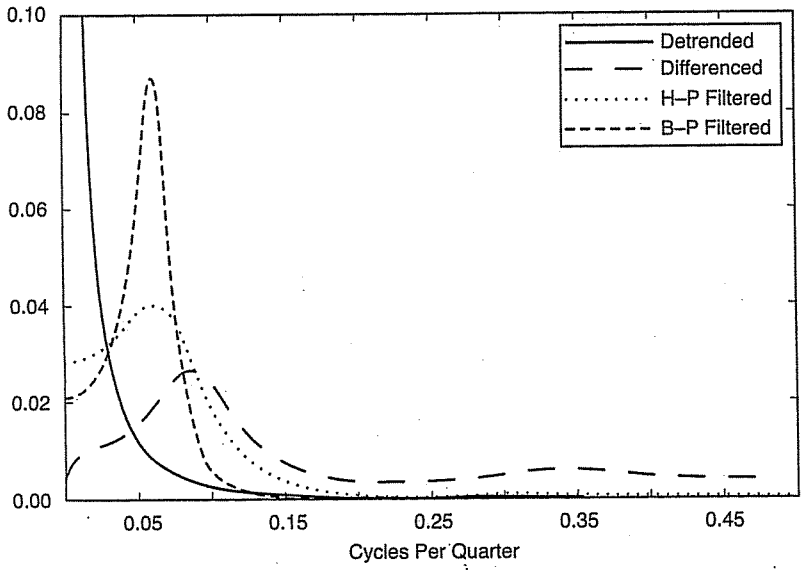

Figure 4.2 Spectra of output.

where

$$e^{-i\omega} = \cos(\omega) - i\sin(\omega),$$

$\sqrt{i} = -1$, and $|\cdot|$ denotes the modulus operator; for example,[4]

$$|\theta(e^{-i\omega})| = \sqrt{\theta(e^{-i\omega})\theta(e^{i\omega})}.$$

Spectra estimated for the four versions of output described above are illustrated in figure 4.2. Also, spectra estimated for the H-P filtered versions of output, consumption, investment, and hours are illustrated in figure 4.3. The estimates were obtained by estimating ARMA models for each series, and constructing $s_y(\omega)$ as in (4.36). 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 4.1 translate into the behavior of the spectra depicted in figure 4.2. The persistence evident in the detrended series translates as a spike in its corresponding spectrum at zero frequency. (The height of this spectrum

[4] See Harvey (1993) for a derivation of (4.36). The procedure `spec_arma.prc` can be used to construct $s_y(\omega)$, taking $\rho(L)$ and $\theta(L)$ as inputs.

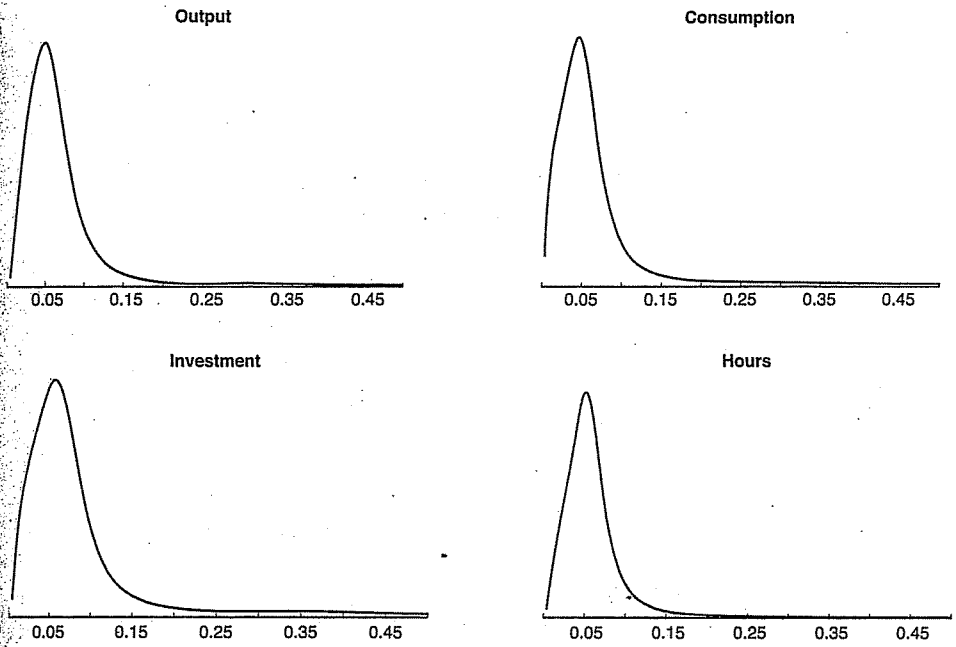

Figure 4.3 Spectra of H-P filtered data.

actually grows unboundedly as ω 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.

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 3, section 3.2 illustrate (see in particular figures 3.6 and 3.10). The spectral peak associated with the differenced series is much less pronounced in comparison, and the non-trivial 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 4.1. 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 ψ_j, $j = 1, 2, \ldots$ of the Wold Representation of y_t, as expressed in (4.8). Impulse response functions are illustrated for the four H-P filtered series in figure 4.4. Each

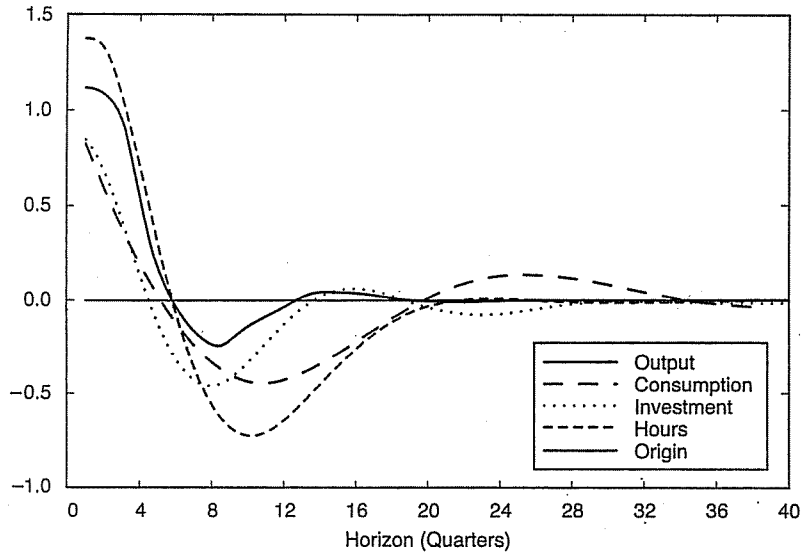

Figure 4.4 Univariate impulse response functions of H-P filtered data.

function was constructed from estimates of an ARMA specification, which were used to construct $\psi_j, j = 1, 2, \ldots$ as in (4.25).[5]

Note how the cyclical nature of the responses mimic 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 (4.29). Patterns of auto- and cross-covariation are usefully characterized using $\Gamma(s) = E(z_t z_t')$, $s = 0, 1, \ldots$. These may be obtained from estimated VAR parameters following (4.30)–(4.32). A collection of statistics frequently of interest in empirical business cycle applications are reported for the differenced, H-P filtered, and B-P filtered data in table 4.1. 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 (4.31) is not possible in this case.

[5] The procedure `armaimp.prc` can be used to calculate impulse response functions in this fashion.

TABLE 4.1
Summary statistics estimated from VAR

			Differenced Data		
j	σ_j	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\varphi_{j,y}(1)$
y	0.0099	1.00	0.36	1.00	0.36
c	0.0051	0.51	0.22	0.54	0.42
i	0.0505	5.10	0.14	0.91	0.21
h	0.0091	0.92	0.60	0.69	0.32
			H-P filtered Data		
j	σ_j	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\varphi_{j,y}(1)$
y	0.0177	1.00	0.86	1.00	0.86
c	0.0081	0.46	0.83	0.82	0.75
i	0.0748	4.23	0.79	0.95	0.80
h	0.0185	1.05	0.90	0.83	0.62
			B-P filtered Data		
j	σ_j	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\varphi_{j,y}(1)$
y	0.0184	1.00	0.94	1.00	0.94
c	0.0084	0.46	0.94	0.90	0.85
i	0.0733	3.98	0.92	0.96	0.89
h	0.0193	1.05	0.94	0.87	0.71

The first column of the table reports standard deviations of the individual series, and the second reports standard deviations relative to the standard deviation of output. Note that under all three versions of the data, investment is far more volatile than output (σ_i/σ_y ranges from approximately 4 to 5), whereas consumption is far smoother (σ_c/σ_y is roughly 0.5); the volatility of hours and output are roughly equal. Measures of first-order serial correlation are quite low among the differenced series (particularly for investment: 0.14); while ranging from 0.79 to 0.94 among the H-P and B-P filtered data. Finally, with $\varphi_{j,y}(s)$ denoting the s^{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 4.3

Construct the collection of summary statistics presented in table 4.1 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–2004:IV. For each, fit a common linear trend to logged output, consumption, and investment; subtract the estimated trend; then recombine the resulting series (use the procedure `ct.prc` for this step).
2. Estimate a VAR representation for the data using $p = 8$, construct the companion matrix A as in (4.29), and then construct $\Gamma(0)$ and $\Gamma(1)$ using (4.30)–(4.32) (use the procedures contained in `var.src` 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 (4.36). Setting $\theta(L) = 1$, (4.36) characterizes how the spectrum of a single variable is obtained using estimates of an AR(p) model. To obtain the multivariate analogue of (4.36), write the VAR of X_t as

$$X_t = \Upsilon_1 X_{t-1} + \Upsilon_2 X_{t-2} + \cdots + \Upsilon_p X_{t-p} + \varepsilon_t, \tag{4.37}$$

where Υ_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[6]

$$S(\omega) = \frac{1}{2\pi}[\Upsilon(e^{-i\omega})'\Sigma^{-1}\Upsilon(e^{-i\omega})]^{-1}. \tag{4.38}$$

The j^{th} diagonal element of $S(\omega)$ contains the spectrum of the j^{th} variable of X_t at frequency ω, $s_j(\omega)$, and the $(i,j)^{\text{th}}$ element of $S(\omega)$ contains the cross spectrum between the i^{th} and j^{th} variables of X_t, $s_{i,j}(\omega)$. Estimates of the spectra for the H-P filtered data produced using (4.38) closely mimic those depicted in figure 4.3, 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 (4.29). Initializing $z_{-1} = 0$, this proceeds by constructing a nonzero initial value e_0, and then constructing $\{z_t\}_{t=0}^{\infty}$ as $z_0 = e_0$, $z_1 = Ae_0$, $z_2 = A^2 e_0, \ldots$ The specification for e_0 is typically chosen to simulate the impact on the system of a one-standard-deviation innovation to the j^{th} variable of the system. If the innovations were all uncorrelated, so that the covariance matrix Σ 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^{th} equation of the VAR. This entry would be set to the square root of the

[6] For a derivation, see Hannan (1970). The procedure `spec_var.prc` is available for use in constructing $S(\omega)$ using (4.38).

j^{th} diagonal element of Σ. However, correlation among the innovations implies that an innovation to the j^{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

$$e_t = [\varepsilon_{1t}\ 0 \ldots 0\ \varepsilon_{2t}\ 0 \ldots 0\ \varepsilon_{mt}\ 0 \ldots 0]';$$

for the moment, it will be convenient to work directly with the vector ε_t. To recap, the problem faced in constructing e_0 is that the elements of ε_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 ε_t. Orthogonalized innovations are innovations that are uncorrelated across VAR equations. Let ν_t represent such innovations, so that $E(\nu_t \nu_t') = I$. Defining a matrix P such that

$$P^{-1}\Sigma P'^{-1} = I, \tag{4.39}$$

which implies

$$\Sigma = PP', \tag{4.40}$$

ν_t may be constructed using

$$\nu_t = P^{-1}\varepsilon_t. \tag{4.41}$$

The only problem at this point is that there are many possible specifications of P^{-1} that can be constructed to satisfy (4.40). Here we discuss a leading specification: the Cholesky decomposition of Σ; see Hamilton (1994) for a discussion of alternative decompositions.

The Cholesky decomposition of Σ is a lower-triangular matrix that satisfies (4.40), with diagonal elements containing the square root of the diagonal elements of Σ (i.e., the standard deviations of the elements of ε_t). Consider the construction of ν_0 obtained by introducing into (4.41) a specification of ε_0 containing 1 in its j^{th} row and zeros elsewhere. The resulting specification of ν_0 will contain the j^{th} column of P^{-1}. Using the Cholesky decomposition of Σ for P^{-1}, ν_0 will contain zeros in the first $j-1$ entries, the standard deviation of the j^{th} element of ε_t in the j^{th} row, and non-zero entries in the remaining rows. These latter entries represent the influence of an innovation to the j^{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^{th} equation is prevented from influencing the 1^{st} through

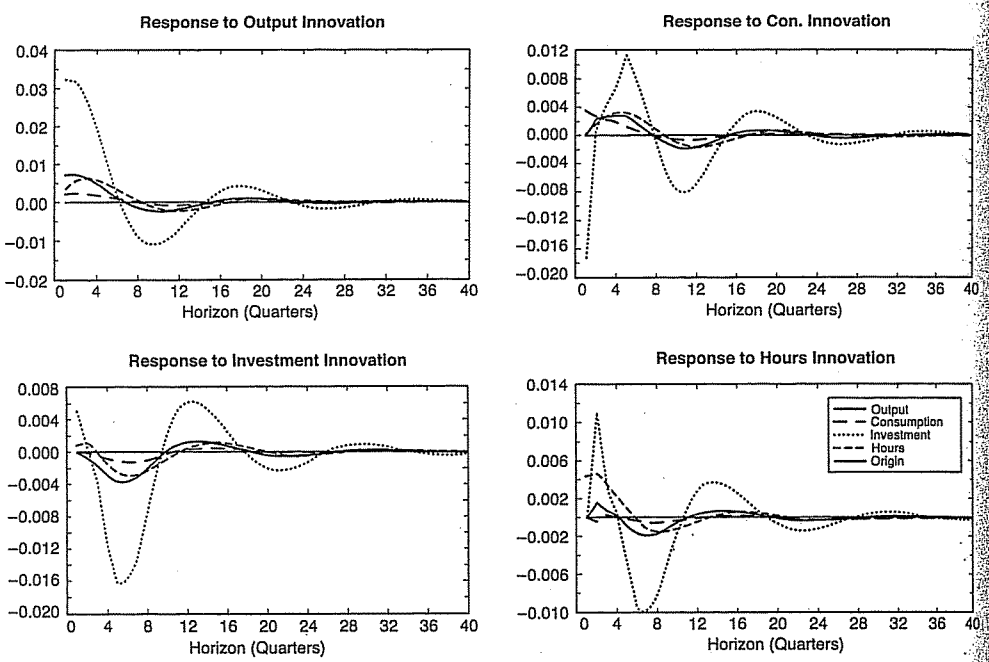

Figure 4.5 System impulse response functions of H-P filtered data.

$(j-1)^{st}$ 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 4.5.[7] 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: note in particular that the variables tend to overshoot their targets in returning to pre-shock levels.

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

[7] A procedure for calculating impulse response functions is included in `var.src`.

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 parameterized model. For example, this could involve individual t-tests of the null hypothesis that the q^{th} lag chosen for the MA component of an ARMA model, or the p^{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 $2\log[L_u - L_r]$, 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

$$AIC = |\Sigma(K)| + \frac{2K}{T}, \tag{4.42}$$

and the Bayesian Information Criterion (BIC) (Schwarz, 1978), which delivers the K that minimizes

$$BIC = |\Sigma(K)| + \frac{K\log(T)}{T}. \tag{4.43}$$

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.

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. Recall that these are contained in the vector x_t, which has a structural representation given by

$$x_t = Fx_{t-1} + e_t, \qquad E(e_t e_t') = Q. \tag{4.44}$$

Note the similarity of this specification relative to the companion form of the VAR specified for z_t. Exploiting this similarity, we simply note here that

each of the summary statistics that can be constructed from VAR estimates can also be constructed for x_t by replacing A and Σ with F and Q. Moreover, because the relationship between model variables and observables is given by the simple linear mapping $X_t = H'x_t$, it is straightforward to align summary statistics obtained from reduced-form specifications with those obtained from structural specifications. The various empirical techniques presented in part II of this book use alternative approaches to aligning statistics and judging their proximity.

We conclude this section with a note regarding measures of the precision with which summary statistics have been estimated. For concreteness, let $\widehat{g(\vartheta)}$ denote the estimate of a given function of the $k \times 1$ vector of parameters ϑ, which summarize a reduced-form model specified for the data. (The replacement of $\widehat{g(\varphi)}$ with $\widehat{g(\mu)}$, 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 $\widehat{g(\vartheta)}$. 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 $\widehat{g(\vartheta)}$.

Beginning with the former, because the data used to construct $\widehat{g(\vartheta)}$ are random, and represent one of many possible realizations that could have been obtained, $\widehat{g(\vartheta)}$ is also random. Its variance is referred to as the sampling variance of $g(\vartheta)$, which must typically be estimated. The square root of this estimate is the standard error of $\widehat{g(\vartheta)}$. In some cases analytical expressions for standard errors are available; often they are not. For example, if the vector ϑ is being estimated using an OLS regression of the form

$$y = X\vartheta + \varepsilon,$$

and $\widehat{g(\vartheta)}$ is a vector representing the OLS estimate of ϑ, then the associated standard errors of the individual elements of $\widehat{g(\vartheta)}$ are the square roots of the diagonal elements of

$$Var(\widehat{\vartheta}) = \widehat{\sigma}^2(X'X)^{-1}, \qquad \widehat{\sigma}^2 = \left(\frac{1}{T}\right)\sum_{t=1}^{T}\widehat{e}_t^2, \qquad \widehat{e}_t = y - X\widehat{\vartheta}.$$

When $g(\vartheta)$ represents a nonlinear function of ϑ, analytical expressions for the standard error of $\widehat{g(\vartheta)}$ 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 ϑ:

$$\widehat{g(\vartheta)} \approx g(\vartheta) + \left(\frac{\partial g(\vartheta)}{\partial \vartheta}\right)'(\widehat{\vartheta} - \vartheta). \tag{4.45}$$

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^{th} row of $g(\vartheta)$ with respect to the j^{th} element of ϑ. By Slutsky's Theorem (e.g., as presented in Greene, 2003), if $\widehat{\vartheta}$ is a consistent estimate of ϑ, then $g(\widehat{\vartheta})$ will be a consistent estimator of $g(\vartheta)$, with variance given by

$$Var[\widehat{g(\vartheta)}] \approx \left(\frac{\partial g(\vartheta)}{\partial \vartheta}\right)' Var(\widehat{\vartheta}) \left(\frac{\partial g(\vartheta)}{\partial \vartheta}\right). \tag{4.46}$$

The standard errors associated with the individual elements of $\widehat{g(\vartheta)}$ are thus the square roots of the diagonal elements of $Var[\widehat{g(\vartheta)}]$.

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 σ_j, $\varphi_j(1)$, and $\varphi_{i,j}(s)$ via the use of a VAR specified for X_t, as demonstrated in table 4.1. In this case, a parameterized version of (4.26), or equivalently (4.29), serves as the model used to generate artificial realizations of $\{X_t\}_{t=1}^T$; this model is known as the data generation process in the experiment. Consider the use of (4.29) 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 (4.29) first by obtaining a drawing of $\{e_t\}_{t=1}^T$ from a specified distribution, inserting e_1 into (4.29) to obtain

$$z_1 = Az_0 + e_1,$$

and then selecting the 1^{st}, $(p+1)^{\text{th}}, \ldots, [(m-1)p+1]^{\text{th}}$ 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 Σ, from which $\Gamma(0)$ and $\Gamma(1)$ are constructed using (4.30)–(4.32). Finally, extraction of σ_j, $\varphi_j(1)$, and $\varphi_{i,j}(s)$ from $\Gamma(0)$ and $\Gamma(1)$ yields a Monte Carlo drawing of these statistics.

Denote the i^{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

$$\overline{g(\hat{\vartheta})} = \frac{1}{N}\sum_{i=1}^{N} \widehat{g(\vartheta)}^{i}; \tag{4.47}$$

$$Var(g(\hat{\vartheta})) = \frac{1}{N}\sum_{i=1}^{N}\left[\widehat{g(\vartheta)}^{i} - \overline{g(\hat{\vartheta})}\right]^{2}. \tag{4.48}$$

The square root of the diagonal elements of $Var(g(\hat{\vartheta}))$ provide a single estimate of the standard error of $\widehat{g(\vartheta)}$. Denote this estimate as $s.e.[g(\hat{\vartheta})]$. Replication of this process using a total of J Monte Carlo experiments yields J drawings of $s.e.[g(\hat{\vartheta})]$. Letting the j^{th} drawing be given by $s.e.[g(\hat{\vartheta})]^{j}$, a natural estimator of $s.e.[g(\hat{\vartheta})]$ is the average computed over the J experiments:

$$\overline{s.e.[g(\hat{\vartheta})]} = \frac{1}{J}\sum_{j=1}^{J} s.e.[g(\hat{\vartheta})]^{j}. \tag{4.49}$$

The variance of this estimator is calculated as

$$Var\left[\overline{s.e.[g(\hat{\vartheta})]}\right] = \frac{1}{J}\sum_{j=1}^{J}\left\{s.e.[g(\hat{\vartheta})]^{j} - \overline{s.e.[g(\hat{\vartheta})]}\right\}^{2}, \tag{4.50}$$

and the standard deviation $s.e.\left[\overline{s.e.[g(\hat{\vartheta})]}\right]$ is once again the square root of the diagonal elements of the variance.

An assessment of the accuracy of $\overline{s.e.[g(\hat{\vartheta})]}$ 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 $\overline{s.e.[g(\hat{\vartheta})]}$ is therefore

$$nse\left\{\overline{s.e.[g(\hat{\vartheta})]}\right\} = \frac{s.e.\left[\overline{s.e.[g(\hat{\vartheta})]}\right]}{\sqrt{J}}. \tag{4.51}$$

A critical step in obtaining Monte Carlo estimates involves the design of the experiment. In the present context the design has three components: 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 use 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 4.4

Construct a Monte Carlo estimate of standard errors associated with estimates of the spectrum of a single variable y_t obtained using (4.36). 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 $\varepsilon_t \sim N(0,1)$.
2. Let $g(\hat{\vartheta})$ represent the $(n+1)\times 1$ vector of values of the spectrum $s_y(\omega)$, $\omega = 0, \pi/n, 2\pi/n, \ldots, \pi$, $n = 40$.
3. For each of 100 realizations of $\{y_t\}$ obtained from the DGP you constructed, estimate an ARMA(2,2) specification, and construct $\widehat{g(\vartheta)}$ by inserting the parameter estimates you obtain in (4.36).
4. Calculate the standard deviation of each element of $\widehat{g(\vartheta)}$ over the 100 drawings you obtained: this yields a single estimate of $s.e.[s_y(\omega)]$, $\omega = 0, \pi/n, 2\pi/n, \ldots, \pi$.
5. Repeat steps 3 and 4 100 times, and use the resulting 100 drawings of $s.e.[s_y(\omega)]$ to construct $\overline{s.e.[g(\hat{\vartheta})]}$, $s.e.\left[\overline{s.e.[g(\hat{\vartheta})]}\right]$, and $nse\left\{\overline{s.e.[g(\hat{\vartheta})]}\right\}$.

Under the Bayesian perspective, ϑ (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 $\widehat{g(\vartheta)}$ 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 9.

4.3 The Kalman Filter

We now turn to the foundation of full-information analyses of structural models: evaluation of the likelihood function via use of the Kalman filter. For convenience, we begin by repeating the characterization of the DSGE model solution as a collection of stochastic difference equations of the form

$$x_t = F(\mu)x_{t-1} + e_t \tag{4.52}$$

$$e_t = G(\mu)v_t \tag{4.53}$$

$$E(e_t e_t') = G(\mu)E(v_t v_t')G(\mu)' = Q(\mu). \tag{4.54}$$

Again, (4.52)–(4.54) comprise the state system, describing the evolution of the $n \times 1$ vector of model variables x_t. These are mapped into the $m \times 1$ vector of observable variables X_t via either

$$X_t = H(\mu)'x_t \tag{4.55}$$

or

$$X_t = H(\mu)'x_t + u_t, \qquad E(u_t u_t') = \Sigma_u, \tag{4.56}$$

both of which are known as measurement equations. The Kalman filter can be used to build likelihood functions based on various specifications of the measurement system, under the assumption of normality for $\{e_t\}$ and $\{u_t\}$.[8] Departures from normality, and from the linear structure considered here, prompt us to an alternative known as the particle filter, which is outlined in chapter 11. Here we present a general overview of the Kalman filter, and then make specific presentations for three leading specifications.

4.3.1 Overview

Full-information analyses entail calculations of the probability or likelihood associated with the realization of an observed sample $X \equiv \{X_t\}_{t=1}^T$. The Kalman filter is an algorithm designed to execute this calculation recursively, following the recursive nature of its associated structural model.

The idea behind the algorithm is to produce assessments of the conditional probability associated with the time-t observation X_t, given the history of past realizations $X^{t-1} \equiv \{X_j\}_{j=1}^{t-1}$. Denote this probability as $L(X_t|X^{t-1})$, with $L(X_1|X^0)$ denoting the unconditional likelihood associated with X_1. The sequence of conditional likelihoods $\{L(X_t|X^{t-1})\}_{t=1}^T$

[8] The GAUSS procedure `kalman.prc` is available for performing the calculations described in this section.

are independent across time, thus the likelihood associated with X is given by the product of the individual conditional likelihoods:

$$L(X) = \prod_{t=1}^{T} L(X_t|X^{t-1}).$$

Regarding the structure of $L(X_t|X^{t-1})$, this is most simply described for the case in which each of the elements of x_t is observable, so that $X_t \equiv x_t$. Conditional on $\{x_j\}_{j=1}^{t-1}$, from (4.52) we observe that the optimal forecast of x_t is given by

$$\widehat{x}_t = F(\mu)x_{t-1},$$

and the difference between the forecasted and observed value of x_t serves as the inferred value of e_t:

$$\widehat{e}_t = x_t - F(\mu)x_{t-1}.$$

The conditional likelihood associated with the observation of x_t can thus be assessed as the likelihood assigned to $\widehat{e}_t$ by its assumed probability distribution (say, p_e):

$$L(X_t|X^{t-1}) = p_e(\widehat{e}_t).$$

The details are slightly more complicated when certain elements of x_t are unobservable, but the basic idea is the same: conditional likelihoods represent probabilities associated with the realization of observables at time t, given the sequence of variables that were observed previously.

4.3.2 The Filter without Measurement Errors

When the measurement system is given by (4.55), then (4.52)–(4.55) are referred to as a state-space representation. The associated likelihood function is obtained by making a distributional assumption for $\{e_t\}$. As noted, we proceed here under the assumption of normality.

Consider the objective of calculating the value of the likelihood function for a given value of μ, which implies values for $[F(\mu), G(\mu), Q(\mu), H(\mu)]$ (hereafter, we revert to the notation under which the dependence of F, etc. on μ is taken as given). To establish intuition for how this is achieved given the presence of unobservable variables in x_t, consider first the simpler case in which each variable is instead observable, as in the previous subsection. In this case, the system consists of (4.52)–(4.54); likelihood evaluation is achieved via the iterative evaluation of a sequence of condi-

tional distributions for x_t, given x_{t-1}. The iteration begins by inserting x_1 into its unconditional distribution, which is $N(0, Q)$:

$$L(x_1|\mu) = (2\pi)^{-n/2}|Q^{-1}|^{1/2}\exp\left[-\frac{1}{2}(x_1'Q^{-1}x_1)\right].$$

Then for x_t, $t = 2, \ldots, T$, the iteration continues by inserting x_t into its conditional (on x_{t-1}) distribution, which is $N(Fx_{t-1}, Q)$:

$$L(x_t|\mu) = (2\pi)^{-n/2}|Q^{-1}|^{1/2} \times \exp\left[-\frac{1}{2}(x_t - Fx_{t-1})'Q^{-1}(x_t - Fx_{t-1})\right].$$

Finally, the sample likelihood is the product of the individual likelihoods:

$$L(x|\mu) = \prod_{t=1}^{T} L(x_t|\mu).$$

Exercise 4.5

Map the $AR(p)$ process specified for the univariate time series y_t in (4.17) into a representation of the form given in (4.52)–(4.54). Next, show that OLS and ML estimates of $\rho(L)$ coincide.

For the case in which x_t includes unobservables, the Kalman filter adds an additional step to this iterative process. This involves making an inference regarding the most likely specification for x_t, given the sequence of observations on $\{X_j\}$, $j = 1, \ldots t$.

Let $x_{t|t-1}$ be the conditional expectation of x_t given observations

$$\{X_1, \ldots, X_{t-1}\},$$

and

$$P_{t|t-1} = E[(x_t - x_{t|t-1})(x_t - x_{t|t-1})']$$

be its associated covariance matrix. The iterative process begins with the calculation of their initial (unconditional) values, given by

$$x_{1|0} = 0, \qquad P_{1|0} = FP_{1|0}F' + Q$$
$$\rightarrow vec(P_{1|0}) = (I - F \otimes F')^{-1}vec(Q). \tag{4.57}$$

These are used to construct associated values for $X_{1|0}$, given by

$$X_{1|0} = H'x_{1|0} = 0, \tag{4.58}$$
$$\Omega_{1|0} = E[(X_1 - X_{1|0})(X_1 - X_{1|0})'] = H'P_{1|0}H. \tag{4.59}$$

These serve as inputs for the likelihood function of X_1, which is $N(X_{1|0}, \Omega_{1|0})$:

$$L(X_1|\mu) = (2\pi)^{-m/2}|\Omega_{1|0}^{-1}|^{1/2}\exp\left[-\frac{1}{2}(X_1'\Omega_{1|0}{}^{-1}X_1)\right]. \tag{4.60}$$

Finally, the unconditional values $[x_{1|0}, P_{1|0}]$ are updated to take into account information conveyed by the observation of X_1. The updates yield conditional values $x_{1|1} \equiv x_1$ and $P_{1|1} \equiv P_1$:

$$x_{1|1} = x_{1|0} + P_{1|0}H\Omega_{1|0}^{-1}(X_1 - X_{1|0}) \tag{4.61}$$
$$P_{1|1} = P_{1|0} - P_{1|0}H\Omega_{1|0}^{-1}H'P_{1|0}. \tag{4.62}$$

Having accomplished initialization, the iterations involving X_t, $t = 2, \ldots T$ are identical. First, $x_{t|t-1}$ and $P_{t|t-1}$ are constructed:

$$x_{t|t-1} = Fx_{t-1} \tag{4.63}$$
$$P_{t|t-1} = FP_{t|t-1}F' + Q$$
$$\rightarrow vec(P_{t|t-1}) = (I - F \otimes F')^{-1}vec(Q). \tag{4.64}$$

These serve as inputs for the construction of $[X_{t|t-1}, \Omega_{t|t-1}]$:

$$X_{t|t-1} = H'x_{t|t-1} \tag{4.65}$$
$$\Omega_{t|t-1} = E[(X_t - X_{t|t-1})(X_t - X_{t|t-1})'] = H'P_{t|t-1}H. \tag{4.66}$$

These then facilitate the calculation of the likelihood function for X_t, which is $N(X_{t|t-1}, \Omega_{t|t-1})$:

$$L(X_t|\mu) = (2\pi)^{-m/2}|\Omega_{t|t-1}^{-1}|^{1/2} \times \exp\left[\frac{-1}{2}(X_t - X_{t|t-1})'\Omega_{t|t-1}{}^{-1}(X_t - X_{t|t-1})\right]. \tag{4.67}$$

Finally, X_t is fed into the updating system, yielding

$$x_{t|t} = x_{t|t-1} + P_{t|t-1}H\Omega_{t|t-1}^{-1}(X_t - X_{t|t-1}) \tag{4.68}$$
$$P_{t|t} = P_{t|t-1} - P_{t|t-1}H\Omega_{t|t-1}^{-1}H'P_{t|t-1}. \tag{4.69}$$

The sample likelihood is once again the product of the individual likelihoods:

$$L(X|\mu) = \prod_{t=1}^{T} L(X_t|\mu). \tag{4.70}$$

It is often the case that the time-series behavior of certain unobserved elements of x_t are of independent interest. It is possible to infer this behavior

as a by-product of the likelihood evaluation process by obtaining what are known as smoothed estimates of the unobservables: $\{\widehat{x}_{t|T}\}_{t=1}^{T}$. As indicated above, the steps of the Kalman filter yield values of

$$\{x_{t|t}\}_{t=1}^{T}, \{x_{t+1|t}\}_{t=0}^{T-1}, \{P_{t|t}\}_{t=1}^{T} \text{ and } \{P_{t+1|t}\}_{t=0}^{T-1}.$$

These values may be used to construct $\{\widehat{x}_{t|T}\}_{t=1}^{T}$, as follows. The smoothed estimate $\widehat{x}_{T|T}$ is simply the last entry of $\{x_{t|t}\}_{t=1}^{T}$. Next, the quantity $J_t = P_{t|t}F'P_{t+1|t}$ is computed so that the series $\{J_t\}_{t=1}^{T-1}$ is available. This allows the computation of

$$\widehat{x}_{T-1|T} = \widehat{x}_{T-1|T-1} + J_{T-1}(\widehat{x}_{T|T} - \widehat{x}_{T|T-1}), \tag{4.71}$$

which can be used for $t = T-2$ to compute

$$\widehat{x}_{T-1|T} = \widehat{x}_{T-1|T-1} + J_{T-1}(\widehat{x}_{T|T} - \widehat{x}_{T|T-1}). \tag{4.72}$$

Proceeding recursively in this manner yields the sequence of interest

$$\{\widehat{x}_{t|T}\}_{t=1}^{T},$$

with associated covariance given by

$$P_{t|T} = P_{t|t} + J_t(P_{t+1|T} - P_{t+1|t})J_t'. \tag{4.73}$$

Exercise 4.6

Repeat Steps 1–3 of the previous exercise using the ARMA model for y_t specified in (4.22), given $p = q = 1$.

4.3.3 *The Filter with Measurement Errors*

When there is measurement error associated with the observation of X_t, the measurement system is given by (4.56). The likelihood function is obtained in this case by making a distributional assumption for $\{e_t\}$ and the $m \times 1$ vector of measurement errors $\{u_t\}$. Once again, we proceed here under the assumption of normality for both e_t and u_t.[9] In addition, we assume that the measurement errors u_t are serially uncorrelated; in section 4.3.4 we consider the case in which this assumption is generalized.

Relative to the case in which there is no measurement error, many of the computations required to build the likelihood function remain the same. The main difference is that the initial unconditional values at which the filter begins are altered slightly. This implies changes in the updating formulae used in the construction of the likelihood function. In particular, $\Omega_{1|0}$ is now given by

$$\begin{aligned}\Omega_{1|0} &= E[(X_1 - X_{1|0})(X_1 - X_{1|0})'] \\ &= H'P_{1|0}H + \Sigma_u;\end{aligned} \tag{4.74}$$

analogously, $\Omega_{t|t-1}$ is given by:

$$\begin{aligned}\Omega_{t|t-1} &= E[(X_t - X_{t|t-1})(X_t - X_{t|t-1})'] \\ &= H'P_{t|t-1}H + \Sigma_u.\end{aligned} \tag{4.75}$$

The calculation of the likelihood function for X_t, which continues to be $N(X_{t|t-1}, \Omega_{t|t-1})$, is facilitated by replacing (4.66) by (4.75). Finally, smoothed estimates are obtained using exactly the same process described above, with (4.75) replacing (4.66) prior to the computation of (4.71)–(4.73).

Exercise 4.7

Consider the following model relating consumption (y_t) to income (z_t), taste shocks (x_t), and subject to an error in measurement (ε_t)

$$y_t = a_1 z_t + a_2 x_t + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma^2).$$

Assume that taste shocks follow an AR(1) specification given by

$$x_t = \rho x_{t-1} + \epsilon_t, \quad \epsilon_t \sim N(0, \varsigma^2).$$

1. Specify the model in state-space form.
2. Initialize the Kalman filter.
3. Specify the likelihood function of the model expressed in Kalman filter form.

4.3.4 *Serially Correlated Measurement Errors*

As a final case, suppose that the measurement errors in (4.56) are serially correlated, obeying:

$$u_t = \Gamma u_{t-1} + \xi_t \tag{4.76}$$

$$E(\xi_t \xi_t') = \Sigma_\xi. \tag{4.77}$$

Sargent (1989) demonstrated the evaluation of DSGE models featuring diagonal specifications for the $m \times m$ matrices Γ and Σ_ξ, and Ireland

[9] The GAUSS procedure `kalmanm.prc` is available for performing the calculations described in this section.

(2004b) demonstrated an extension to the non-diagonal case. In either case, the evaluation of the likelihood function associated with (4.52)–(4.54), (4.56), and (4.76)–(4.77) turns out to be identical to the case in which the observation equation is given by (4.55), and thus the system has no measurement error. To see why, consider an augmentation of the state system given by

$$\zeta_t = \begin{bmatrix} x_t \\ u_t \end{bmatrix} \quad \text{and} \quad \gamma_t = \begin{bmatrix} G\upsilon_t \\ \xi_t \end{bmatrix}. \tag{4.78}$$

Given this augmentation, the state-space representation may be written as

$$\zeta_t = \Theta_0 \zeta_{t-1} + \gamma_t \tag{4.79}$$

$$X_t = \Theta_1 \zeta_t \tag{4.80}$$

$$\Theta_0 = \begin{bmatrix} F & 0 \\ 0 & \Gamma \end{bmatrix}, \quad \Theta_1 = \begin{bmatrix} H \\ I \end{bmatrix}' \tag{4.81}$$

$$Q' = E(\gamma_t \gamma_t') = \begin{bmatrix} Q & 0 \\ 0 & \Sigma_\xi \end{bmatrix}, \tag{4.82}$$

which is exactly in the form of (4.52)–(4.55). Thus replacing F and H with Θ_0 and Θ_1 in section 4.3.2 enables the construction of the likelihood function; likewise, this replacement enables the calculation of smoothed estimates of ζ_t.

As a preview of issues to be discussed in chapter 8, we conclude this chapter 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 μ. 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 μ, 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 8, paying particular attention to peculiarities that arise in the context of the analysis of DSGE models.

Chapter 5

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 PROVIDED background for preparing structural models for empirical analysis. Recall that the first step of the preparation stage is the construction of a linear approximation of the structural model under investigation, which takes the form

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

This chapter demonstrates the completion of this first step for three prototypical model environments that will serve as examples throughout the remainder of the text. This will set the stage for part II, which outlines and demonstrates alternative approaches to pursuing empirical analysis. (For guidance regarding the completion of this step for a far 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 (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 collection of 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 neo-Keynesian tradition. Its empirical purpose is to simultaneously 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).

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,b), Stokey and Lucas (1989), Blanchard and Fischer (1998), Romer (2001), and Ljungqvist and Sargent (2004).

5.1 Model I: A Real Business Cycle Model

5.1.1 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:

$$\max_{c_t, l_t} \quad U = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t, l_t). \tag{5.1}$$

In (5.1), 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

$$y_t = z_t f(k_t, n_t), \tag{5.2}$$

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:

$$1 = n_t + l_t. \tag{5.3}$$

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:

$$y_t = c_t + i_t, \tag{5.4}$$

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

$$k_{t+1} = i_t + (1-\delta)k_t, \tag{5.5}$$

where $\delta \in (0,1)$ denotes the depreciation rate. The household's problem is to maximize (5.1) subject to (5.2)–(5.5), 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 (5.4), higher consumption today implies lower investment (savings), and thus from (5.5), less capital available for production tomorrow. The other is a labor/leisure trade-off: from (5.3), 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

$$\lim_{c\to 0} \frac{\partial u(c_t, l_t)}{\partial c_t} = \lim_{l\to 0} \frac{\partial u(c_t, l_t)}{\partial l_t} = \infty. \tag{5.6}$$

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 II.

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:

$$u(c_t, l_t) = \left(\frac{c_t^{\varphi} l_t^{1-\varphi}}{1-\phi} \right)^{1-\phi}, \tag{5.7}$$

or when $\phi = 1$, $u(\cdot) = \log(\cdot)$. 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 ϕ, the more intense is the household's interest in maintaining a smooth consumption/leisure profile. Also, $\varphi \in (0,1)$ indicates the

importance of consumption relative to leisure in determining instantaneous utility.

Next, the production function is of the Cobb-Douglas variety:

$$y_t = z_t k_t^{\alpha} n_t^{1-\alpha}, \tag{5.8}$$

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:

$$\log z_t = (1-\rho)\log(\bar{z}) + \rho \log z_{t-1} + \varepsilon_t \tag{5.9}$$

$$\varepsilon_t \sim NID(0, \sigma^2), \quad \rho \in (-1, 1). \tag{5.10}$$

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 10 of this book). Necessary conditions associated with the household's problem expressed in general terms are given by

$$\frac{\partial u(c_t, l_t)}{\partial l_t} = \left\{ \frac{\partial u(c_t, l_t)}{\partial c_t} \right\} \times \left\{ \frac{\partial f(k_t, n_t)}{\partial n_t} \right\} \tag{5.11}$$

$$\frac{\partial u(c_t, l_t)}{\partial c_t} = \beta E_t \left\{ \frac{\partial u(c_{t+1}, l_{t+1})}{\partial c_{t+1}} \left[\frac{\partial f(k_{t+1}, n_{t+1})}{\partial k_{t+1}} + (1-\delta) \right] \right\}. \tag{5.12}$$

The intratemporal optimality condition (5.11) 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 (5.12) 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 (5.11) and (5.12) for the impact of a positive productivity shock on the household's labor/leisure and consumption/savings decisions. From (5.11), higher labor productivity implies a higher opportunity cost of leisure, prompting a reduction in leisure time in favor of labor time. From (5.12), 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 con-

sumption 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 (5.11) by the expression for the marginal utility of consumption, and using the functional forms introduced above, these conditions can be written as

$$\left(\frac{1-\varphi}{\varphi}\right)\frac{c_t}{l_t} = (1-\alpha)z_t\left(\frac{k_t}{n_t}\right)^{\alpha} \tag{5.13}$$

$$c_t^{\varphi(1-\phi)-1} l_t^{(1-\varphi)(1-\phi)} = \beta E_t\left\{c_{t+1}^{\varphi(1-\phi)-1} l_{t+1}^{(1-\varphi)(1-\phi)}\left[\alpha z_{t+1}\left(\frac{n_{t+1}}{k_{t+1}}\right)^{1-\alpha} + (1-\delta)\right]\right\}. \tag{5.14}$$

5.1.2 The Nonlinear System

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

$$\left(\frac{1-\varphi}{\varphi}\right)\frac{c_t}{l_t} = (1-\alpha)z_t\left(\frac{k_t}{n_t}\right)^{\alpha} \tag{5.15}$$

$$c_t^{\kappa} l_t^{\lambda} = \beta E_t\left\{c_{t+1}^{\kappa} l_{t+1}^{\lambda}\left[\alpha z_{t+1}\left(\frac{n_{t+1}}{k_{t+1}}\right)^{1-\alpha} + (1-\delta)\right]\right\} \tag{5.16}$$

$$y_t = z_t k_t^{\alpha} n_t^{1-\alpha} \tag{5.17}$$

$$y_t = c_t + i_t \tag{5.18}$$

$$k_{t+1} = i_t + (1-\delta)k_t \tag{5.19}$$

$$1 = n_t + l_t \tag{5.20}$$

$$\log z_t = (1-\rho)\log(\bar{z}) + \rho \log z_{t-1} + \varepsilon_t, \tag{5.21}$$

where

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

and

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

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:

$$\begin{aligned}
\frac{\bar{y}}{\bar{n}} &= \eta, \\
\frac{\bar{c}}{\bar{n}} &= \eta - \delta\theta, \\
\frac{\bar{i}}{\bar{n}} &= \delta\theta, \\
\bar{n} &= \frac{1}{1 + \left(\frac{1}{1-\alpha}\right)\left(\frac{1-\varphi}{\varphi}\right)\left[1 - \delta\theta^{1-\alpha}\right]}, \\
\bar{l} &= 1 - \bar{n}, \\
\frac{\bar{k}}{\bar{n}} &= \theta,
\end{aligned} \tag{5.22}$$

where

$$\theta = \left(\frac{\alpha}{1/\beta - 1 + \delta}\right)^{\frac{1}{1-\alpha}}$$

$$\eta = \theta^{\alpha}.$$

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:

$$z_t = z_0(1+g)^t e^{\omega_t}, \tag{5.23}$$

$$\omega_t = \rho\omega_{t-1} + \varepsilon_t. \tag{5.24}$$

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 $\log z_t$ given by (5.21). 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 (5.19). Trend removal here involves division of both

sides of (5.19) by $\left(1+\frac{g}{1-\alpha}\right)^{t}$; however, the trend component associated with k_{t+1} is $\left(1+\frac{g}{1-\alpha}\right)^{t+1}$, so the specification in terms of detrended variables is

$$\left(1+\frac{g}{1-\alpha}\right)k_{t+1} = i_t + (1-\delta)k_t. \tag{5.25}$$

Likewise, a residual trend factor will be associated with c_{t+1} in the intertemporal optimality condition (5.16). Because c_{t+1} is raised to the power

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

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

$$c_t^{\kappa} l_t^{\lambda} = \beta E_t \times \left\{\left(1+\frac{g}{1-\alpha}\right)^{\kappa} c_{t+1}^{\kappa} l_{t+1}^{\lambda}\left[\alpha z_{t+1}\left(\frac{n_{t+1}}{k_{t+1}}\right)^{1-\alpha} + (1-\delta)\right]\right\}. \tag{5.26}$$

With κ negative (insured 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 5.1

Rederive the steady state expressions (5.22) by replacing (5.19) with (5.25), and (5.16) with (5.26). Interpret the intuition behind the impact of g on the expressions you derive.

5.1.3 Linearization

The linearization step involves taking a log-linear approximation of the model at steady state values. In this case, the objective is to map (5.15)–(5.21) into the linearized system

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

for eventual empirical evaluation. Regarding D, dropping E_t from the Euler equation (5.16) 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 (5.21) implies

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

Regarding A and B, using the solution methodology discussed in chapter 2, these can be constructed by introducing the following system of equations into a gradient procedure (where time subscripts are dropped so that, e.g., $y = y_t$ and $y' = y_{t+1}$):

$$0 = \log\left(\frac{1-\varphi}{\varphi}\right) + \log c' - \log l' - \log(1-\alpha) - \log z' - \alpha\log k + \alpha\log n' \tag{5.27}$$

$$0 = \kappa\log c + \lambda\log l - \log\beta - \kappa\log c' - \lambda\log l' - \log\left[\alpha\exp(\log z')\frac{\exp[(1-\alpha)\log n']}{\exp[(1-\alpha)\log k']} + (1-\delta)\right] \tag{5.28}$$

$$0 = \log y' - \log z' - \alpha\log k - (1-\alpha)\log n' \tag{5.29}$$

$$0 = \log y' - \log\{\exp[\log(c')] + \exp[\log(i')]\} \tag{5.30}$$

$$0 = \log k' - \log\{\exp[\log(i')] + (1-\delta)\exp[\log(k)]\} \tag{5.31}$$

$$0 = -\log\{\exp[\log(n')] + \exp[\log(l')]\} \tag{5.32}$$

$$0 = \log z' - \rho\log z. \tag{5.33}$$

The mapping from (5.15)–(5.21) to (5.27)–(5.33) 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 $y = \exp(\log(y))$; 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 $\log y'$, and so on evaluated at steady state values yield A, and derivatives taken with respect to $\log y$, and so on yield $-B$. Note that capital installed at time t is not productive until period $t+1$; thus k rather than k' appears in (5.29).

Having obtained A, B, C, and D, the system can be solved using any of the solution methods outlined in chapter 2 to obtain a system of the form

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

This system can then be evaluated empirically using any of the methods described in part II.

Exercise 5.2

With x_t given by

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

and

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

show that the steady state values of the model are $\bar{y} = 0.9$, $\bar{c} = 0.7$, $\bar{i} = 0.2$, $\bar{n} = 0.47$, $\bar{l} = 0.53$, and $\bar{k} = 3.5$ (and take as granted $\bar{z} = 1$). Next, use a numerical gradient procedure to derive

$$A = \begin{bmatrix} 0 & 1 & 0 & 0.33 & -1 & 0 & -1 \\ 0 & 1.5 & 0 & -0.12 & 0.5 & 0 & -0.17 \\ 1 & 0 & 0 & -0.67 & 0 & 0 & -1 \\ 1 & -0.77 & -0.23 & 0 & 0 & 0 & 0 \\ 0 & 0 & -0.18 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & -0.47 & -0.53 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{bmatrix}$$

$$B = \begin{bmatrix} 0 & 0 & 0 & 0 & 0 & 0.33 & 0 \\ 0 & 1.5 & 0 & 0 & 0.5 & -0.9 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.33 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.77 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.9 \end{bmatrix}.$$

Exercise 5.3

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 5.1, and using (5.25) and (5.26) in place of (5.19) and (5.16).

5.2 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 from chapter 2 that it is possible to represent a system of arbitrary order using the first-order form taken by

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

given appropriate specification of the elements of x_t. Second, the problem of mapping implications carried by a stationary model into the behavior of non-stationary data is revisited from an alternative perspective than 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.

5.2.1 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.

HOUSEHOLDS

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

$$\max_{c_t, m_t, n_t} \quad U = E_0 \sum_{t=0}^{\infty} \beta^t \left\{ a_t \log c_t + \log \frac{m_t}{p_t} - \frac{n_t^{\xi}}{\xi} \right\} \tag{5.34}$$

$$s.t. \quad p_t c_t + \frac{b_t}{r_t} + m_t = m_{t-1} + b_{t-1} + \tau_t + w_t n_t + d_t, \tag{5.35}$$

where $\beta \in (0,1)$ and $\xi \geq 1$. According to the budget constraint (5.35), 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 τ_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 (5.34) 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

$$\left(\frac{w_t}{p_t}\right)\left(\frac{a_t}{c_t}\right) = n_t^{\xi-1} \tag{5.36}$$

$$\beta E_t\left\{\left(\frac{1}{p_{t+1}}\right)\left(\frac{a_{t+1}}{c_{t+1}}\right)\right\} = \left(\frac{1}{r_t p_t}\right)\left(\frac{a_t}{c_t}\right) \tag{5.37}$$

$$\left(\frac{m_t}{p_t}\right)^{-1} + \beta E_t\left\{\left(\frac{1}{p_{t+1}}\right)\left(\frac{a_{t+1}}{c_{t+1}}\right)\right\} = \left(\frac{1}{p_t}\right)\left(\frac{a_t}{c_t}\right). \tag{5.38}$$

Exercise 5.4

Interpret how (5.36)–(5.38) 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^{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:

$$\max_{y_{it}} \quad \Pi_t^F = p_t y_t - \int_0^1 p_{it} y_{it}\, di \tag{5.39}$$

$$s.t. \quad y_t = \left\{\int_0^1 y_{it}^{\frac{\theta_t-1}{\theta_t}}\, di\right\}^{\frac{\theta_t}{\theta_t-1}}. \tag{5.40}$$

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

$$y_{it} = y_t\left\{\frac{p_{it}}{p_t}\right\}^{-\theta_t} \tag{5.41}$$

$$p_t = \left\{\int_0^1 p_{it}^{1-\theta_t}\, di\right\}^{\frac{1}{1-\theta_t}}. \tag{5.42}$$

Notice that θ_t is the markup of price above marginal cost; randomness in θ_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:

$$\max_{p_{it}} \quad \Pi_{it}^I = E_0\sum_{t=0}^{\infty}\beta^t\left(\frac{a_t}{c_t}\right)\left(\frac{d_t}{p_t}\right), \tag{5.43}$$

$$s.t. \quad y_{it} = z_t n_{it} \tag{5.44}$$

$$y_{it} = y_t\left\{\frac{p_{it}}{p_t}\right\}^{-\theta_t} \tag{5.45}$$

$$\chi(p_{it}, p_{it-1}) = \frac{\phi}{2}\left[\frac{p_{it}}{\overline{\pi} p_{it-1}} - 1\right]^2 y_t, \quad \phi > 0, \tag{5.46}$$

where $\overline{\pi}$ is the gross inflation rate targeted by the monetary authority (described below), and the real value of dividends in (5.43) is given by

$$\frac{d_t}{p_t} = \left\{\frac{p_{it} y_{it} - w_t n_{it}}{p_t} - \chi(p_{it}, p_{it-1})\right\}. \tag{5.47}$$

The associated first-order condition can be written as

$$\begin{aligned}(\theta_t - 1)&\left(\frac{p_{it}}{p_t}\right)^{-\theta_t}\frac{y_t}{p_t} \\ &= \theta_t\left(\frac{p_{it}}{p_t}\right)^{-\theta_t-1}\frac{w_t}{p_t}\frac{y_t}{z_t}\frac{1}{p_t} - \left\{\phi\left[\frac{p_{it}}{\overline{\pi} p_{it-1}} - 1\right]\frac{y_t}{\overline{\pi} p_{it-1}} - \beta\phi\right. \\ &\quad \left.\times E_t\left(\frac{a_{t+1}}{a_t}\frac{c_t}{c_{t+1}}\left(\frac{p_{it+1}}{\overline{\pi} p_{it}} - 1\right)\frac{y_{t+1} p_{it+1}}{\overline{\pi} p_{it}^2}\right)\right\}.\end{aligned} \tag{5.48}$$

The left-hand side of (5.48) 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 ϕ 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

$$\widetilde{r}_t = \rho_r \widetilde{r}_{t-1} + \rho_\pi \widetilde{\pi}_t + \rho_g \widetilde{g}_t + \rho_o \widetilde{o}_t + \varepsilon_{rt}, \qquad \varepsilon_{rt} \sim iidN(0, \sigma_r^2), \tag{5.49}$$

where $\widetilde{\pi}_t$ is the gross inflation rate, $\widetilde{g}_t$ is the gross growth rate of output, and $\widetilde{o}_t$ is the output gap (defined below). The ρ_i parameters denote elasticities. The inclusion of $\widetilde{r}_{t-1}$ 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 $\widehat{y}_t$. 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:

$$\max_{\widehat{y}_t, n_{it}} \quad U^S = E_0 \sum_0^\infty \beta^t \left\{ a_t \log \widehat{y}_t - \frac{1}{\xi} \left(\int_0^1 n_{it}\, di \right)^\xi \right\} \tag{5.50}$$

$$s.t. \quad \widehat{y}_t = z_t \left(\int_0^1 n_{it}^{\frac{\theta_t - 1}{\theta_t}}\, di \right)^{\frac{\theta_t}{\theta_t - 1}}. \tag{5.51}$$

The symmetric solution to this problem is simply

$$\widehat{y}_t = a_t^{\frac{1}{\xi}} z_t. \tag{5.52}$$

STOCHASTIC SPECIFICATION

In addition to the monetary policy shock ε_{rt} introduced in (5.49), the model features a demand shock a_t, a technology shock z_t, and a cost-push shock θ_t. The former is *iid*; the latter three evolve according to

$$\log(a_t) = (1 - \rho_a)\log(\overline{a}) + \rho_a \log(a_{t-1}) + \varepsilon_{at}, \qquad \overline{a} > 1 \tag{5.53}$$

$$\log(z_t) = \log(\overline{z}) + \log(z_{t-1}) + \varepsilon_{zt}, \qquad \overline{z} > 1 \tag{5.54}$$

$$\log(\theta_t) = (1 - \rho_\theta)\log(\overline{\theta}) + \rho_\theta \log(\theta_{t-1}) + \varepsilon_{\theta t}, \qquad \overline{\theta} > 1, \tag{5.55}$$

with $|\rho_i| < 1$, $i = a, \theta$. Note that the technology shock is non-stationary: 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:

$$y_{it} = y_t, \qquad n_{it} = n_t, \qquad p_{it} = p_t, \qquad d_{it} = d_t. \tag{5.56}$$

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

$$m_t = m_{t-1} + \tau_t \tag{5.57}$$

$$b_t = b_{t-1} = 0. \tag{5.58}$$

5.2.2 *The Nonlinear System*

In its current form, the model consists of twelve 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 (5.49)), 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:

$$\ddot{y}_t = \frac{y_t}{z_t}, \quad \ddot{c}_t = \frac{c_t}{z_t}, \quad \ddot{\widehat{y}}_t = \frac{\widehat{y}_t}{z_t}, \qquad \pi_t = \frac{p_t}{p_{t-1}},$$

$$\ddot{d}_t = \frac{(d_t/p_t)}{z_t}, \qquad \ddot{w}_t = \frac{(w_t/p_t)}{z_t}, \qquad \ddot{m}_t = \frac{(m_t/p_t)}{z_t}, \quad \ddot{z}_t = \frac{z_t}{z_{t-1}}.$$

Using the expression for real dividends given by (5.47), the household's budget constraint in equilibrium is rewritten as

$$\ddot{y}_t = \ddot{c}_t + \frac{\phi}{2} \left(\frac{\pi_t}{\overline{\pi}} - 1 \right)^2 \ddot{y}_t. \tag{5.59}$$

Next, the household's first-order condition (5.37) is written in normalized form as

$$\frac{a_t}{\ddot{c}_t} = \beta r_t E_t \left\{ \frac{a_{t+1}}{\ddot{c}_{t+1}} \times \frac{1}{\ddot{z}_{t+1}} \times \frac{1}{\pi_{t+1}} \right\}. \tag{5.60}$$

Next, the household's remaining first-order conditions, the expression for the real dividend payment (5.47) 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:

$$o_t \equiv \frac{y_t}{\widehat{y}_t} = \frac{\ddot{y}_t}{a_t^{\frac{1}{\xi}}}. \tag{5.61}$$

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

$$\ddot{y}_t = \ddot{c}_t + \frac{\phi}{2}\left(\frac{\pi_t}{\overline{\pi}} - 1\right)^2 \ddot{y}_t \tag{5.62}$$

$$\frac{a_t}{\ddot{c}_t} = \beta r_t E_t \left\{ \frac{a_{t+1}}{\ddot{c}_{t+1}} \times \frac{1}{\ddot{z}_{t+1}} \times \frac{1}{\pi_{t+1}} \right\} \tag{5.63}$$

$$\begin{aligned} 0 = 1 - \theta_t + \theta_t \frac{\ddot{c}_t}{a_t} \ddot{y}_t^{\xi-1} - \phi\left(\frac{\pi_t}{\overline{\pi}} - 1\right)\frac{\pi_t}{\overline{\pi}} \\ + \beta\phi E_t \left\{ \frac{\ddot{c}_t a_{t+1}}{\ddot{c}_{t+1} a_t} \left(\frac{\pi_{t+1}}{\overline{\pi}} - 1\right) \frac{\pi_{t+1}}{\overline{\pi}} \frac{\ddot{y}_{t+1}}{\ddot{y}_t} \right\} \end{aligned} \tag{5.64}$$

$$g_t = \frac{\ddot{z}_t \ddot{y}_t}{\ddot{y}_{t-1}} \tag{5.65}$$

$$o_t = \frac{y_t}{\widehat{y}_t} = \frac{\ddot{y}_t}{a_t^{\frac{1}{\xi}}} \tag{5.66}$$

$$\log(a_t) = (1-\rho_a)\log(\overline{a}) + \rho_a \log(a_{t-1}) + \varepsilon_{at} \tag{5.67}$$

$$\log(\theta_t) = (1-\rho_\theta)\log(\overline{\theta}) + \rho_\theta \log(\theta_{t-1}) + \varepsilon_{\theta t} \tag{5.68}$$

$$\log(\ddot{z}_t) = \log(\overline{z}) + \varepsilon_{zt} \tag{5.69}$$

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

5.2.3 Linearization

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

$$\begin{aligned} \overline{r} &= \frac{\overline{z}}{\beta}\overline{\pi}, \\ \overline{c} = \overline{y} &= \left(\overline{a}\frac{\overline{\theta}-1}{\overline{\theta}}\right)^{\frac{1}{\xi}}, \\ \overline{o} &= \left(\frac{\overline{\theta}-1}{\overline{\theta}}\right)^{\frac{1}{\xi}}; \end{aligned} \tag{5.70}$$

(5.62)–(5.69) are then log-linearized around these values. As with model I, 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 re-configuration 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.

As a first step, the variables appearing in (5.62)–(5.69) are written in logged form. Log-linearization of (5.62) then yields

$$\widetilde{y}_t \equiv \log\left(\frac{\ddot{y}_t}{\overline{y}}\right) = \widetilde{c}_t,$$

because the partial derivative of $\widetilde{y}_t$ with respect to $\widetilde{\pi}_t$ (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 $\widetilde{c}_t$ is replaced by $\widetilde{y}_t$ in the remaining equations.

Next, recalling that $E_t(\widetilde{z}_{t+1}) = 0$, log-linearization of (5.63) yields

$$0 = \widetilde{r}_t - E_t\widetilde{\pi}_{t+1} - (E_t\widetilde{y}_{t+1} - \widetilde{y}_t) + E_t\widetilde{y}_{t+1} - \widetilde{a}_t. \tag{5.71}$$

Relating output and the output gap via the log-linearization of (5.66),

$$\widetilde{y}_t = \frac{1}{\xi}\widetilde{a}_t + \widetilde{o}_t, \tag{5.72}$$

the term $E_t(\widetilde{y}_{t+1}) - \widetilde{y}_t$ can be substituted out of (5.71), yielding the IS curve:

$$\widetilde{o}_t = E_t\widetilde{o}_{t+1} - (\widetilde{r}_t - E_t\widetilde{\pi}_{t+1}) + (1-\xi^{-1})(1-\rho_a)\widetilde{a}_t. \tag{5.73}$$

Similarly, log-linearizing (5.64) and eliminating $\widetilde{y}_t$ using (5.72) yields the Phillips curve:

$$\widetilde{\pi}_t = \beta E_t \widetilde{\pi}_{t+1} + \psi \widetilde{o}_t - \widetilde{e}_t, \tag{5.74}$$

where $\psi = \frac{\xi(\theta-1)}{\phi}$ and $\widetilde{e}_t = \frac{1}{\phi}\widetilde{\theta}_t$. 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, $\widetilde{o}_{t-1}$ as $\widetilde{o}^-$, the system is given by

$$\widetilde{o} = \alpha_o \widetilde{o}^- + (1-\alpha_o) E_t \widetilde{o}' - (\widetilde{r} - E_t \widetilde{\pi}') + (1-\xi^{-1})(1-\rho_a)\widetilde{a} \tag{5.75}$$

$$\widetilde{\pi} = \beta\alpha_\pi \widetilde{\pi}^- + \beta(1-\alpha_\pi) E_t \widetilde{\pi}' + \psi \widetilde{o} - \widetilde{e} \tag{5.76}$$

$$\widetilde{g}' = \widetilde{y}' - \widetilde{y} + \widetilde{z}' \tag{5.77}$$

$$\widetilde{o}' = \widetilde{y}' - \xi^{-1}\widetilde{a}' \tag{5.78}$$

$$\widetilde{r}' = \rho_r \widetilde{r} + \rho_\pi \widetilde{\pi}' + \rho_g \widetilde{g}' + \rho_o \widetilde{o}' + \varepsilon_r' \tag{5.79}$$

$$\widetilde{a}' = \rho_a \widetilde{a} + \varepsilon_a' \tag{5.80}$$

$$\widetilde{e}' = \rho_e \widetilde{e} + \varepsilon_e' \tag{5.81}$$

$$\widetilde{z}' = \varepsilon_z' \tag{5.82}$$

where the structural shocks

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

are *iidN* with diagonal covariance matrix Σ. The additional parameters introduced are $\alpha_o \in [0,1]$ and $\alpha_\pi \in [0,1]$; setting $\alpha_o = \alpha_\pi = 0$ 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

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

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

$$x_t \equiv [\widetilde{o}_t \quad \widetilde{o}_{t-1} \quad \widetilde{\pi}_t \quad \widetilde{\pi}_{t-1} \quad \widetilde{y}_t \quad \widetilde{r}_t \quad \widetilde{g}_t \quad \widetilde{a}_t \quad \widetilde{e}_t \quad \widetilde{z}_t]'.$$

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

$$A = \begin{bmatrix}
-(1-\alpha_0) & 1 & -1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & -\psi & -\beta(1-\alpha_\pi) & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 1 & 0 & 0 & -1 \\
1 & 0 & 0 & 0 & -1 & 0 & 0 & \xi^{-1} & 0 & 0 \\
-\rho_0 & 0 & -\rho_\pi & 0 & 0 & 1 & -\rho_g & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}, \tag{5.83}$$

$$B = \begin{bmatrix}
0 & \alpha_o & 0 & 0 & 0 & -1 & 0 & (1-\xi^{-1})(1-\rho_a) & 0 & 0 \\
0 & 0 & 0 & \beta\alpha_\pi & 0 & 0 & 0 & 0 & -1 & 0 \\
0 & 0 & 0 & 0 & -1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & \rho_r & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho_a & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & \rho_e & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0
\end{bmatrix}. \tag{5.84}$$

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

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

and

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

the matrices C and D are given by

$$C = \begin{bmatrix}
 & 0_{4\times 4} & & \\
1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 \\
 & 0_{2\times 4} & &
\end{bmatrix}, \qquad D = \begin{bmatrix}
-(1-\alpha_0) & -1 \\
0 & -\beta(1-\alpha_\pi) \\
0_{2\times 8} &
\end{bmatrix}. \tag{5.85}$$

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 π_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 π_t are stationary; the additional assumption of stationarity for r_t is all that is necessary to proceed with the analysis.

Exercise 5.5

Solve the linearized system (5.75)–(5.82) using any of the methods outlined in chapter 2. Note that the vector of deep parameters is now given by:

$$\mu = [\overline{z}\ \overline{\pi}\ \beta\ \omega\ \theta\ \phi\ \alpha_x\ \alpha_\pi\ \rho_r\ \rho_\pi\ \rho_g\ \rho_x\ \rho_a\ \rho_\theta\ \sigma_a\ \sigma_\theta\ \sigma_z\ \sigma_r]'.$$

Exercise 5.6

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

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

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

5.3 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, alternative functional forms are introduced. 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).

5.3.1 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

$$\max_{c_t} \quad U = E_0 \sum_{t=0}^{\infty} \beta^t u(c_t), \tag{5.86}$$

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

$$c_t + p_t(s_t - s_{t-1}) = d_t s_{t-1} + q_t. \tag{5.87}$$

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

$$p_t = \beta E_t\left[\frac{u'(d_{t+1} + q_{t+1})}{u'(d_t + q_t)}(d_{t+1} + p_{t+1})\right]. \tag{5.88}$$

From (5.88), 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 (5.88) 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

$$\log d_t = (1-\rho_d)\log(\overline{d}) + \rho_d \log(d_{t-1}) + \varepsilon_{dt} \tag{5.89}$$

$$\log q_t = (1-\rho_q)\log(\overline{q}) + \rho_q \log(q_{t-1}) + \varepsilon_{qt}, \tag{5.90}$$

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

$$\begin{bmatrix} \varepsilon_{dt} \\ \varepsilon_{qt} \end{bmatrix} \sim iidN(0, \Sigma). \tag{5.91}$$

5.3.2 *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

$$c_t + p_t^e(s_t^e - s_{t-1}^e) + p_t^f s_t^f = d_t s_{t-1}^e + s_{t-1}^f + q_t, \tag{5.92}$$

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 (5.88) established in the single-asset specification. Rearranging slightly:

$$\beta E_t\left\{\frac{u'(c_{t+1})}{u'(c_t)} \times \frac{p_{t+1}^e + d_t}{p_t^e}\right\} = 1 \tag{5.93}$$

$$\beta E_t\left\{\frac{u'(c_{t+1})}{u'(c_t)} \times \frac{1}{p_t^f}\right\} = 1. \tag{5.94}$$

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

$$\beta E_t\left\{\frac{u'(c_{t+1})}{u'(c_t)} r_{t+1}^f\right\} = 1 \tag{5.95}$$

$$E_t\left\{\frac{u'(c_{t+1})}{u'(c_t)}\left[r_{t+1}^e - r_{t+1}^f\right]\right\} = 0, \tag{5.96}$$

where (5.96) is derived by subtracting (5.94) from (5.93). 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 (5.96), and again taking $\{c_t\}$ as given, the average value observed for r^f is far too low to reconcile with (5.95). This second component, as emphasized by Weil (1989), is the risk-free rate puzzle.

5.3.3 *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

$$u(c_t) = \frac{c_t^{1-\gamma}}{1-\gamma}, \tag{5.97}$$

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

$$p_t = \beta E_t\left[\frac{(d_{t+1} + q_{t+1})^{-\gamma}}{(d_t + q_t)^{-\gamma}}(d_{t+1} + p_{t+1})\right]. \tag{5.98}$$

Notice that, *ceteris paribus*, a relatively large value of γ 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 $\overline{d}$ and $\overline{q}$ are simply parameters. Normalizing $\overline{d}$ to 1 and defining $\eta = \frac{\overline{q}}{\overline{d}}$, so that $\eta = \overline{q}$, the steady state value of consumption (derived from the budget constraint) is $\overline{c} = 1 + \eta$. And from the pricing equation,

$$\begin{aligned}\overline{p} &= \frac{\beta}{1-\beta}\overline{d} \\ &= \frac{\beta}{1-\beta}. \end{aligned} \tag{5.99}$$

Letting $\beta = 1/(1+r)$, where r denotes the household's discount rate, (5.99) implies $\overline{p}/\overline{d} = 1/r$. 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 β under this specification of preferences.

Exercise 5.7

Linearize the pricing equation (5.98) around the model's steady state values.

HABIT/DURABILITY

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

$$u(h_t) = \frac{h_t^{1-\gamma}}{1-\gamma}, \tag{5.100}$$

with

$$h_t = h_t^d - \alpha h_t^h, \tag{5.101}$$

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

$$h_t^d = \sum_{j=0}^{\infty} \delta^j c_{t-j} \tag{5.102}$$

$$\begin{aligned} h_t^h &= (1-\theta)\sum_{j=0}^{\infty}\theta^j h_{t-1-j}^d \\ &= (1-\theta)\sum_{j=0}^{\infty}\theta^j \sum_{i=0}^{\infty}\delta^i c_{t-1-i} \end{aligned} \tag{5.103}$$

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 δ. 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 consumption complementarity. The variable h_t represents the current level of durable services net of the average of past services; the parameter α 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, h_t becomes

$$\begin{aligned} h_t &= c_t + \sum_{j=1}^{\infty}\left[\delta^j - \alpha(1-\theta)\sum_{i=0}^{j-1}\delta^i\theta^{j-i-1}\right]c_{t-j} \\ &\equiv \sum_{j=0}^{\infty}\Phi_j c_{t-j}, \end{aligned} \tag{5.104}$$

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

$$p_t = \beta E_t \frac{\sum_{j=0}^{\infty}\beta^j\Phi_j\left(\sum_{i=0}^{\infty}\Phi_i c_{t+1+j-i}\right)^{-\gamma}}{\sum_{j=0}^{\infty}\beta^j\Phi_j\left(\sum_{i=0}^{\infty}\Phi_i c_{t+j-i}\right)^{-\gamma}}(d_{t+1}+p_{t+1}), \tag{5.105}$$

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

$$\begin{aligned} p_t = \beta E_t &\frac{(c_{t+1}+\Phi_1 c_t + \Phi_2 c_{t-1} + \cdots)^{-\gamma}}{(c_t + \Phi_1 c_{t-1} + \Phi_2 c_{t-2} + \cdots)^{-\gamma}} \\ &\frac{+\beta\Phi_1(c_{t+2}+\Phi_1 c_{t+1} + \Phi_2 c_t + \cdots)^{-\gamma} + \cdots}{+\beta\Phi_1(c_{t+1}+\Phi_1 c_t + \Phi_2 c_{t-1} + \cdots)^{-\gamma} + \cdots}(d_{t+1}+p_{t+1}). \end{aligned} \tag{5.106}$$

When there is a positive shock to say q_t, c_t increases by the amount of the shock, say σ_q. Given (5.89)–(5.90), 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 parenthesis in both the numerator and the denominator. First, in the denominator c_t will grow by σ_q. Second, in the numerator $c_{t+1} + \Phi_1 c_t$ goes

up by $(\rho_q + \Phi_1)\sigma_q \lesseqgtr \sigma_q$. Thus, whether the share price p_t increases by more than in the standard CRRA case depends ultimately on whether $\rho_q + \Phi_1 \lesseqgtr 1$. Notice that if $\Phi_j = 0$ for $j > 0$, (5.106) 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 $\rho + \Phi_1 \lesseqgtr 1$. Thus with only durability, we cannot judge how the volatility of p_t would be affected: this will depend upon the sizes of ρ and Φ_1. Finally, we also face indeterminacy under a combination of both durability and habit: if α is large and δ 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 (5.106) that it is identical to the CRRA case.

Exercise 5.8

Given that the pricing equation under Habit/Durability involves an infinite number of lags, truncate the lags to 3 and linearize the pricing equation (5.106) around its steady state.

SELF-CONTROL PREFERENCES

Consider next a household that 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

$$W(s,P) = \max_{s'}\{u(c) + v(c) + \beta EW(s',P')\} - \max_{\tilde{s}'} v(\tilde{c}), \qquad (5.107)$$

where $P = (p,d,q)$, $u(\cdot)$ and $v(\cdot)$ are Von Neuman-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:

$$c = ds + q - p(s' - s) \qquad (5.108)$$

$$\tilde{c} = ds + q - p(\tilde{s}' - s). \qquad (5.109)$$

In (5.107), $v(c) - \max_{\tilde{s}'} v(\tilde{c}) \leq 0$ represents the disutility of self-control given that the agent has chosen c. With $v(c)$ specified as strictly increasing,

the solution for $\max_{\tilde{s}'} v(\tilde{c})$ 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

$$W(s,P) = \max_{s'}\{u(c) + v(c) + \beta EW(s',P')\} - v(ds + q + ps) \qquad (5.110)$$

subject to

$$c = ds + q - p(s' - s). \qquad (5.111)$$

The optimality condition reads

$$\left[u'(c) + v'(c)\right]p = \beta EW'(s',P'), \qquad (5.112)$$

and since

$$W'(s,P) = \left[u'(c) + v'(c)\right](d + p) - v'(ds + q + ps)(d + p), \qquad (5.113)$$

the optimality condition becomes

$$\left[u'(c) + v'(c)\right]p = \beta E\left[u'(c') + v'(c') - v'(d's' + q' + p's')\right](d' + p'). \qquad (5.114)$$

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

$$p = \beta E\left\{(d' + p')\left[\frac{u'(d' + q') + v'(d' + q') - v'(d' + q' + p')}{u'(d + q) + v'(d + q)}\right]\right\}. \qquad (5.115)$$

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:

$$u(c) = \frac{c^{1-\gamma}}{1-\gamma} \qquad (5.116)$$

$$v(c) = \lambda\frac{c^{\phi}}{\phi}, \qquad (5.117)$$

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

$$p = \beta E\left\{[d' + p']\left[\frac{(d'+q')^{-\gamma} + \lambda(d'+q')^{\phi-1} - \lambda(d'+q'+p')^{\phi-1}}{(d+q)^{-\gamma} + \lambda(d+q)^{\phi-1}}\right]\right\}. \tag{5.118}$$

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 (5.118) as

$$p = \beta E\left\{[d' + p']\right.$$
$$\left.\times\left[\frac{\frac{(d'+q')^{-\gamma}}{(d+q)^{-\gamma}} + \lambda(d+q)^{\gamma}\left[(d'+q')^{\phi-1} - (d'+q'+p')^{\phi-1}\right]}{1 + \lambda(d+q)^{\phi-1+\gamma}}\right]\right\}. \tag{5.119}$$

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:

$$-\gamma(d+q)^{-\gamma-1} + \lambda(\phi-1)[(d+q)^{\phi-2} - (d+q+p)^{\phi-2}] < 0 \tag{5.120}$$

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

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

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

From (5.118), the steady-state price in this case is given by

$$\bar{p} = \beta(1+\bar{p})$$
$$\times\left[\frac{(1+\eta)^{-\gamma} + \lambda(1+\eta)^{\phi-1} - \lambda(1+\eta+\bar{p})^{\phi-1}}{(1+\eta)^{-\gamma} + \lambda(1+\eta)^{\phi-1}}\right]. \tag{5.121}$$

Regarding (5.121), 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 (5.121) yields a unique positive solution for $\overline{p}$ for any admissible parameterization of the model. (In practice, (5.121) 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 λ causes the function of $\overline{p}$ on the right-hand side of (5.121) to shift down and flatten, thus $\overline{p}$ is decreasing in λ. The intuition for this is again straightforward: an increase in λ represents an intensification of the household's temptation to liquidate its asset holdings. This drives down its demand for asset shares, and thus $\overline{p}$. Note the parallel between this effect and that generated by an increase in r, or a decrease in β, which operates analogously in both (5.99) and (5.121).

Exercise 5.9

Solve for $\overline{p}$ in (5.121) using $\beta = 0.96, \gamma = 2, \lambda = 0.01, \eta = 10, \phi = 0.4$. Linearize the asset-pricing equation (5.119) using the steady state values for $(\overline{p}, \overline{d}, \overline{q})$ implied by these parameter values.

Part II

Empirical Methods

When you come to a fork in the road, take it.
—Yogi Berra

Chapter 6

Calibration

Models are to be used, not believed.
—Henri Theil, *Principles of Econometrics*

6.1 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.[1] 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:

[1] 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)

identities (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 chapter 5 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, p. 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 percent 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. [p. 1]

The second development was theoretical: the underlying assumption that the parameters of behavioral equations in such models are fixed was

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. [p. 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. [pp. 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." [p. 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." [p. 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,b). 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. [p. 1]

Such studies personified Frisch's vision of econometrics: "This mutual penetration of quantitative economic theory and statistical observation is the essence of econometrics." [p. 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." [p. 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. [p. 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.'" [p. 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. [p. 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.

6.2 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. [pp. 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." [p. 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." [p. 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." [p. 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." [p. 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." [p. 65] So in close relation with step two, 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. [p. 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." [p. 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. [p. 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 2. Using the notation established in part I of this book, the model solution yields a state-space representation of the form

$$x_{t+1} = F(\mu)x_t + G(\mu)v_{t+1} \tag{6.1}$$

$$X_t = H(\mu)'x_t \tag{6.2}$$

$$E(e_t e_t') = G(\mu)E(v_t v_t')G(\mu)' \equiv Q(\mu). \tag{6.3}$$

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 μ 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 μ. Representing the real-world criteria used to specify μ as $\Omega(\{X_t\}_{t=1}^T)$, the calibration step involves the choice of μ such that

$$\Omega(\{X_t^M\}_{t=1}^T) = \Omega(\{X_t\}_{t=1}^T). \tag{6.4}$$

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." [p. 70] As will be apparent in chapter 7, 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 μ such that (6.4) is satisfied. In this case, given that μ contains as a subset of elements parameters characterizing the nature of the policy under investigation, then the policy change will be reflected by a new specification μ', 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.

6.3 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 λ such that

$$U(\{(1+\lambda)c_t^A\}) = U(\{c_t^B\}). \tag{6.5}$$

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

$$c_t = Ae^{\mu t}e^{-\frac{1}{2}\sigma^2}\varepsilon_t, \tag{6.6}$$

where $\log \varepsilon_t$ is distributed as $N(0, \sigma^2)$. Under this distributional assumption for ε_t,

$$E(\varepsilon_t^m) = e^{\frac{1}{2}\sigma^2 m^2}, \tag{6.7}$$

and thus

$$E(c_t) = Ae^{\mu t}, \tag{6.8}$$

the growth rate of which is μ. Under the second, consumption is deterministic, and at time t is given by $Ae^{\mu t}$, as in (6.8).

Modeling the lifetime utility generated by a given consumption stream $\{c_t\}$ as

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{c_t^{1-\gamma}}{1-\gamma}, \tag{6.9}$$

where β is the consumer's discount factor and γ measures the consumer's degree of relative risk aversion, Lucas' welfare comparison involves the calculation of λ such that

$$E_0 \sum_{t=0}^{\infty} \beta^t \frac{[(1+\lambda)c_t]^{1-\gamma}}{1-\gamma} = \sum_{t=0}^{\infty} \beta^t \frac{(Ae^{\mu t})^{1-\gamma}}{1-\gamma}, \tag{6.10}$$

with the behavior of c_t given by (6.6). Given (6.7),

$$E_0(\varepsilon_t^{1-\gamma}) = e^{\frac{1}{2}(1-\gamma)^2\sigma^2},$$

and thus (6.10) simplifies to

$$(1+\lambda)^{1-\gamma} e^{\frac{1}{2}(1-\gamma)^2\sigma^2} = 1. \tag{6.11}$$

Finally, taking logs and using the approximation $\log(1+\lambda) \approx \lambda$, the comparison reduces to

$$\lambda = \frac{1}{2}\sigma^2\gamma. \tag{6.12}$$

TABLE 6.1
Welfare Costs, CRRA Preferences

γ	λ	\$/person, 2004
0.5	0.00023	\$5.86
1	0.00045	\$11.72
1.5	0.00068	\$17.59
2	0.00090	\$23.45
2.5	0.00113	\$29.31

Thus the calculation of λ boils down to this simple relationship involving two parameters: σ^2 and γ. 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 σ^2 as $(0.032)^2$. Extending these data through 2004 yields an estimate of $(0.030)^2$; the data used for this purpose are contained in the file `weldat.txt`, available at the textbook Web site.

Regarding the specification of γ, Lucas (2003) appeals to an intertemporal optimality condition for consumption growth μ that is a standard feature of theoretical growth models featuring consumer optimization (e.g., see Barro and Sala-i-Martin, 2004):

$$\mu = \frac{1}{\gamma}(r-\rho), \tag{6.13}$$

where r is the after-tax rate of return generated by physical capital and ρ 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, μ estimated as 0.023 in the regression of $\log c_t$ on a constant and trend, and ρ restricted to be positive, an upper bound for γ is approximately 2.2, and a value of 1 is often chosen as a benchmark specification.

Table 6.1 reports values of λ calculated for alternative specifications of γ, based on the estimate $\sigma^2 = (0.030)^2$. It also reports the associated (chain-weighted 2000) dollar value of per capita consumption in 2004.

As these figures indicate, potential welfare gains offered by the complete elimination of cyclical fluctuations in consumption are strikingly low: as high as only \$29.31 per person given $\gamma = 2.5$, and only \$11.71 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.

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 5. 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 9. Here, we adopt Otrok's estimates of the habit/durability parameters in pursuit of calculations of λ akin to those made by Lucas (1987, 2003) for the CRRA case.

Recall that under habit/durability preferences, instantaneous utility is given by

$$u(h_t) = \frac{h_t^{1-\gamma}}{1-\gamma}, \tag{6.14}$$

with

$$h_t = h_t^d - \alpha h_t^h, \tag{6.15}$$

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

$$h_t^d = \sum_{j=0}^{\infty} \delta^j c_{t-j} \tag{6.16}$$

$$\begin{aligned} h_t^h &= (1-\theta)\sum_{j=0}^{\infty} \theta^j h_{t-1-j}^d \\ &= (1-\theta)\sum_{j=0}^{\infty} \theta^j \sum_{i=0}^{\infty} \delta^i c_{t-1-i} \end{aligned} \tag{6.17}$$

where $\delta \in (0,1)$ and $\theta \in (0,1)$. Substituting for h_t^d and h_t^h in (6.15) thus yields

$$\begin{aligned} h_t &= c_t + \sum_{j=1}^{\infty} \left[\delta^j - \alpha(1-\theta)\sum_{i=0}^{j-1} \delta^i \theta^{j-i-1} \right] c_{t-j} \\ &\equiv \sum_{j=0}^{\infty} \Phi_j c_{t-j}, \end{aligned} \tag{6.18}$$

where $\Phi_0 \equiv 1$.

Combining (6.14) and (6.18) with (6.10), Lucas' welfare comparison in this case involves the calculation of λ such that

$$\begin{aligned} E_0 \sum_{t=0}^{\infty} \beta^t \frac{\left[(1+\lambda)\left\{\sum_{j=0}^{\infty} \Phi_j c_{t-j}\right\}\right]^{1-\gamma}}{1-\gamma} \\ = \sum_{t=0}^{\infty} \beta^t \frac{\left(\sum_{j=0}^{\infty} \Phi_j A e^{\mu t - j}\right)^{1-\gamma}}{1-\gamma}, \end{aligned} \tag{6.19}$$

where once again the behavior of c_t is given by (6.6). In this case, a convenient analytical expression for λ is unavailable, due to the complicated nature of the integral needed to calculate the expectation on the left-hand side of (6.19). Hence we proceed in this case by approximating λ numerically.

The algorithm we use for this purpose takes the collection of parameters that appear in (6.19) and (6.6), including λ, as given. The parameters (A, μ, σ^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 Otrok (2001), and a grid of values was specified for λ. Because Otrok's estimates were obtained using postwar quarterly data on per capita consumption, defined as the consumption of nondurables and services, we re-estimated (A, μ, σ^2) using this alternative measure of consumption. The series is contained in `ycih.txt`, available at the textbook Web site. It is slightly more volatile than the annual measure of consumption considered by Lucas: the estimate of σ^2 it yields is $(0.035)^2$.

Given a specification of parameters, the right-hand side of (6.19) 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 Otrok, 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 (6.19). This is accomplished via use of a technique known as numerical integration. Full details on the general use of this technique are provided in chapter 9. 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 (6.6) 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 λ that satisfies (6.19), the left-hand side of this equation was approximated over a sequence of values specified for λ; 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 (6.19) is guaranteed to be less than the right-hand side given $\lambda = 0$; and as λ increases, the left-hand side increases, ultimately reaching the value of the right-hand side. The value of λ we seek in the approximation is the value that generates equality between the two sides.

Results of this exercise are reported in table 6.2 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 Otrok calculated for each parameter: 0.9878, 0.7228, 0.446, 0.1533, 0.1299. Next, sensitivity to the specification of γ is demonstrated by re-specifying γ 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 re-specification of γ. Sensitivity to the specifications of α, δ, and θ is demonstrated analogously. Finally, a modification of the baseline under which $\gamma = 1$ is reported, to facilitate a closer comparison of Lucas' estimate of λ obtained under CRRA preferences given $\gamma = 1$.

Under Otrok's baseline parameterization, λ is estimated as 0.00275. In terms of the annual version of consumption considered by Lucas,

TABLE 6.2
Welfare Costs, Habit/Durability Preferences

Parameterization	λ	*$/Person, 2004*
Baseline	0.000275	$7.15
$\gamma = 0.5363$	0.000225	$5.85
$\gamma = 0.9471$	0.000400	$10.40
$\theta = 0.0178$	0.000375	$9.75
$\theta = 0.3039$	0.000200	$5.20
$\alpha = 0.2618$	0.000275	$7.15
$\alpha = 0.6133$	0.000400	$10.40
$\delta = 0.0223$	0.000425	$11.05
$\delta = 0.3428$	0.000175	$4.55
$\gamma = 1$	0.000550	$14.30

Note: Baseline estimates are $(\beta, \gamma, \alpha, \delta, \theta) = (0.9878, 0.7228, 0.446, 0.1533, 0.1299)$.

this corresponds with a consumption cost of $7.15/person in 2004. For the modification of the baseline under which $\gamma = 1$, the cost rises to $14.30/person, which slightly exceeds the cost of $11.72/person calculated using CRRA preferences given $\gamma = 1$. The estimated value of λ is most sensitive to changes in δ, 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: $4.55/person given $\delta = 0.3428$; $11.05/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 Otrok: "... it is a result that casts doubt that empirically plausible modifications to preferences alone could lead to large costs of consumption volatility." [p. 88]

Exercise 6.1

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 `weldat.txt`, regress the log of consumption on a constant and trend. Use the resulting parameter estimates to specify (A, μ, σ^2) in (6.6), 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 (6.6) to obtain $\{c_t\}_{t=0}^{T}$.

3. Using the drawing $\{\varepsilon_t\}_{t=0}^T$, calculate $\sum_{t=0}^T \beta^t \frac{[(1+\lambda)c_t]^{1-\gamma}}{1-\gamma}$ 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 $\lambda = 0,\ 0.000025, \ldots, 0.0001$.
4. Repeat steps 2 and 3 1,000 times, and record the average values of $\sum_{t=0}^T \beta^t \frac{[(1+\lambda)c_t]^{1-\gamma}}{1-\gamma}$ you obtain as approximations of $E_0 \sum_{t=0}^T \beta^t \frac{[(1+\lambda)c_t]^{1-\gamma}}{1-\gamma}$.
5. Calculate the right-hand side of (6.10) using the estimates of A and μ obtained in step 1. Do so for each value of γ considered in step 3.
6. For each value of γ, find the value of λ that most closely satisfies (6.10). Compare the values you obtain with those reported in table 6.1. Are Lucas' original calculations robust to departures from the log-Normality assumption adopted for $\{\varepsilon_t\}$?

Exercise 6.2

Using the baseline estimates of the habit/durability specification reported in table 6.2, evaluate the robustness of the estimates of λ reported in table 6.2 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 the exercise 6.1. Once again, are the results reported in table 6.2 robust to departures from the log-Normality assumption adopted for $\{\varepsilon_t\}$?

6.4 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 5, section 5.1.1, 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 3 and 4 (the data are contained in `ycih.txt`, available at the textbook Web site). Reverting to the notation used in the specification of the model, the parameters to be calibrated are as follows: capital's share of output α; the subjective discount factor $\beta = \frac{1}{1+\varrho}$, where ϱ is the subjective discount rate; the degree of relative risk aversion ϕ; consumption's share (relative to leisure) of instantaneous utility φ; the depreciation rate of physical capital δ; the AR parameter specified for the productivity shock ρ; and the standard deviation of innovations to the productivity shock σ.

Standard specifications of β result from the association of the subjective discount rate ϱ 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 α and δ. Recall that steady state values of $\frac{\bar{y}}{\bar{n}}$ and $\frac{\bar{i}}{\bar{n}}$ are given by

$$\frac{\bar{y}}{\bar{n}} = \eta,$$
$$\frac{\bar{i}}{\bar{n}} = \delta\theta,$$

where

$$\theta = \left(\frac{\alpha}{\varrho + \delta}\right)^{\frac{1}{1-\alpha}},$$
$$\eta = \theta^{\alpha}.$$

Combining these expressions yields the relationship

$$\alpha = \left(\frac{\delta + \varrho}{\delta}\right)\frac{\bar{i}}{\bar{y}}. \tag{6.20}$$

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 α and δ.

Before illustrating this relationship explicitly, we use a similar step to identify a plausible range of values for φ. Here we use the steady state expression for $\bar{n}$, the fraction of discretionary time spent on job-market activities:

$$\bar{n} = \frac{1}{1 + \left(\frac{1}{1-\alpha}\right)\left(\frac{1-\varphi}{\varphi}\right)[1 - \delta\theta^{1-\alpha}]}. \tag{6.21}$$

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 (6.21), and exploiting the fact that $\frac{\bar{i}}{\bar{y}} = [1 - \delta\theta^{1-\alpha}]$ (the sample average of which is 0.825), we obtain the following relationship between α and φ:

$$\varphi = \frac{1}{1 + \frac{2(1-\alpha)}{\bar{i}/\bar{y}}}. \tag{6.22}$$

In sum, using sample averages to pin down $\frac{\bar{i}}{\bar{y}}$, $\frac{\bar{c}}{\bar{y}}$, and $\bar{n}$, we obtain the relationships between α, δ, and φ characterized by (6.20) and (6.22).

TABLE 6.3
Trade-Offs Between δ, α, and φ

δ	α	φ
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 $\frac{\bar{i}}{\bar{y}} = 0.175$, $\frac{\bar{c}}{\bar{y}} = 0.825$, and $\bar{n} = 1/3$.

Table 6.3 presents combinations of these parameters for a grid of values specified for δ 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 $\frac{\bar{i}}{\bar{y}}$, $\frac{\bar{c}}{\bar{y}}$, and $\bar{n}$. Regarding the specification of α, 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 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 9.

The final parameters to be established are those associated with the behavior of the productivity shock z_t: ρ and σ. 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

$$k_{t+1} = i_t + (1 - \delta)k_t \tag{6.23}$$

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 (6.23) by y_t; use beginning-of-sample averages to measure the resulting ratios $\frac{\bar{x}}{\bar{y}}$, $\bar{x} = \bar{i}, \bar{k}$; solve for

$$\frac{\bar{k}}{\bar{y}} = \frac{1}{\delta}\frac{\bar{i}}{\bar{y}}; \tag{6.24}$$

and finally, multiply $\frac{1}{\delta}\frac{\bar{i}}{\bar{y}}$ by y_0. The results reported here are based on the specification of $\frac{\bar{i}}{\bar{y}}$ 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 (1957) as the unexplained component of $\log y_t$ given the input of $\log k_t$ and $\log n_t$ in the production function:

$$\log z_t = \log y_t - \alpha \log k_t - (1 - \alpha)\log n_t. \tag{6.25}$$

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 ρ and σ 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 4.1. Moments calculated using both the model and data are reported in table 6.4.

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, it performs poorly in characterizing the relative volatility of hours, which are roughly equally as volatile as output in the data, but only 1/3 as volatile in the model. Figure 6.1 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,

TABLE 6.4
Moment comparison

			H-P filtered Data		
j	σ_j	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\varphi_{j,y}(1)$
y	0.0177	1.00	0.86	1.00	0.86
c	0.0081	0.46	0.83	0.82	0.75
i	0.0748	4.23	0.79	0.95	0.80
n	0.0185	1.05	0.90	0.83	0.62
			RBC Model		
	σ_j	$\frac{\sigma_j}{\sigma_y}$	$\varphi(1)$	$\varphi_{j,y}(0)$	$\varphi_{j,y}(1)$
y	0.0207	1.00	0.87	1.00	0.87
c	0.0101	0.48	0.94	0.96	0.93
i	0.0752	3.63	0.81	0.98	0.79
n	0.0076	0.366	0.78	0.97	0.76

Notes: $\varphi(1)$ denotes first-order serial correlation; $\varphi_{j,y}(l)$ denotes l^{th} order correlation between variables j and y. Model moments based on the parameterization

$$\mu = [\alpha\ \beta\ \phi\ \varphi\ \delta\ \rho\ \sigma]' = [0.24\ 0.99\ 1.5\ 0.35\ 0.025\ 0.78\ 0.0067]'.$$

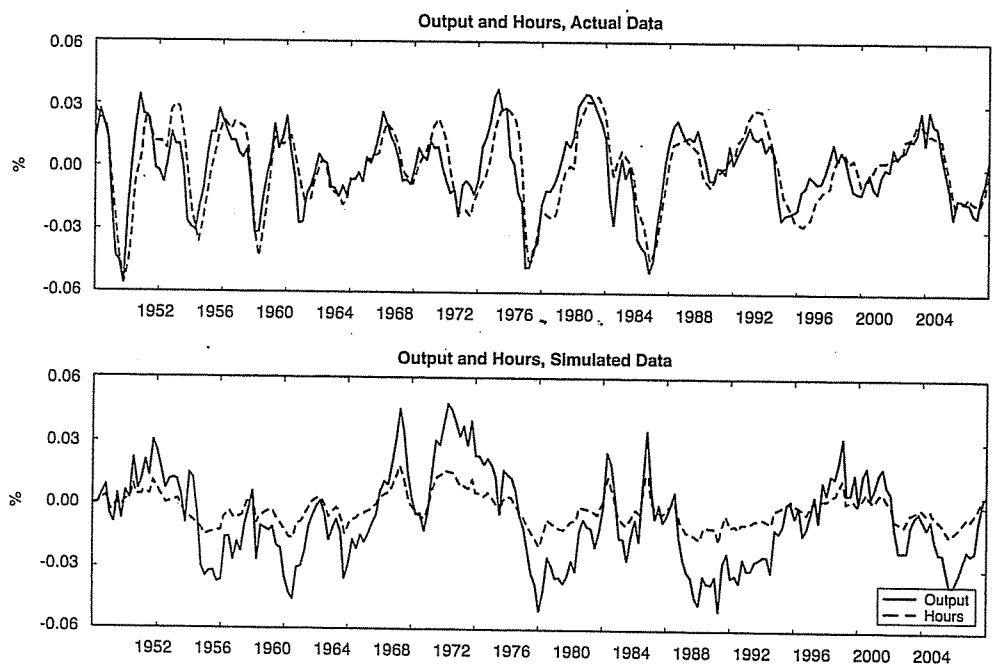

Figure 6.1 Comparisons of output and hours.

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, p. 70).

Exercise 6.3

Reconstruct table 6.4 by using the alternative combinations of values specified for δ, α, and φ listed in table 6.3. Also, consider 0.5 and 2.5 as alternative specifications of ϕ. Are the results of table 6.4 materially altered by any of the alternative choices you considered?

Exercise 6.4

Reconduct the calibration exercise of this section using the extension of the RBC model outlined in chapter 5, section 5.1.1 that explicitly incorporates long-term growth. Pay careful attention to the fact that equations (6.20) and (6.22), used to restrict the parameterizations of δ, α, and φ, will be altered given this extension. So too will be (6.24), which was used to specify k_0. Once again, are the results of table 6.4 materially altered given this extension of the model?

6.5 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 5, section 5.3.2, 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 non-durables and services. The time span of their data is 1889–1978; here we demonstrate a brief replication of their study using data updated through 2004. 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 Web site.

As characterized in chapter 5, section 5.3.2, 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:

$$\beta E_t \left\{ \frac{u'(c_{t+1})}{u'(c_t)} r^f_{t+1} - 1 \right\} = 0 \tag{6.26}$$

$$E_t \left\{ \frac{u'(c_{t+1})}{u'(c_t)} \left[r^e_{t+1} - r^f_{t+1} \right] \right\} = 0, \tag{6.27}$$

where r^f_{t+1} and r^e_{t+1} denote the gross returns associated with a risk-free and risky asset. The difference $r^e_{t+1} - r^f_{t+1}$ is referred to as the equity premium. Sample averages (standard deviations) of r^f_{t+1} and r^e_{t+1} are 0.8% (5.67%) and 6.98% (16.54%) in the data ending in 1978, and 1.11% (5.61%) and 7.48% (16.04%) in the data ending through 2004. 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 γ representing the risk-aversion parameter, the equations are given by

$$\beta E_t \left\{ \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} r^f_{t+1} - 1 \right\} = 0 \tag{6.28}$$

$$E_t \left\{ \left(\frac{c_{t+1}}{c_t} \right)^{-\gamma} \left[r^e_{t+1} - r^f_{t+1} \right] \right\} = 0. \tag{6.29}$$

Here, we use sample averages to estimate the expectations represented in (6.28) and (6.29), and seek to determine whether it is possible to specify γ in order to jointly account for each equation, given $\beta = 0.99$. The results of this exercise are presented in table 6.5.

Note that for a specification of γ 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, γ plays 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 1.79% in the data ending in 2004), 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 γ. Of course, the specification of a smaller value for β requires an even lower value for γ; thus $\beta = 0.99$ represents nearly a best-case scenario along this dimension.

In turn, a large specification of γ is required to account for the second equation. In fact, the sample average first becomes negative given $\gamma = 20$.

TABLE 6.5
The Equity Premium and Risk-Free Rate Puzzles

γ	$\beta\left(\frac{c_{t+1}}{c_t}^{-\gamma} r^f_{t+1}\right) - 1$	$\left(\frac{c_{t+1}}{c_t}\right)^{-\gamma} \left[r^e_{t+1} - r^f_{t+1} \right]$
0	0.001	0.064
0.5	−0.007	0.062
1	−0.015	0.060
1.5	−0.023	0.059
2	−0.030	0.057
2.5	−0.037	0.056
3	−0.044	0.054

Notes: $\overline{x}$ denotes the sample mean of x. Results obtained using $\beta = 0.99$.

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 6.5

Replicate the calculations presented in table 6.5 using the baseline parameterization of the habit/durability specification presented in section 6.3. Does this modification represent progress in resolving the puzzle? Ponder the intuition behind your findings.

6.6 Critiques and Extensions

6.6.1 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 6.2) 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.* [p. iii]

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, p. 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. [p. 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. [p. 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. [p. 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 epigram 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 9.

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." [p. 91] Their reasoning fits nicely with the description provided in section 6.1 of the means by which the specification of μ is achieved via use of (6.4); and the means by which judgements 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. [p. 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." [p. 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. [p. 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. [p. 109]

It is precisely in the spirit of these sentiments that we present the additional material contained in Part II of this book.

6.6.2 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 Υ_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 Υ_t?

To quantify this question, recall from chapter 4 that the s^{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^{th}-order autocovariance of $U_t = \Upsilon_t - X_t$ is given by

$$\Gamma_U(s) = \Gamma_\Upsilon(s) + \Gamma_X(s) - \Gamma_{X\Upsilon}(s) - \Gamma_{\Upsilon X}(s), \tag{6.30}$$

where

$$\Gamma_{X\Upsilon}(s) = E(X_t \Upsilon_{t-s}'),$$

and

$$\Gamma_{X\Upsilon}(s) = \Gamma_{\Upsilon X}(-s)'.$$

With the autocovariance generating function (ACGF) of z_t given by

$$A_z(e^{-i\omega}) = \sum_{s=-\infty}^{\infty} \Gamma_z(s) e^{-i\omega s}, \tag{6.31}$$

where i is complex and $\omega \in [0,\ 2\pi]$ represents a particular frequency, the ACGF of U_t implied by (6.30) is given by

$$\begin{aligned} A_U(e^{-i\omega}) = {} & A_\Upsilon(e^{-i\omega}) + A_X(e^{-i\omega}) \\ & - A_{X\Upsilon}(e^{-i\omega}) - A_{X\Upsilon}(e^{i\omega})', \end{aligned} \tag{6.32}$$

where

$$A_{XY}(e^{i\omega})' = A_{YX}(e^{-i\omega}).$$

As discussed in chapter 4, 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

$$\Sigma_U = \Sigma_Y + \Sigma_X - \Sigma_{XY} - \Sigma_{YX}. \tag{6.33}$$

Because there is no unique measure of the size of Σ_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

$$tr(G\Sigma_U G') = tr(G'G\Sigma_U).$$

Given W, Watson shows that

$$\Sigma_{XY} = C_X V U' C_Y' \tag{6.34}$$

is the unique specification of Σ_{XY} that minimizes $tr(W\Sigma_U)$. In (6.34), C_X and C_Y are arbitrary $m \times m$ matrix square roots of Σ_X and Σ_Y (e.g., $\Sigma_X = C_X C_X'$, 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

$$C_Y' W C_X = USV', \tag{6.35}$$

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.[2]

[2] In GAUSS, Cholesky decompositions can be obtained using the command `chol`, and singular value decompositions can be obtained using the command `svd1`.

For the general case in which X_t and Y_t are serially correlated, Watson's minimization objective translates directly from Σ_{XY} to $A_{XY}(e^{-i\omega})$. Recall from chapter 4 the relationship between ACGFs and spectra; in the multivariate case of interest here, the relationship is given by

$$s_z(\omega) = \left(\frac{1}{2\pi}\right) \sum_{s=-\infty}^{\infty} \Gamma_z(s) e^{-i\omega s}, \qquad \omega \in [-\pi, \pi]. \tag{6.36}$$

With $C_X(\omega)$ now denoting the matrix square root of $s_X(\omega)$ calculated for a given specification of ω, etc., the specification of $A_{XY}(e^{-i\omega})$ that minimizes $tr(W\Sigma_U)$ is given by

$$A_{XY}(e^{-i\omega}) = C_X(\omega) V(\omega) U'(\omega) C_Y'(\omega), \tag{6.37}$$

where $U(\omega)$ and $V(\omega)$ are as indicated in (6.35).

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):

$$r_j(\omega) = \frac{A_U(e^{-i\omega})_{jj}}{A_Y(e^{-i\omega})_{jj}}, \tag{6.38}$$

where $A_U(e^{-i\omega})_{jj}$ denotes the j^{th} diagonal element of $A_U(e^{-i\omega})$.

As an illustration, figure 6.2 closely follows Watson by demonstrating the application of his procedure to the RBC model presented in section 6.4, parameterized as indicated in table 6.4. 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 and 1/6 cycles per quarter). The data were also HP-filtered, and W was specified as the 4×4 identity matrix.[3]

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 0.18, 0.79, 0.15, and 0.66; thus the model's characterization of consumption and hours over this range is seen to be particularly poor.

[3] Three GAUSS procedures were used to produce this figure: `modspec.prc`, `dataspec.prc`, and `errspec.prc`. All are available at the textbook Web site.

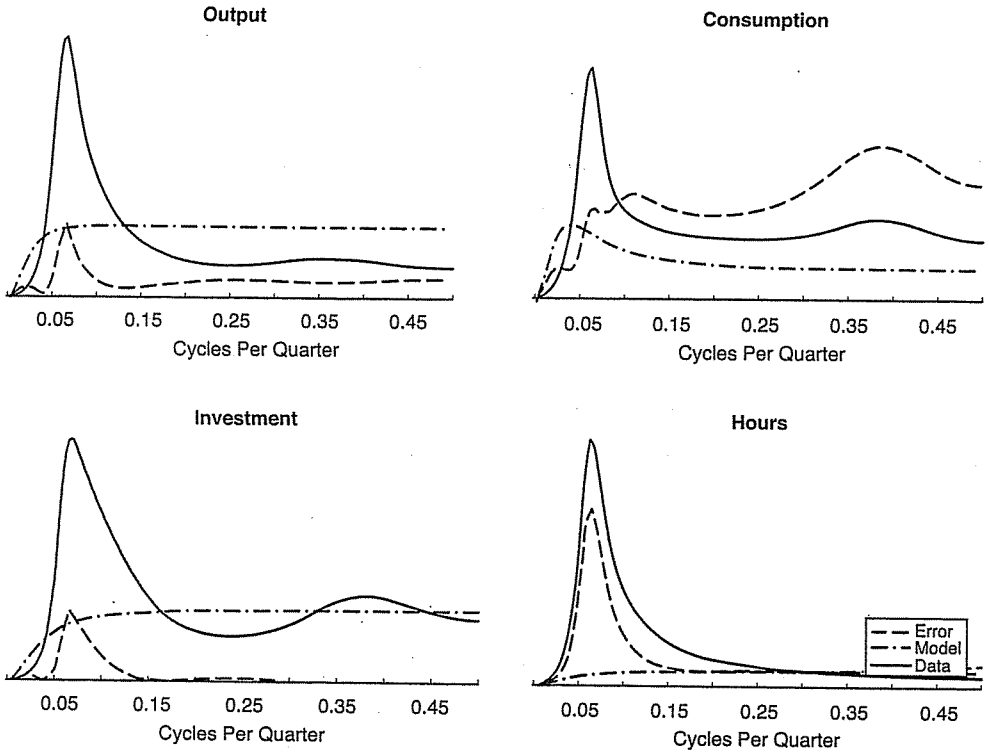

Figure 6.2 Decomposition of spectra.

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 employment has an external affect on sectoral output; and the specification of habit formation in the enjoyment of leisure activities. So too did Otrok (2001), for the model (described in section 6.3) he used to analyze the welfare cost of business cycles. Thus the results of figure 6.2 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 μ 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 6.4 used to evaluate the RBC model presented in section 6.4 represents a common choice of empirical targets. Reverting to the notation used in section 6.2, hereafter we represent the selected targets as Ω.

The difference between these studies lies in the empirical distributions over Ω they use. DeJong et al. worked with posterior distributions over Ω 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 9, 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 Ω 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 4, section 4.2, as a means of approximating standard errors numerically.

Using the notation of chapter 4, let the VAR specified for X_t be given by

$$X_t = \Upsilon_1 X_{t-1} + \Upsilon_2 X_{t-2} + \cdots + \Upsilon_p X_{t-p} + \varepsilon_t, \qquad E(\varepsilon_t \varepsilon_t') = \Sigma. \tag{6.39}$$

The algorithm takes as inputs OLS estimates $\widehat{\Upsilon}_j, j = 1, \ldots, p$, $\widehat{\Sigma}$, estimated residuals $\{\widehat{\varepsilon}_t\}$, and the first p observations of X_t, which serve as starting values. Given these inputs, a simulated drawing $\widetilde{X}_{p+1}$ can be constructed by obtaining an artificial drawing $\widetilde{\varepsilon}_p$ from a given distribution, and evaluating the right-hand side of (6.39) using $\{X_1, X_2, \ldots, X_{p-1}, X_p\}$ and $\widetilde{\varepsilon}_p$. Next, $\widetilde{X}_{p+2}$ is constructed by obtaining a second drawing $\widetilde{\varepsilon}_{p+1}$, and evaluating the right-hand side of (6.39) using $\{X_2, X_3, \ldots, X_p, \widetilde{X}_{p+1}\}$ and $\widetilde{\varepsilon}_p$. Performing T replications of this process yields an artificial drawing $\{\widetilde{X}_t\}$. (The influence of the initial observations $\{X_1, X_2, \ldots, X_{p-1}, X_p\}$ can be reduced by obtaining $T + T'$ drawings of $\widetilde{X}_t$, and discarding the first T' drawings in constructing $\{\widetilde{X}_t\}$.) For each of J replications of this process, a collection of J estimates of Ω is calculated; the resulting distribution of Ω approximates the sampling distribution we seek.

There are two common general methods for obtaining artificial drawings $\widetilde{\varepsilon}_t$. One method involves a distributional assumption for $\{\varepsilon_t\}$, parameterized by $\widehat{\Sigma}$. For example, let $\widehat{S}$ denote the Cholesky decomposition of $\widehat{\Sigma}$. Then under the assumption of Normality for $\{\varepsilon_t\}$, drawings $\widetilde{\varepsilon}_t$ may be obtained using $\widetilde{\varepsilon}_t = \widehat{S}\widetilde{u}$, where $\widetilde{u}$ represents an $m \times 1$ vector of independent $N(0,1)$ random variables. Alternatively, $\widetilde{\varepsilon}_t$ may be obtained as a drawing (with replacement) from the collection of residuals $\{\widehat{\varepsilon}_t\}$.

Let the distribution obtained for the i^{th} element of Ω be given by $S(\Omega_i)$, and let $[a, b]_i$ denote the range of values for Ω_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 Ω_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 $[a, b]_i$:

$$f_i = \int_{a_i}^{b_i} \pi(\Omega_i) d\Omega_i. \tag{6.40}$$

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 Ω_i.

Exercise 6.6

Using the RBC model presented in section 6.4, calculate f_i for the collection of moments analyzed in table 6.4. Use a 6-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 $\widetilde{\varepsilon}_t$ are obtained as drawings (with replacement) from the collection of residuals $\{\widehat{\varepsilon}_t\}$. Finally, use the following prior distribution specified over the elements of μ : $\beta \sim U[0.988, 0.9925]$; $\delta \sim U[0.01, 0.04]$; $\rho \sim U[0.75, 0.81]$, and U denoting the uniform distribution. (Uniform drawings of, e.g., β over $[\underline{\beta}, \overline{\beta}]$ may be obtained using the transformation $\widetilde{\beta} = \underline{\beta} + (\overline{\beta} - \underline{\beta})\widetilde{u}$, where $\widetilde{u}$ is a drawing from a $U[0, 1]$ distribution.) Given a drawing of δ and $\varrho = 1/\beta - 1$, calculate the corresponding value of α using (6.20), and the corresponding value of φ using (6.22). Throughout, hold σ fixed at 0.0067.

Chapter 7

Matching Moments

> Know the right moment.
> —The Seven Sages, from Diogenes Luertius,
> *Lives of Eminent Philosophers*

7.1 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 4, 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 μ that best enables the underlying structural model to match a collection of preselected moments. The parameterization $\widehat{\mu}$ that accomplishes this objective is interpreted as an estimate of the actual value of these parameters. Because the estimate $\widehat{\mu}$ 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