HAND- BOOKS

IN ECONOMICS

John B. Taylor
Harald Uhlig

Handbook of Macroeconomics

VOLUME 2A

CHAPTER 9

Solution and Estimation Methods for DSGE Models

J. Fernández-Villaverde*, J.F. Rubio-Ramírez[†,‡,§,¶], F. Schorfheide*
*University of Pennsylvania, Philadelphia, PA, United States
†Emory University, Atlanta, GA, United States
‡Federal Reserve Bank of Atlanta, Atlanta, GA, United States
§BBVA Research, Madrid, Madrid, Spain
¶Fulcrum Asset Management, London, England, United Kingdom

Contents

Abstract

This chapter provides an overview of solution and estimation techniques for dynamic stochastic general equilibrium models. We cover the foundations of numerical approximation techniques as well as statistical inference and survey the latest developments in the field.

Keywords

Approximation error analysis, Bayesian inference, DSGE model, Frequentist inference, GMM estimation, Impulse response function matching, Likelihood-based inference, Metropolis-Hastings algorithm, Minimum distance estimation, Particle filter, Perturbation methods, Projection methods, Sequential Monte Carlo.

JEL Classification Codes

C11, C13, C32, C52, C61, C63, E32, E52

1. INTRODUCTION

The goal of this chapter is to provide an illustrative overview of the state-of-the-art solution and estimation methods for dynamic stochastic general equilibrium (DSGE) models. DSGE models use modern macroeconomic theory to explain and predict comovements of aggregate time series over the business cycle. The term *DSGE model* encompasses a broad class of macroeconomic models that spans the standard neoclassical growth model discussed in King et al. (1988) as well as New Keynesian monetary models with numerous real and nominal frictions along the lines of Christiano et al. (2005) and Smets and Wouters (2003). A common feature of these models is that decision rules of economic agents are derived from assumptions about preferences, technologies, information, and the prevailing fiscal and monetary policy regime by solving intertemporal optimization problems. As a consequence, the DSGE model paradigm delivers empirical models with a strong degree of theoretical coherence that are attractive as a laboratory for policy experiments. Modern DSGE models are flexible enough to accurately track and forecast macroeconomic time series fairly well. They have become one of the workhorses of monetary policy analysis in central banks.

The combination of solution and estimation methods in a single chapter reflects our view of the central role of the tight integration of theory and data in macroeconomics. Numerical solution methods allow us to handle the rich DSGE models that are needed for business cycle analysis, policy analysis, and forecasting. Estimation methods enable us to take these models to the data in a rigorous manner. DSGE model solution and estimation techniques are the two pillars that form the basis for understanding the behavior of aggregate variables such as GDP, employment, inflation, and interest rates, using the tools of modern macroeconomics.

Unfortunately for PhD students and fortunately for those who have worked with DSGE models for a long time, the barriers to entry into the DSGE literature are quite high. The solution of DSGE models demands familiarity with numerical approximation techniques and the estimation of the models is nonstandard for a variety of reasons, including a state-space representation that requires the use of sophisticated filtering techniques to evaluate the likelihood function, a likelihood function that depends in a

complicated way on the underlying model parameters, and potential model misspecification that renders traditional econometric techniques based on the "axiom of correct specification" inappropriate. The goal of this chapter is to lower the barriers to entry into this field by providing an overview of what have become the "standard" methods of solving and estimating DSGE models in the past decade and by surveying the most recent technical developments. The chapter focuses on methods more than substantive applications, though we provide detailed numerical illustrations as well as references to applied research. The material is grouped into two parts. Part I: Solving DSGE Models (Sections 2–7) is devoted to solution techniques, which are divided into perturbation and projection techniques. Part II: Estimating DSGE Models (Sections 8–12) focuses on estimation. We cover both Bayesian and frequentist estimation and inference techniques.

PART I. SOLVING DSGE MODELS
2. SOLUTION METHODS FOR DSGE MODELS

DSGE models do not admit, except in a few cases, a closed-form solution to their equilibrium dynamics that we can derive with "paper and pencil." Instead, we have to resort to numerical methods and a computer to find an approximated solution.

However, numerical analysis and computer programming are not a part of the standard curriculum for economists at either the undergraduate or the graduate level. This educational gap has created three problems. The first problem is that many macroeconomists have been reluctant to accept the limits imposed by analytic results. The cavalier assumptions that are sometimes taken to allow for closed-form solutions may confuse more than clarify. While there is an important role for analytic results for building intuition, for understanding economic mechanisms, and for testing numerical approximations, many of the questions that DSGE models are designed to address require a quantitative answer that only numerical methods can provide. Think, for example, about the optimal response of monetary policy to a negative supply shock. Suggesting that the monetary authority should lower the nominal interest rate to smooth output is not enough for real-world advice. We need to gauge the magnitude and the duration of such an interest rate reduction. Similarly, showing that an increase in government spending raises output does not provide enough information to design an effective countercyclical fiscal package.

The second problem is that the lack of familiarity with numerical analysis has led to the slow diffusion of best practices in solution methods and little interest in issues such as the assessment of numerical errors. Unfortunately, the consequences of poor approximations can be severe. Kim and Kim (2003) document how inaccurate solutions may cause spurious welfare reversals. Similarly, the identification of parameter values may depend on the approximated solution. For instance, van Binsbergen et al. (2012) show that a

DSGE model with recursive preferences needs to be solved with higher-order approximations for all parameters of interest to be identified. Although much progress in the quality of computational work has been made in the last few years, there is still room for improvement. This is particularly important as essential nonlinearities—such as those triggered by nonstandard utility functions, time-varying volatility, or occasionally binding constraints—are becoming central to much research on the frontier of macroeconomics. Nonstandard utility functions such as the very popular Epstein–Zin preferences (Epstein and Zin, 1989) are employed in DSGE models by Tallarini (2000), Piazzesi and Schneider (2006), Rudebusch and Swanson (2011, 2012), van Binsbergen et al. (2012), and Fernández-Villaverde et al. (2014), among many others. DSGE models with time-varying volatility include Fernández-Villaverde and Rubio-Ramírez (2007), Justiniano and Primiceri (2008), Bloom (2009), Fernández-Villaverde et al. (2011, 2015b), also among many others. Occasionally binding constraints can be caused by many different mechanisms. Two popular ones are the zero lower bound (ZLB) of nominal interest rates (Eggertsson and Woodford, 2003; Christiano et al., 2011; Fernández-Villaverde et al., 2015a; Aruoba and Schorfheide, 2015; and Gust et al., 2016) and financial frictions (such as in Bernanke and Gertler, 1989; Carlstrom and Fuerst, 1997; Bernanke et al., 1999; Fernández-Villaverde, 2010; Christiano et al., 2014; and dozens of others). Inherent nonlinearities force macroeconomists to move beyond traditional linearization methods.

The third problem is that, even within the set of state-of-the-art solution methods, researchers have sometimes been unsure about the trade-offs (for example, regarding speed vs accuracy) involved in choosing among different algorithms.

Part I of the chapter covers some basic ideas about solution methods for DSGE models, discusses the trade-offs created by alternative algorithms, and introduces basic concepts related to the assessment of the accuracy of the solution. Throughout the chapter, we will include remarks with additional material for those readers willing to dig deeper into technical details.

Because of space considerations, there are important topics we cannot cover in what is already a lengthy chapter. First, we will not deal with value and policy function iteration. Rust (1996) and Cai and Judd (2014) review numerical dynamic programming in detail. Second, we will not discuss models with heterogeneous agents, a task already well accomplished by Algan et al. (2014) and Nishiyama and Smetters (2014) (the former centering on models in the Krusell and Smith (1998) tradition and the latter focusing on overlapping generations models). Although heterogeneous agent models are, indeed, DSGE models, they are often treated separately for simplicity. For the purpose of this chapter, a careful presentation of issues raised by heterogeneity will consume many pages. Suffice it to say, nevertheless, that most of the ideas in our chapter can also be applied, with suitable modifications, to models with heterogeneous agents. Third, we will not spend much time explaining the peculiarities of Markov-switching

regime models and models with stochastic volatility. Finally, we will not explore how the massively parallel programming allowed by graphic processor units (GPUs) is a game-changer that opens the door to the solution of a much richer class of models. See, for example, Aldrich et al. (2011) and Aldrich (2014). Finally, for general background, the reader may want to consult a good numerical analysis book for economists. Judd (1998) is still the classic reference.

Two additional topics—a survey of the evolution of solution methods over time and the contrast between the solution of models written in discrete and continuous time—are briefly addressed in the next two remarks.

Remark 1 (*The evolution of solution methods*) We will skip a detailed historical survey of methods employed for the solution of DSGE models (or more precisely, for their ancestors during the first two decades of the rational expectations revolution). Instead, we will just mention four of the most influential approaches.

Fair and Taylor (1983) presented an extended path algorithm. The idea was to solve, for a terminal date sufficiently far into the future, the path of endogenous variables using a shooting algorithm. Recently, Maliar et al. (2015) have proposed a promising derivation of this idea, the extended function path (EFP), to analyze applications that do not admit stationary Markov equilibria.

Kydland and Prescott (1982) exploited the fact that the economy they were analyzing was Pareto optimal to solve the social planner's problem instead of the recursive equilibrium of their model. To do so, they substituted a linear quadratic approximation to the original social planner's problem and exploited the fast solution algorithms existing for that class of optimization problems. We will discuss this approach and its relation with perturbation in Remark 13.

King, Plosser, and Rebelo (in the widely disseminated technical appendix, not published until 2002), building on Blanchard and Kahn (1980)'s approach, linearized the equilibrium conditions of the model (optimality conditions, market clearing conditions, etc.), and solved the resulting system of stochastic linear difference equations. We will revisit linearization below by interpreting it as a first-order perturbation.

Christiano (1990) applied value function iteration to the social planner's problem of a stochastic neoclassical growth model.

Remark 2 (*Discrete vs continuous time*) In this chapter, we will deal with DSGE models expressed in discrete time. We will only make passing references to models in continuous time. We do so because most of the DSGE literature is in discrete time. This, however, should not be a reason to forget about the recent advances in the computation of DSGE models in continuous time (see Parra–Alvarez, 2015) or to underestimate the analytic power of continuous time. Researchers should be open to both specifications and opt, in each particular application, for the time structure that maximizes their ability to analyze the model and take it to the data successfully.

3. A GENERAL FRAMEWORK

A large number of solution methods have been proposed to solve DSGE models. It is, therefore, useful to have a general notation to express the model and its solution. This general notation will make the similarities and differences among the solution methods clear and will help us to link the different approaches with mathematics, in particular with the well–developed study of functional equations.

Indeed, we can cast numerous problems in economics in the form of a functional equation.[a] Let us define a functional equation more precisely. Let J^1 and J^2 be two functional spaces, $\Omega \subseteq \mathbb{R}^n$ (where Ω is the state space), and $\mathcal{H}: J^1 \to J^2$ be an operator between these two spaces. A *functional equation problem* is to find a function $d \subseteq J^1: \Omega \to \mathbb{R}^m$ such that:

$$\mathcal{H}(d) = \mathbf{0}. \tag{1}$$

From Eq. (1), we can see that regular equations are nothing but particular examples of functional equations. Also, note that $\mathbf{0}$ is the space zero, different in general than the zero in the real numbers.

Examples of problems in macroeconomics that can be framed as a functional equation include value functions, Euler equations, and conditional expectations. To make this connection explicit, we introduce first the stochastic neoclassical growth model, the ancestor of all modern DSGE models. Second, we show how we can derive a functional equation problem that solves for the equilibrium dynamics of the model in terms of either a value function, an Euler equation, or a conditional expectation. After this example, the reader will be able to extend the steps in our derivations to her application.

3.1 The Stochastic Neoclassical Growth Model

We have an economy with a representative household that picks a sequence of consumption c_t and capital k_t to solve

$$\max_{\{c_t, k_{t+1}\}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t) \tag{2}$$

where $\mathbb{E}_t$ is the conditional expectation operator evaluated at period t, β is the discount factor, and u is the period utility function. For simplicity, we have eliminated the labor supply decision.

---

[a] Much of we have to say in this chapter is not, by any means, limited to macroeconomics. Similar problems appear in fields such as finance, industrial organization, international finance, etc.

The resource constraint of the economy is given by

$$c_t + k_{t+1} = e^{z_t} k_t^{\alpha} + (1-\delta) k_t \tag{3}$$

where δ is the depreciation rate and z_t is an AR(1) productivity process:

$$z_t = \rho z_{t-1} + \sigma \varepsilon_t, \varepsilon_t \sim N(0,1) \text{ and } |\rho| < 1. \tag{4}$$

Since both fundamental welfare theorems hold in this economy, we can jump between the social planner's problem and the competitive equilibrium according to which approach is more convenient in each moment. In general, this would not be possible, and some care is required to stay on either the equilibrium problem or the social planner's problem according to the goals of the exercise.

3.2 A Value Function

Under standard technical conditions (Stokey et al., 1989), we can transform the sequential problem defined by Eqs. (2)–(4) into a recursive problem in terms of a value function $V(k_t, z_t)$ for the social planner that depends on the two state variables of the economy, capital, k_t, and productivity, z_t. More concretely, $V(k_t, z_t)$ is defined by the Bellman operator:

$$V(k_t, z_t) = \max_{k_{t+1}} \left[u \left(e^{z_t} k_t^{\alpha} + (1-\delta) k_t - k_{t+1} \right) + \beta \mathbb{E}_t V(k_{t+1}, z_{t+1}) \right] \tag{5}$$

where we have used the resource constraint (3) to substitute for c_t in the utility function and the expectation in (5) is taken with respect to (4). This value function has an associated decision rule $g : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$:

$$k_{t+1} = g(k_t, z_t)$$

that maps the states k_t and z_t into optimal choices of k_{t+1} (and, therefore, optimal choices of $c_t = e^{z_t} k_t^{\alpha} + (1-\delta) k_t - g(k_t, z_t)$).

Expressing the model as a value function problem is convenient for several reasons. First, we have many results about the properties of value functions and the decision rules associated with them (for example, regarding their differentiability). These results can be put to good use both in the economic analysis of the problem and in the design of numerical methods. The second reason is that, as a default, we can use value function iteration (as explained in Rust, 1996 and Cai and Judd, 2014), a solution method that is particularly reliable, although often slow.

We can rewrite the Bellman operator as:

$$V(k_t, z_t) - \max_{k_{t+1}} \left[u \left(e^{z_t} k_t^{\alpha} + (1-\delta) k_t - k_{t+1} \right) + \beta \mathbb{E}_t V(k_{t+1}, z_{t+1}) \right] = 0,$$

for all k_t and z_t. If we define:

$$\mathcal{H}(d) = V(k_t, z_t) - \max_{k_{t+1}} \left[u\left(e^{z_t}k_t^\alpha + (1-\delta)k_t - k_{t+1}\right) + \beta \mathbb{E}_t V(k_{t+1}, z_{t+1}) \right] = 0, \quad (6)$$

for all k_t and z_t, where $d(\cdot, \cdot) = V(\cdot, \cdot)$, we see how the operator $\mathcal{H}$, a rewrite of the Bellman operator, takes the value function $V(\cdot, \cdot)$ and obtains a zero. More precisely, Eq. (6) is an integral equation given the presence of the expectation operator. This can lead to some nontrivial measure theory considerations that we leave aside.

3.3 Euler Equation

We have outlined several reasons why casting the problem in terms of a value function is attractive. Unfortunately, this formulation can be difficult. If the model does not satisfy the two fundamental welfare theorems, we cannot easily move between the social planner's problem and the competitive equilibrium. In that case, also, the value function of the household and firms will require laws of motion for individual and aggregate state variables that can be challenging to characterize.[b]

   An alternative is to work directly with the set of equilibrium conditions of the model. These include the first-order conditions for households, firms, and, if specified, government, budget and resource constraints, market clearing conditions, and laws of motion for exogenous processes. Since, at the core of these equilibrium conditions, we will have the Euler equations for the agents in the model that encode optimal behavior (with the other conditions being somewhat mechanical), this approach is commonly known as the Euler equation method (sometimes also referred to as solving the equilibrium conditions of the models). This solution strategy is extremely general and it allows us to handle non-Pareto efficient economies without further complications.

   In the case of the stochastic neoclassical growth model, the Euler equation for the sequential problem defined by Eqs. (2)–(4) is:

$$u'(c_t) = \beta \mathbb{E}_t \left[u'(c_{t+1}) \left(\alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} + 1 - \delta \right) \right]. \quad (7)$$

Again, under standard technical conditions, there is a decision rule $g : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+^2$ for the social planner that gives the optimal choice of consumption ($g^1(k_t, z_t)$) and capital tomorrow ($g^2(k_t, z_t)$) given capital, k_t, and productivity, z_t, today. Then, we can rewrite the first-order condition as:

$$u'\left(g^1(k_t, z_t)\right) = \beta \mathbb{E}_t \left[u'\left(g^1\left(g^2(k_t, z_t), z_{t+1}\right)\right) \left(\alpha e^{\rho z_t + \sigma \varepsilon_{t+1}} \left(g^2(k_t, z_t)\right)^{\alpha-1} + 1 - \delta \right) \right],$$

---

[b] See Hansen and Prescott (1995), for examples of how to recast a non-Pareto optimal economy into the mold of an associated Pareto-optimal problem.

for all k_t and z_t, where we have used the law of motion for productivity (4) to substitute for z_{t+1} or, alternatively:

$$\left(\begin{array}{c} u'\left(g^1(k_t, z_t)\right) \\ -\beta\mathbb{E}_t\left[u'\left(g^1\left(g^2(k_t, z_t), z_{t+1}\right)\right)\left(\alpha e^{\rho z_t + \sigma\varepsilon_{t+1}}\left(g^2(k_t, z_t)\right)^{\alpha-1} + 1 - \delta\right)\right] \end{array} \right) = 0, \quad (8)$$

for all k_t and z_t (note the composition of functions $g^1\left(g^2(k_t, z_t), z_{t+1}\right)$ when evaluating consumption at $t + 1$). We also have the resource constraint:

$$g^1(k_t, z_t) + g^2(k_t, z_t) = e^{z_t}k_t^\alpha + (1 - \delta)k_t \quad (9)$$

Then, we have a functional equation where the unknown object is the decision rule g. Mapping Eqs. (8) and (9) into our operator $\mathcal{H}$ is straightforward:

$$\mathcal{H}(d) = \left\{ \begin{array}{c} u'\left(g^1(k_t, z_t)\right) \\ -\beta\mathbb{E}_t\left[u'\left(g^1\left(g^2(k_t, z_t), z_{t+1}\right)\right)\left(\alpha e^{\rho z_t + \sigma\varepsilon_{t+1}}\left(g^2(k_t, z_t)\right)^{\alpha-1} + 1 - \delta\right)\right] = \mathbf{0}, \\ g^1(k_t, z_t) + g^2(k_t, z_t) - e^{z_t}k_t^\alpha - (1 - \delta)k_t \end{array} \right.$$

for all k_t and z_t, where $d = g$.

In this simple model, we could also have substituted the resource constraint in Eq. (8) and solved for a one-dimensional decision rule, but by leaving Eqs. (8) and (9), we illustrate how to handle cases where this substitution is either infeasible or inadvisable.

An additional consideration that we need to take care of is that the Euler equation (7) is only a necessary condition. Thus, after finding $g(\cdot, \cdot)$, we would also need to ensure that a transversality condition of the form:

$$\lim_{t\to\infty} \beta^t \frac{u'(c_t)}{u'(c_0)} k_t = 0$$

(or a related one) is satisfied. We will describe below how we build our solution methods to ensure that this is, indeed, the case.

3.4 Conditional Expectations

We have a considerable degree of flexibility in how we specify $\mathcal{H}$ and d. For instance, if we go back to the Euler equation (7):

$$u'(c_t) = \beta\mathbb{E}_t\left[u'(c_{t+1})\left(\alpha e^{z_{t+1}}k_{t+1}^{\alpha-1} + 1 - \delta\right)\right]$$

we may want to find the unknown conditional expectation:

$$\mathbb{E}_t\left[u'(c_{t+1})\left(\alpha e^{z_{t+1}}k_{t+1}^{\alpha-1} + 1 - \delta\right)\right].$$

This may be the case either because the conditional expectation is the object of interest in the analysis or because solving for the conditional expectation avoids problems associated

with the decision rule. For example, we could enrich the stochastic neoclassical growth model with additional constraints (such as a nonnegative investment: $k_{t+1} \geq (1 - \delta)k_t$) that induce kinks or other undesirable properties in the decision rules. Even when those features appear, the conditional expectation (since it smooths over different realizations of the productivity shock) may still have properties such as differentiability that the researcher can successfully exploit either in her numerical solution or later in the economic analysis.[c]

To see how this would work, we can define $g : \mathbb{R}_+ \times \mathbb{R} \to \mathbb{R}_+$:

$$g(k_t, z_t) = \mathbb{E}_t \left[u'(c_{t+1}) \left(\alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} + 1 - \delta \right) \right] \tag{10}$$

where we take advantage of $\mathbb{E}_t$ being a function of the states of the economy. Going back to our the Euler equation (7) and the resource constraint (3), if we have access to g, we can find:

$$c_t = u'(\beta g(k_t, z_t))^{-1} \tag{11}$$

and

$$k_{t+1} = e^{z_t} k_t^\alpha + (1 - \delta)k_t - u'(\beta g(k_t, z_t))^{-1}.$$

Thus, knowledge of the conditional expectation allows us to recover all the other endogenous variables of interest in the model. To save on notation, we write $c_t = c_{g,t}$ and $k_{t+1} = k_{g,t}$ to denote the values of c_t and k_{t+1} implied by g. Similarly:

$$c_{t+1} = c_{g,t+1} = u'(\beta g(k_{t+1}, z_{t+1}))^{-1} = u'\left(\beta g\left(k_{g,t}, z_{t+1}\right)\right)^{-1}$$

is the value of c_{t+1} implied by the recursive application of g.

To solve for g, we use its definition in Eq. (10):

$$g(k_t, z_t) = \beta \mathbb{E}_t \left[u'\left(c_{g,t+1}\right) \left(\alpha e^{\rho z_t + \sigma \varepsilon_{t+1}} k_{g,t}^{\alpha-1} + 1 - \delta \right) \right]$$

and write:

$$\mathcal{H}(d) = g(k_t, z_t) - \beta \mathbb{E}_t \left[u'\left(c_{g,t+1}\right) \left(\alpha e^{\rho z_t + \sigma \varepsilon_{t+1}} k_{g,t}^{\alpha-1} + 1 - \delta \right) \right] = \mathbf{0}$$

where $d = g$.

---

[c] See Fernández-Villaverde et al. (2015a) for an example. The paper is interested in solving a New Keynesian business cycle model with a zero lower bound (ZLB) on the nominal interest rate. This ZLB creates a kink on the function that maps states of the model into nominal interest rates. The paper gets around this problem by solving for consumption, inflation, and an auxiliary variable that encodes information similar to that of a conditional expectation. Once these functions have been found, the rest of the endogenous variables of the model, including the nominal interest rate, can be derived without additional approximations. In particular, the ZLB is always satisfied.

3.5 The Way Forward

We have argued that a large number of problems in macroeconomics can be expressed in terms of a functional equation problem

$$\mathcal{H}(d) = \mathbf{0}$$

and we have illustrated our assertion by building the operator $\mathcal{H}$ for a value function, for an Euler equation problem, and for a conditional expectation problem. Our examples, though, do not constitute an exhaustive list. Dozens of other cases can be constructed following the same ideas.

We will move now to study the two main families of solution methods for functional equation problems: perturbation and projection methods. Both families replace the unknown function d for an approximation $d^j(\mathbf{x}, \theta)$ in terms of the state variables of the model $\mathbf{x}$ and a vector of coefficients θ and a degree of approximation j (we are deliberately being ambiguous about the interpretation of that degree). We will use the terminology "parameters" to refer to objects describing the preferences, technology, and information sets of the model. The discount factor, risk aversion, the depreciation rate, or the persistence of the productivity shock are examples of parameters. We will call the numerical terms "coefficients" in the numerical solution. While the "parameters" usually have a clear economic interpretation associated with them, the "coefficients" will, most of the time, lack such interpretation.

Remark 3 (Structural parameters?) We are carefully avoiding the adjective "structural" when we discuss the parameters of the model. Here we follow Hurwicz (1962), who defined a "structural parameter" as a parameter that was invariant to a class of policy interventions the researcher is interested in analyzing. Many parameters of interest may not be "structural" in Hurwicz's sense. For example, the persistence of a technology shock may depend on the barriers to entry/exit in the goods and services industries and how quickly technological innovations can diffuse. These barriers may change with variations in competition policy. See a more detailed discussion on the "structural" character of parameters in DSGE models as well as empirical evidence in Fernández-Villaverde and Rubio-Ramírez (2008).

The states of the model will be determined by the structure of the model. Even if, in the words of Thomas Sargent, "finding the states is an art" (meaning both that there is no constructive algorithm to do so and that the researcher may be able to find different sets of states that accomplish the goal of fully describing the situation of the model, some of which may be more useful than the others in one context but less so in another one), determining the states is a step previous to the numerical solution of the model and, therefore, outside the purview of this chapter.

4. PERTURBATION

Perturbation methods build approximate solutions to a DSGE economy by starting from the exact solution of a particular case of the model or from the solution of a nearby model whose solution we have access to. Perturbation methods are also known as asymptotic methods, although we will avoid such a name because it risks confusion with related techniques regarding the large sample properties of estimators as the ones we will introduce in Part II of the chapter. In their more common incarnation in macroeconomics, perturbation algorithms build Taylor series approximations to the solution of a DSGE model around its deterministic steady state using implicit-function theorems. However, other perturbation approaches are possible, and we should always talk about *a* perturbation of the model instead of *the* perturbation. With a long tradition in physics and other natural sciences, perturbation theory was popularized in economics by Judd and Guu (1993) and it has been authoritatively presented by Judd (1998), Judd and Guu (2001), and Jin and Judd (2002).[d] Since there is much relevant material about perturbation problems in economics (including a formal mathematical background regarding solvability conditions, and more advanced perturbation techniques such as gauges and Padé approximants) that we cannot cover in this chapter, we refer the interested reader to these sources.

Over the last two decades, perturbation methods have gained much popularity among researchers for four reasons. First, perturbation solutions are accurate around an approximation point. Perturbation methods find an approximate solution that is inherently local. In other words, the approximated solution is extremely close to the exact, yet unknown, solution around the point where we take the Taylor series expansion. However, researchers have documented that perturbation often displays good global properties along a wide range of state variable values. See the evidence in Judd (1998); Aruoba et al. (2006) and Caldara et al. (2012). Also, as we will discuss below, the perturbed solution can be employed as an input for other solution methods, such as value function iteration. Second, the structure of the approximate solution is intuitive and easily interpretable. For example, a second-order expansion of a DSGE model includes a term that corrects for the standard deviation of the shocks that drive the stochastic dynamics of the economy. This term, which captures precautionary behavior, breaks the certainty equivalence of linear approximations that makes the discussion of welfare and risk in a linearized world challenging. Third, as we will explain below, a traditional linearization is nothing but a first-order perturbation. Hence, economists can import into perturbation theory much of their knowledge and practical experience while, simultaneously, being able to incorporate the formal results developed in applied mathematics. Fourth, thanks

---

[d] Perturbation approaches were already widely used in physics in the 19th century. They became a central tool in the natural sciences with the development of quantum mechanics in the first half of the 20th century. Good general references on perturbation methods are Simmonds and Mann (1997) and Bender and Orszag (1999).

to open-source software such as Dynare and Dynare++ (developed by Stéphane Adjemian, Michel Juillard, and their team of collaborators), or Perturbation AIM (developed by Eric Swanson, Gary Anderson, and Andrew Levin) higher-order perturbations are easy to compute even for practitioners less familiar with numerical methods.[e]

4.1 The Framework

Perturbation methods solve the functional equation problem:

$$\mathcal{H}(d) = \mathbf{0}$$

by specifying a Taylor series expansion to the unknown function $d : \Omega \to \mathbb{R}^m$ in terms of the n state variables of the model $\mathbf{x}$ and some coefficients θ. For example, a second-order Taylor expansion has the form:

$$d_i^2(\mathbf{x}, \theta) = \theta_{i,0} + \theta_{i,1}(\mathbf{x} - \mathbf{x}_0)' + (\mathbf{x} - \mathbf{x}_0)\theta_{i,2}(\mathbf{x} - \mathbf{x}_0)', \text{ for } i = 1, \dots, m \qquad (12)$$

where $\mathbf{x}'$ is the transpose of $\mathbf{x}$, $\mathbf{x}_0$ is the point around which we build our perturbation solution, $\theta_{i,0}$ is a scalar, $\theta_{i,1}$ is an n-dimensional vector, $\theta_{i,2}$ is a $n \times n$ matrix, and where $\theta_{i,0}$, $\theta_{i,1}$, and $\theta_{i,2}$ depend on the derivatives of d that we will find using implicit-function theorems.[f]

In comparison, the traditional linearization approach popularized by King et al. (2002) delivers a solution of the form:

$$d_i^1(\mathbf{x}, \theta) = \widetilde{\theta}_{i,0} + \theta_{i,1}(\mathbf{x} - \mathbf{x}_0)'$$

where the vector $\theta_{i,1}$ is the same as in Eq. (12) and $\widetilde{\theta}_{i,0} = \theta_{i,0}$ if $j = 1$. In other words, linearization is nothing more than a first-order perturbation. Higher-order approximations generalize the structure of the linearized solution by including additional terms. Instead of being an *ad hoc* procedure (as it was sometimes understood in the 1980s and 1990s), linearization can borrow from a large set of well-established results in perturbation theory. But the direction of influence also goes in the opposite direction: we can use much of our accumulated understanding on linearized DSGE models (such as how to efficiently solve for the coefficients $\theta_{i,0}$ and $\theta_{i,1}$ and how to interpret their economic meaning) in perturbation.

---

[e] Dynare (a toolbox for Matlab) and Dynare++ (a stand-alone application) allow the researcher to write, in a concise and transparent language, the equilibrium conditions of a DSGE model and find a perturbation solution to it, up to the third order in Dynare and an arbitrary order in Dynare++. See http://www.dynare.org/. Perturbation AIM follows a similar philosophy, but with the additional advantage of being able to rely on Mathematica and its efficient use of arbitrary-precision arithmetic. This is important, for example, in models with extreme curvature such as those with Epstein–Zin preferences or habit persistence. See http://www.ericswanson.us/perturbation.html.

[f] Strictly speaking, the order of the approximation is given by the first nonzero or dominant term, but since in DSGE models the $\theta_{i,1}$ are typically different from zero, we can proceed without further qualifications.

Remark 4 (Linearization vs loglinearization) Linearization and, more generally, perturbation, can be performed in the level of the state variables or after applying some change of variables to any (or all) the variables of the model. Loglinearization, for example, approximates the solution of the model in terms of the log-deviations of the variables with respect to their steady state. That is, for a variable $x \in \mathbf{x}$, we define:

$$\hat{x} = \log \frac{x}{\bar{x}}$$

where $\bar{x}$ is its steady-state value, and then we find a second-order approximation:

$$d_i^2(\hat{\mathbf{x}}, \theta) = \theta_{i,0} + \theta_{i,1}(\hat{\mathbf{x}} - \hat{\mathbf{x}}_0)' + (\hat{\mathbf{x}} - \hat{\mathbf{x}}_0)\theta_{i,2}(\hat{\mathbf{x}} - \hat{\mathbf{x}}_0)', \text{ for } i = 1, \ldots, m.$$

If $\mathbf{x}_0$ is the deterministic steady state (this is more often than not the case), $\hat{\mathbf{x}}_0 = \mathbf{0}$, since for all variables $x \in \mathbf{x}$

$$\hat{x}_0 = \log \frac{x}{\bar{x}} = 0.$$

This result provides a compact representation:

$$d_i^2(\hat{\mathbf{x}}, \theta) = \theta_{i,0} + \theta_{i,1}\hat{\mathbf{x}}' + \hat{\mathbf{x}}\theta_{i,2}\hat{\mathbf{x}}', \text{ for } i = 1, \ldots, m.$$

Loglinear solutions are easy to read (the loglinear deviation is an approximation of the percentage deviation with respect to the steady state) and, in some circumstances, they can improve the accuracy of the solution. We will revisit the change of variables later in the chapter.

Before getting into technical details of how to implement perturbation methods, we will briefly distinguish between regular and singular perturbations. A regular perturbation is a situation where a *small* change in the problem induces a *small* change in the solution. An example is a standard New Keynesian model (Woodford, 2003). A small change in the standard deviation of the monetary policy shock will lead to a small change in the properties of the equilibrium dynamics (ie, the standard deviation and autocorrelation of variables such as output or inflation). A singular perturbation is a situation where a *small* change in the problem induces a *large* change in the solution. An example can be an excess demand function. A small change in the excess demand function may lead to an arbitrarily large change in the price that clears the market.

Many problems involving DSGE models will result in regular perturbations. Thus, we will concentrate on them. But this is not necessarily the case. For instance, introducing a new asset in an incomplete market model can lead to large changes in the solution. As researchers pay more attention to models with financial frictions and/or market incompleteness, this class of problems may become common. Researchers will need to learn more about how to apply singular perturbations. See, for pioneering work,

Judd and Guu (1993), and a presentation of bifurcation methods for singular problems in Judd (1998).

4.2 The General Case

We are now ready to deal with the details of how to implement a perturbation. We present first the general case of how to find a perturbation solution of a DSGE model by (1) using the equilibrium conditions of the model and (2) by finding a higher-order Taylor series approximation. Once we have mastered this task, it would be straightforward to extend the results to other problems, such as the solution of a value function, and to conceive other possible perturbation schemes. This section follows much of the structure and notation of section 3 in Schmitt-Grohé and Uribe (2004).

We start by writing the equilibrium conditions of the model as

$$\mathbb{E}_t \mathcal{H}(\mathbf{y}, \mathbf{y}', \mathbf{x}, \mathbf{x}') = 0, \tag{13}$$

where $\mathbf{y}$ is an $n_y \times 1$ vector of controls, $\mathbf{x}$ is an $n_x \times 1$ vector of states, and $n = n_x + n_y$. The operator $\mathcal{H} : \mathbb{R}^{n_y} \times \mathbb{R}^{n_y} \times \mathbb{R}^{n_x} \times \mathbb{R}^{n_x} \to \mathbb{R}^n$ stacks all the equilibrium conditions, some of which will have expectational terms, some of which will not. Without loss of generality, and with a slight change of notation with respect to Section 3, we place the conditional expectation operator outside $\mathcal{H}$: for those equilibrium conditions without expectations, the conditional expectation operator will not have any impact. Moving $\mathbb{E}_t$ outside $\mathcal{H}$ will make some of the derivations below easier to follow. Also, to save on space, when there is no ambiguity, we will employ the recursive notation where x represents a variable at period t and x' a variable at period $t + 1$.

It will also be convenient to separate the endogenous state variables (capital, asset positions, etc.) from the exogenous state variables (productivity shocks, preference shocks, etc.). In that way, it will be easier to see the variables on which the perturbation parameter that we will introduce below will have a direct effect. Thus, we partition the state vector $\mathbf{x}$ (and taking transposes) as

$$\mathbf{x} = [\mathbf{x}_1'; \mathbf{x}_2']'.$$

where $\mathbf{x}_1$ is an $(n_x - n_\epsilon) \times 1$ vector of endogenous state variables and $\mathbf{x}_2$ is an $n_\epsilon \times 1$ vector of exogenous state variables. Let $\tilde{n} = n_x - n_\epsilon$.

4.2.1 Steady State

If we suppress the stochastic component of the model (more details below), we can define the deterministic steady-state of the model as vectors $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ such that:

$$\mathcal{H}(\bar{\mathbf{y}}, \bar{\mathbf{y}}, \bar{\mathbf{x}}, \bar{\mathbf{x}}) = 0. \tag{14}$$

The solution $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ of this problem can often be found analytically. When this cannot be done, it is possible to resort to a standard nonlinear equation solver.

The previous paragraph glossed over the possibility that the model we are dealing with either does not have a steady state or that it has several of them (in fact, we can even have a continuum of steady states). Given our level of abstraction with the definition of Eq. (13), we cannot rule out any of these possibilities. Galor (2007) discusses in detail the existence and stability (local and global) of steady states in discrete time dynamic models.

A case of interest is when the model, instead of having a steady state, has a balanced growth path (BGP): that is, when the variables of the model (with possibly some exceptions such as labor) grow at the same rate (either deterministic or stochastic). Given that perturbation is an inherently local solution method, we cannot deal directly with solving such a model. However, on many occasions, we can rescale the variables x_t in the model by the trend μ_t:

$$\hat{x}_t = \frac{x_t}{\mu_t}$$

to render them stationary (the trend itself may be a complicated function of some technological processes in the economy, as when we have both neutral and investment-specific technological change; see Fernández-Villaverde and Rubio-Ramírez, 2007). Then, we can undertake the perturbation in the rescaled variable $\hat{x}_t$ and undo the rescaling when using the approximated solution for analysis and simulation.[g]

Remark 5 (Simplifying the solution of $(\bar{x},\bar{y})$) Finding the solution $(\bar{x},\bar{y})$ can often be made much easier by using two "tricks." One is to substitute some of the variables away from the operator $\mathcal{H}(\cdot)$ and reduce the system from being one of n equations in n unknowns into a system of $n' < n$ equations in n' unknowns. For example, if we have a law of motion for capital involving capital next period, capital next period, and investment:

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

we can substitute out investment throughout the whole system just by writing:

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

Since the complexity of solving a nonlinear system of equations grows exponentially in the dimension of the problem (see Sikorski, 1985, for classic results on computational complexity), even a few substitutions can produce considerable improvements.

A second possibility is to select parameter values to pin down one or more variables of the model and then to solve all the other variables as a function of the fixed variables. To illustrate this point, let us consider a simple stochastic neoclassical growth model with a representative household with utility function:

---

[g] This rescaling is also useful with projection methods since they need a bounded domain of the state variables.

$$\mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \left(\log c_t - \psi \frac{l_t^{1+\eta}}{1+\eta} \right)$$

where the notation is the same as in Section 3 and a production function:

$$output_t = A_t k_t^{\alpha} l_t^{1-\alpha}$$

where A_t is the productivity level and a law of motion for capital:

$$k_{t+1} = output_t + (1-\delta)k_t - c_t.$$

This model has a static optimality condition for labor supply of the form:

$$\psi c_t l_t^{\eta} = w_t$$

where w_t is the wage. Since with the log-CRRA utility function that we selected l_t does not have a natural unit, we can fix its deterministic steady-state value, for example, $\bar{l} = 1$. This normalization is as good as any other and the researcher can pick the normalization that best suits her needs.

Then, we can analytically solve the rest of the equilibrium conditions of the model for all other endogenous variables as a function of $\bar{l} = 1$. After doing so, we return to the static optimality condition to obtain the value of the parameter ψ as:

$$\psi = \frac{\bar{w}}{\bar{c}\bar{l}^{\eta}} = \frac{\bar{w}}{\bar{c}}$$

where $\bar{c}$ and $\bar{w}$ are the deterministic steady-state values of consumption and wage, respectively. An alternative way to think about this procedure is to realize that it is often easier to find parameter values that imply a particular endogenous variable value than to solve for those endogenous variable values as a function of an arbitrary parameter value.

Although not strictly needed to find $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, other good practices include picking units that make algebraic and numerical computations convenient to handle. For example, we can pick units to make $\overline{output} = 1$. Again, in the context of the stochastic neoclassical growth model, we will have:

$$\overline{output} = 1 = \overline{A}\bar{k}^{\alpha}\bar{l}^{1-\alpha} = \overline{A}\bar{k}^{\alpha}.$$

Then, we can find:

$$\overline{A} = \frac{1}{\bar{k}^{\alpha}}$$

and wages:

$$\overline{w} = (1-\alpha)\frac{\overline{output}}{\bar{l}} = 1 - \alpha.$$

Going back to the intertemporal Euler equation:

$$\frac{1}{\bar{c}} = \frac{1}{\bar{c}}\beta(1 + \bar{r} - \delta)$$

where r is the rental rate of capital and δ is depreciation, we find:

$$\bar{r} = \frac{1}{\beta} - 1 + \delta.$$

Since:

$$\bar{r} = \alpha\frac{\overline{output}}{\bar{k}} = \frac{\alpha}{\bar{k}}$$

we get:

$$\bar{k} = \frac{\alpha}{\frac{1}{\beta} - 1 + \delta}$$

and:

$$\bar{c} = \overline{output} - \delta\bar{k} = 1 - \delta\frac{\alpha}{\frac{1}{\beta} - 1 + \delta},$$

from which:

$$\psi = \frac{\bar{w}}{\bar{c}} = \frac{1 - \alpha}{1 - \delta\dfrac{\alpha}{\dfrac{1}{\beta} - 1 + \delta}}$$

In this example, two judicious choices of units ($\bar{l} = \overline{output} = 1$) render the solution of the deterministic steady state a straightforward exercise. While the deterministic steady state of more complicated models would be harder to solve, experience suggests that following the advice in this remark dramatically simplifies the task in many situations.

The deterministic steady state $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ is different from a fixed point $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ of (13):

$$\mathbb{E}_t\mathcal{H}(\hat{\mathbf{y}}, \hat{\mathbf{y}}, \hat{\mathbf{x}}, \hat{\mathbf{x}}) = 0,$$

because in the former case we eliminate the conditional expectation operator while in the latter we do not. The vector $(\hat{\mathbf{x}}, \hat{\mathbf{y}})$ is sometimes known as the stochastic steady state (although, since we find the idea of mixing the words "stochastic" and "steady state" in the same term confusing, we will avoid that terminology).

4.2.2 Exogenous Stochastic Process
For the exogenous stochastic variables, we specify a stochastic process of the form:

$$\mathbf{x}_2' = \mathbf{C}(\mathbf{x}_2) + \sigma\eta_\epsilon\epsilon' \tag{15}$$

where $\mathbf{C}$ is a potentially nonlinear function. At our current level of abstraction, we are not imposing much structure on $\mathbf{C}$, but in concrete applications, we will need to add more constraints. For example, researchers often assume that all the eigenvalues of the Hessian matrix of $\mathbf{C}$ evaluated at the steady state $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ lie within the unit circle. The vector ϵ' contains the n_ϵ exogenous zero-mean innovations. Initially, we only assume that ϵ' is independent and identically distributed with finite second moments, meaning that we do not rely on any distributional assumption. Thus, the innovations may be non-Gaussian. This is denoted by $\epsilon' \sim iid(\mathbf{0}, \mathbf{I})$. Additional moment restrictions will be introduced as needed in each concrete application. Finally, η_ϵ is an $n_\epsilon \times n_\epsilon$ matrix that determines the variances-covariances of the innovations, and $\sigma \geq 0$ is a perturbation parameter that scales η_ϵ.

Often, it will be the case that $\mathbf{C}$ is linear:

$$x'_2 = Cx_2 + \sigma\eta_\epsilon\epsilon'$$

where C is an $n_\epsilon \times n_\epsilon$ matrix, with all its eigenvalues with modulus less than one.

Remark 6 (Linearity of innovations) The assumption that innovations enter linearly in Eq. (15) may appear restrictive, but it is without loss of generality. Imagine that instead of Eq. (15), we have:

$$\mathbf{x}_{2,t} = \mathbf{D}(\mathbf{x}_{2,t-1}, \sigma\eta_\epsilon\epsilon_t).$$

This richer structure can be handled by extending the state vector by incorporating the innovations ϵ in the state vector. In particular, let

$$\tilde{\mathbf{x}}_{2,t} = \begin{bmatrix} \mathbf{x}_{2,t-1} \\ \epsilon_t \end{bmatrix}$$

and

$$\tilde{\epsilon}_{t+1} = \begin{bmatrix} \mathbf{0}_{n_\epsilon \times 1} \\ \epsilon_{t+1} \end{bmatrix}$$

Then, we can write

$$\mathbf{x}_{2,t} = \tilde{\mathbf{D}}\left(\tilde{\mathbf{x}}_{2,t}, \sigma\eta_\epsilon\right).$$

The new stochastic process is given by:

$$\begin{bmatrix} \mathbf{x}_{2,t} \\ \epsilon_{t+1} \end{bmatrix} = \begin{bmatrix} \tilde{\mathbf{D}}\left(\tilde{\mathbf{x}}_{2,t}, \sigma\eta_\epsilon\right) \\ 0 \end{bmatrix} + \begin{bmatrix} \mathbf{0}_{n_\epsilon \times 1} \\ \epsilon_{t+1} \end{bmatrix}$$

where $\mathbf{u}_{t+1} \sim iid(\mathbf{0}, \mathbf{I})$ or, switching back to the recursive notation:

$$\tilde{\mathbf{x}}'_2 = \mathbf{C}(\tilde{\mathbf{x}}_2) + \tilde{\epsilon}'$$

To illustrate this point, we use the popular case of time-varying volatility, which, it has been argued, is of considerable importance to understand the dynamics of aggregate variables (see Bloom, 2009 and Fernández-Villaverde et al., 2011). Imagine that we have a stochastic volatility process for productivity a_t:

$$\log a_t = \rho_a \log a_{t-1} + \lambda_t v_t, v_t \sim N(0,1)$$

where λ_t is the standard deviation of the innovation v_t. The standard deviation follows another autoregressive process:

$$\log \lambda_t = \overline{\lambda} + \rho_\lambda \log \lambda_{t-1} + \psi \eta_t, \eta_t \sim N(0,1).$$

To fit this system into our notation, we only need to define:

$$\tilde{\mathbf{x}}_{2,t} = \begin{bmatrix} \log a_{t-1} \\ \log \lambda_{t-1} \\ v_t \\ \eta_t \end{bmatrix}$$

and

$$\tilde{\epsilon}_{t+1} = \begin{bmatrix} \mathbf{0}_{2\times 1} \\ \epsilon_{t+1} \end{bmatrix}.$$

Note, also, how the perturbation parameter controls both the innovation v_t and its standard deviation λ_t.

Perturbation methods are well suited to the solution of models with time-varying volatility because these models have a richness of state variables: for each stochastic process, we need to keep track of the level of the process and its variance. The projection methods that we will describe in the next section will have problems dealing with this large number of state variables.

Only one perturbation parameter appears in Eq. (15), even if we have a model with many innovations. The matrix η_ϵ takes account of relative sizes (and comovements) of the different innovations. If we set $\sigma = 0$, we have a deterministic model.

Remark 7 (Perturbation parameter) In the main text, we introduced the perturbation parameter as controlling the standard deviation of the stochastic process:

$$\mathbf{x}'_2 = \mathbf{C}(\mathbf{x}_2) + \sigma \eta_\epsilon \epsilon'.$$

However, we should not hew too closely to this choice. First, there may be occasions where placing the perturbation in another parameter could offer better accuracy and/or deeper insights into the behavior of the model. For example, in models with Epstein–Zin preferences, Hansen et al. (2008) perform a perturbation around an elasticity of intertemporal

substitution equal to 1. Also, the choice of perturbation would be different in a continuous time model, where it is usually more convenient to control the variance.

We depart from Samuelson (1970) and Jin and Judd (2002), who impose a bounded support for the innovations of the model. By doing so, these authors avoid problems with the stability of the simulations coming from the perturbation solution that we will discuss below. Instead, we will introduce pruning as an alternative strategy to fix these problems.

4.2.3 Solution of the Model

The solution of the model will be given by a set of decision rules for the control variables

$$\mathbf{y} = \mathbf{g}(\mathbf{x}; \sigma), \tag{16}$$

and for the state variables

$$\mathbf{x}' = \mathbf{h}(\mathbf{x}; \sigma) + \sigma \eta \epsilon', \tag{17}$$

where $\mathbf{g}$ maps $\mathbb{R}^{n_x} \times \mathbb{R}^+$ into R^{n_y} and $\mathbf{h}$ maps $\mathbb{R}^{n_x} \times \mathbb{R}^+$ into $\mathbb{R}^{n_x}$. Note our timing convention: controls depend on current states, while states next period depend on states today and the innovations tomorrow. By defining additional state variables that store the information of states with leads and lags, this structure is sufficiently flexible to capture rich dynamics. Also, we separate states $\mathbf{x}$ and the perturbation parameter σ by a semicolon to emphasize the difference between both elements.

The $n_x \times n_\epsilon$ matrix η is:

$$\eta = \begin{bmatrix} \emptyset \\ \eta_\epsilon \end{bmatrix}$$

where the first n_x rows come from the states today determining the endogenous states tomorrow and the last n_ϵ rows come from the exogenous states tomorrow depending on the states today and the innovations tomorrow.

The goal of perturbation is to find a Taylor series expansion of the functions $\mathbf{g}$ and $\mathbf{h}$ around an appropriate point. A natural candidate for this point is the deterministic steady state, $\mathbf{x}_t = \bar{\mathbf{x}}$ and $\sigma = 0$. As we argued above, we know how to compute this steady state and, consequently, how to evaluate the derivatives of the operator $\mathcal{H}(\cdot)$ that we will require.

First, note by the definition of the deterministic steady state (14) we have that

$$\bar{\mathbf{y}} = \mathbf{g}(\bar{\mathbf{x}}; 0) \tag{18}$$

and

$$\bar{\mathbf{x}} = \mathbf{h}(\bar{\mathbf{x}}; 0). \tag{19}$$

Second, we plug-in the unknown solution on the operator $\mathcal{H}$ and define the new operator $F : \mathbb{R}^{n_x + 1} \to \mathbb{R}^n$:

$$F(\mathbf{x}; \sigma) \equiv \mathbb{E}_t \mathcal{H}(\mathbf{g}(\mathbf{x}; \sigma), \mathbf{g}(\mathbf{h}(\mathbf{x}; \sigma) + \sigma \eta \epsilon', \sigma), \mathbf{x}, \mathbf{h}(\mathbf{x}; \sigma) + \sigma \eta \epsilon') = \mathbf{0}.$$

Since $F(\mathbf{x};\sigma) = 0$ for any values of $\mathbf{x}$ and σ , any derivatives of F must also be zero:

$$F_{x_i^k \sigma^j}(\mathbf{x}; \sigma) = 0, \forall \mathbf{x}, \sigma, i, k, j,$$

where $F_{x_i^k \sigma^j}(\mathbf{x}; \sigma)$ is the derivative of F with respect to the i-th component x_i of $\mathbf{x}$ taken k times and with respect to σ taken j times evaluated at $(\mathbf{x};\sigma)$. Intuitively, the solution of the model must satisfy the equilibrium conditions for all possible values of the states and σ. Thus, any change in the values of the states or of σ must still keep the operator F exactly at $\mathbf{0}$. We will exploit this important fact repeatedly.

Remark 8 (Existence of derivatives) We will assume, without further discussion, that all the relevant derivatives of the operator F exist in a neighborhood of $\bar{\mathbf{x}}$. These differentiability assumptions may be hard to check in concrete applications and more research in the area would be welcomed (see the classic work of Santos, 1992). However, the components that enter into F (utility functions, production functions, etc.) are usually smooth when we deal with DSGE models, which suggest that the existence of these derivatives is a heroic assumption (although the examples in Santos, 1993 are a cautionary sign). Judd (1998, p. 463) indicates, also, that if the derivative conditions were violated, our computations would display telltale signs that would alert the researcher to the underlying problems.

The derivative assumption, however, traces the frontiers of problems suitable for perturbation: if, for example, some variables are discrete or the relevant equilibrium conditions are nondifferentiable, perturbation cannot be applied. Two caveats about the previous statement are, nevertheless, worthwhile to highlight. First, the presence of expectations often transforms problems that appear discrete into continuous ones. For example, deciding whether or not to go to college can be "smoothed out" by a stochastic shock to college costs or by an effort variable that controls how hard the prospective student is applying to college or searching for funding. Second, even if the derivative assumption breaks down and the perturbation solution is not valid, it may still be an excellent guess for another solution method.

Remark 9 (Taking derivatives) The previous exposition demonstrates the central role of derivatives in perturbation methods. Except for simple examples, manually calculating these derivatives is too onerous. Thus, researchers need to rely on computers. A first possibility, numerical derivatives, is inadvisable Judd (1998, chapter 7). The errors created by numerical derivatives quickly accumulate and, after the second or third derivative, the perturbation solution is too contaminated by them to be of any real use. A second possibility is to exploit software that takes analytic derivatives, such as Mathematica or the symbolic toolbox of Matlab. This route is usually straightforward, but it may slow down the computation and require an inordinate amount of memory. A third final alternative

is to employ automatic differentiation, a technique that takes advantage of the application of the chain rule to a series of elementary arithmetic operations and functions (for how automatic differentiation can be applied to DSGE models, see Bastani and Guerrieri, 2008).

4.2.4 First-Order Perturbation

A first–order perturbation approximates $\mathbf{g}$ and $\mathbf{h}$ around $(\mathbf{x}; \sigma) = (\bar{\mathbf{x}}; 0)$ as:

$$\begin{aligned}
\mathbf{g}(\mathbf{x}; \sigma) &= \mathbf{g}(\bar{\mathbf{x}}; 0) + \mathbf{g_x}(\bar{\mathbf{x}}; 0)(\mathbf{x} - \bar{\mathbf{x}})' + \mathbf{g}_\sigma(\bar{\mathbf{x}}; 0)\sigma \\
\mathbf{h}(\mathbf{x}; \sigma) &= \mathbf{h}(\bar{\mathbf{x}}; 0) + \mathbf{h_x}(\bar{\mathbf{x}}; 0)(\mathbf{x} - \bar{\mathbf{x}})' + \mathbf{h}_\sigma(\bar{\mathbf{x}}; 0)\sigma
\end{aligned}$$

where $\mathbf{g_x}$ and $\mathbf{h_x}$ are the gradients of $\mathbf{g}$ and $\mathbf{h}$, respectively (including only the partial derivatives with respect to components of $\mathbf{x}$) and $\mathbf{g}_\sigma$ and $\mathbf{h}_\sigma$ the derivatives of $\mathbf{g}$ and $\mathbf{h}$ with respect to the perturbation parameter σ.

Using Eqs. (18) and (19), we can write

$$\begin{aligned}
\mathbf{g}(\mathbf{x}; \sigma) - \bar{\mathbf{y}} &= \mathbf{g_x}(\bar{\mathbf{x}}; 0)(\mathbf{x} - \bar{\mathbf{x}})' + \mathbf{g}_\sigma(\bar{\mathbf{x}}; 0)\sigma \\
\mathbf{h}(\mathbf{x}; \sigma) - \bar{\mathbf{x}} &= \mathbf{h_x}(\bar{\mathbf{x}}; 0)(\mathbf{x} - \bar{\mathbf{x}})' + \mathbf{h}_\sigma(\bar{\mathbf{x}}; 0)\sigma.
\end{aligned}$$

Since we know $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$, we only need to find $\mathbf{g_x}(\bar{\mathbf{x}}; 0)$, $\mathbf{g}_\sigma(\bar{\mathbf{x}}; 0)$, $\mathbf{h_x}(\bar{\mathbf{x}}; 0)$, and $\mathbf{h}_\sigma(\bar{\mathbf{x}}; 0)$ to evaluate the approximation at any arbitrary point $(\mathbf{x}, \sigma)$. We are searching for $n \times (n_x + 1)$ coefficients (the $n_x \times n_y$ terms in $\mathbf{g_x}(\bar{\mathbf{x}}; 0)$, the $n_x \times n_x$ terms in $\mathbf{h_x}(\bar{\mathbf{x}}; 0)$, the n_y terms in $\mathbf{g}_\sigma(\bar{\mathbf{x}}; 0)$, and the n_x terms in $\mathbf{h}_\sigma(\bar{\mathbf{x}}; 0)$).

These coefficients can be found by using:

$$F_{x_i}(\bar{\mathbf{x}}; 0) = 0, \forall i,$$

which gives us $n \times n_x$ equations and

$$F_\sigma(\bar{\mathbf{x}}; 0) = 0,$$

which gives us n equations.

But before doing so, and to avoid runaway notation, we need to introduce the use of tensors.

Remark 10 (Tensor notation) Tensor notation (or Einstein summation notation), commonly used in physics, keeps the algebra required to perform a perturbation at a manageable level by eliminating $\sum$ and ∂ signs. To further reduce clutter, the points of evaluation of a derivative are skipped when they are unambiguous from context. An nth-rank tensor in an m-dimensional space is an operator that has n indices and m^n components and obeys certain transformation rules. In our environment, $[\mathcal{H}_y]^i_\alpha$ is the (i, α) element of the derivative of $\mathcal{H}$ with respect to y:

1. The derivative of $\mathcal{H}$ with respect to y is an $n \times n_y$ matrix.
2. Thus, $[\mathcal{H}_y]^i_\alpha$ is the i-th row and α-th column element of this matrix.

3. When a subindex appears as a superindex in the next term, we are omitting a sum operator. For example,

$$[\mathcal{H}_y]^i_\alpha [\mathbf{g}_x]^\alpha_\beta [\mathbf{h}_x]^\beta_j = \sum_{\alpha=1}^{n_y} \sum_{\beta=1}^{n_x} \frac{\partial \mathcal{H}^i}{\partial y^\alpha} \frac{\partial \mathbf{g}^\alpha}{\partial x^\beta} \frac{\partial \mathbf{h}^\beta}{\partial x^j}.$$

4. The generalization to higher derivatives is direct. If we have $[\mathcal{H}_{y'y'}]^i_{\alpha\gamma}$:

    (a) $\mathcal{H}_{y'y'}$ is a three-dimensional array with n rows, n_y columns, and n_y pages.

    (b) Thus, $[\mathcal{H}_{y'y'}]^i_{\alpha\gamma}$ denotes the i-th row, α-th column element, and γ-th page of this matrix.

With the tensor notation, we can get into solving the system. First, $\mathbf{g_x}(\bar{\mathbf{x}}; 0)$ and $\mathbf{h_x}(\bar{\mathbf{x}}; 0)$ are the solution to:

$$[F_x(\bar{\mathbf{x}}; 0)]^i_j = [\mathcal{H}_{y'}]^i_\alpha [\mathbf{g}_x]^\alpha_\beta [\mathbf{h}_x]^\beta_j + [\mathcal{H}_y]^i_\alpha [\mathbf{g}_x]^\alpha_j + [\mathcal{H}_{x'}]^i_\beta [\mathbf{h}_x]^\beta_j + [\mathcal{H}_x]^i_j = \mathbf{0};$$

$$i = 1, \ldots, n; \quad j, \beta = 1, \ldots, n_x; \quad \alpha = 1, \ldots, n_y.$$

(20)

The derivatives of $\mathcal{H}$ evaluated at $(\mathbf{y}, \mathbf{y}', \mathbf{x}, \mathbf{x}') = (\bar{\mathbf{y}}, \bar{\mathbf{y}}, \bar{\mathbf{x}}, \bar{\mathbf{x}})$ are known. Therefore, we have a system of $n \times n_x$ quadratic equations in the $n \times n_x$ unknowns given by the elements of $\mathbf{g_x}(\bar{\mathbf{x}}; 0)$ and $\mathbf{h_x}(\bar{\mathbf{x}}; 0)$. After some algebra, the system (20) can be written as:

$$AP^2 - BP - C = \mathbf{0}$$

where the $\tilde{n} \times \tilde{n}$ matrix A, the $\tilde{n} \times \tilde{n}$ matrix B and the $\tilde{n} \times \tilde{n}$ matrix C involve terms from $[\mathcal{H}_{y'}]^i_\alpha$, $[\mathcal{H}_y]^i_\alpha$, $[\mathcal{H}_{x'}]^i_\beta$, and $[\mathcal{H}_x]^i_j$ and the $\tilde{n} \times \tilde{n}$ matrix P the terms $[\mathbf{h}_x]^\beta_j$ related to the law of motion of $\mathbf{x}_1$ (in our worked-out example of the next subsection, we will make this algebra explicit). We can solve this system with a standard quadratic matrix equation solver.

Remark 11 (Quadratic equation solvers) The literature has proposed several procedures to solve quadratic systems. Without being exhaustive, we can list Blanchard and Kahn (1980), King and Watson (1998), Uhlig (1999), Klein (2000), and Sims (2002). These different approaches vary in the details of how the solution to the system is found and how general they are (regarding the regularity conditions they require). But, conditional on applicability, all methods find the same policy functions since the linear space approximating a nonlinear space is unique.

For concision, we will only present one of the simplest of these procedures, as discussed by Uhlig (1999, pp. 43–45). Given

$$AP^2 - BP - C = \mathbf{0},$$

define the $2\tilde{n} \times 2\tilde{n}$ matrix:

$$D = \begin{bmatrix} A & \mathbf{0}_{\tilde{n}} \\ \mathbf{0}_{\tilde{n}} & I_{\tilde{n}} \end{bmatrix}$$

where $I_{\tilde{n}}$ is the $\tilde{n} \times \tilde{n}$ identity matrix and $\mathbf{0}_{\tilde{n}}$ the $\tilde{n} \times \tilde{n}$ zero matrix, and the $2\tilde{n} \times 2\tilde{n}$ matrix:

$$F = \begin{bmatrix} B & C \\ I_n & \mathbf{0}_n \end{bmatrix}$$

Let Q and Z be unitary matrices (ie, $Q^H Q = Z^H Z = I_{2\tilde{n}}$ where H is the complex Hermitian transposition operator). Let $\mathbf{\Phi}$ and $\mathbf{\Sigma}$ be upper triangular matrices with diagonal elements ϕ_{ii} and σ_{ii}. Then, we find the generalized Schur decomposition (QZ) of D and F:

$$Q'\Sigma Z = D$$
$$Q'\Phi Z = F$$

such that Σ and $\mathbf{\Phi}$ are diagonal and the ratios of diagonal elements $|\phi_{ii}/\sigma_{ii}|$ are in increasing order (there exists a QZ decomposition for every ordering of these ratios). In such a way, the stable (smaller than one) generalized eigenvalues of F with respect to D would come first and the unstable generalized eigenvalues (exceeding one and infinite) would come last. QZ decompositions are performed by standard numerical software such as Matlab and many programs exist to achieve the QZ decomposition with the desired ordering of ratios.

Then, if we partition:

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

where each submatrix Z_{ii} has a size $\tilde{n} \times \tilde{n}$, we can find:

$$P = -Z_{21}^{-1} Z_{22}.$$

If the number of ratios of diagonal elements with absolute value less than 1 (ie, we have enough stable generalized eigenvalues of F with respect to D), then we can select a P such that $P^m x \to \mathbf{0}$ as $m \to \infty$ for any $\tilde{n}$-dimensional vector. If the number of ratios of diagonal elements with absolute value less than 1 is larger than $\tilde{n}$, there may be more than one possible choice of P such that $P^m x \to \mathbf{0}$ as $m \to \infty$ for any $\tilde{n}$-dimensional vector.

The reason a quadratic system appears is that, in general, we will have multiple possible paths for the endogenous variables of the model that would satisfy the equilibrium conditions (Uhlig, 1999 and Galor, 2007). Some of these paths (the stable manifolds) will be stable and satisfy appropriate transversality conditions (although they might imply limit cycles). The other paths (the unstable manifolds) will not. We will need to select the right eigenvalues that induce stability. For many DSGE models, we will have exactly $\tilde{n}$ stable generalized eigenvalues and the stable solution would also be unique. If we have too few stable generalized eigenvalues, the equilibrium dynamics will be inherently unstable. If we have too many, we can have sunspots (Lubik and Schorfheide, 2003). Suffice it to

note here that all these issues would depend only on the first-order approximation and that going to higher-order approximations would not change the issues at hand. If we have uniqueness of equilibrium in the first-order approximation, we will also have uniqueness in the second-order approximation. And if we have multiplicity of equilibria in the first-order approximation, we will also have multiplicity in the second-order approximation.

Remark 12 (Partitioning the quadratic system) The quadratic system (20) can be further divided into two parts to get a recursive solution. The system:

$$[F_x(\bar{\mathbf{x}};0)]^i_j = [\mathcal{H}_{y'}]^i_\alpha [\mathbf{g}_x]^\alpha_\beta [\mathbf{h}_x]^\beta_j + [\mathcal{H}_y]^i_\alpha [\mathbf{g}_x]^\alpha_j + [\mathcal{H}_{x'}]^i_\beta [\mathbf{h}_x]^\beta_j + [\mathcal{H}_x]^i_j = \mathbf{0};$$

$$i=1,\dots,n; \ \ j,\beta=1,\dots,\tilde{n}; \ \ \alpha=1,\dots,n_y. \tag{21}$$

only involves the $\tilde{n} \times n_y$ elements of $\mathbf{g_x}(\bar{\mathbf{x}};0)$ and the $\tilde{n} \times n_x$ elements of $\mathbf{h_x}(\bar{\mathbf{x}};0)$ related to the $\tilde{n}$ endogenous state variables $\mathbf{x}_1$. Once we have solved the $\tilde{n} \times (n_y + n_x)$ unknowns in this system, we can plug them into the system:

$$[F_x(\bar{\mathbf{x}};0)]^i_j = [\mathcal{H}_{y'}]^i_\alpha [\mathbf{g}_x]^\alpha_\beta [\mathbf{h}_x]^\beta_j + [\mathcal{H}_y]^i_\alpha [\mathbf{g}_x]^\alpha_j + [\mathcal{H}_{x'}]^i_\beta [\mathbf{h}_x]^\beta_j + [\mathcal{H}_x]^i_j = \mathbf{0};$$

$$i=1,\dots,n; \ \ j,\beta=\tilde{n}+1,\dots,n_x; \ \ \alpha=1,\dots,n_y. \tag{22}$$

and solve for the $n_\epsilon \times n_y$ elements of $\mathbf{g_x}(\bar{\mathbf{x}};0)$ and the $n_\epsilon \times n_x$ elements of $\mathbf{h_x}((\bar{\mathbf{x}};0)$ related to the n_ϵ stochastic variables $\mathbf{x}_2$.

This recursive solution has three advantages. The first, and most obvious, is that it simplifies computations. The system (20) has $n_x \times (n_y + n_x)$ unknowns, while the system (21) has $\tilde{n} \times (n_y + n_x)$. The difference, $n_\epsilon \times (n_y + n_x)$, makes the second system considerably smaller. Think, for instance, about the medium-scale New Keynesian model in Fernández-Villaverde and Rubio-Ramírez (2008). In the notation of this chapter, the model has $n_x = 20$, $n_y = 1$, and $n_\epsilon = 5$. Thus, by partitioning the system, we go from solving for 420 unknowns to solve a first system of 315 unknowns and, later, a second system of 105 unknowns. The second advantage, which is not obvious in our compact notation, is that system (22) is linear and, therefore, much faster to solve and with a unique solution. In the next subsection, with our worked-out example, we will see this more clearly. The third advantage is that, in some cases, we may only care about the coefficients associated with the $\tilde{n}$ endogenous state variables $\mathbf{x}_1$. This occurs, for example, when we are interested in computing the deterministic transitional path of the model toward a steady state given some initial conditions or when we are plotting impulse response functions generated by the first-order approximation.

The coefficients $\mathbf{g}_\sigma(\bar{\mathbf{x}};0)$ and $\mathbf{h}_\sigma(\bar{\mathbf{x}};0)$ are the solution to the n equations:

$$[F_\sigma(\bar{\mathbf{x}};0)]^i = \mathbb{E}_t\{[\mathcal{H}_{y'}]^i_\alpha [\mathbf{g}_x]^\alpha_\beta [\mathbf{h}_\sigma]^\beta + [\mathcal{H}_{y'}]^i_\alpha [\mathbf{g}_x]^\alpha_\beta [\eta]^\beta_\phi [\epsilon']^\phi + [\mathcal{H}_{y'}]^i_\alpha [\mathbf{g}_\sigma]^\alpha$$
$$+ [\mathcal{H}_y]^i_\alpha [\mathbf{g}_\sigma]^\alpha + [\mathcal{H}_{x'}]^i_\beta [\mathbf{h}_\sigma]^\beta + [\mathcal{H}_{x'}]^i_\beta [\eta]^\beta_\phi [\epsilon']^\phi\}$$
$$i=1,\dots,n; \ \ \alpha=1,\dots,n_y; \ \ \beta=1,\dots,n_x; \ \ \phi=1,\dots,n_\epsilon.$$

Then:

$$[F_\sigma(\bar{\mathbf{x}};0)]^i = [\mathcal{H}_{y'}]^i_\alpha [\mathbf{g}_x]^\alpha_\beta [\mathbf{h}_\sigma]^\beta + [\mathcal{H}_{y'}]^i_\alpha [\mathbf{g}_\sigma]^\alpha + [\mathcal{H}_y]^i_\alpha [\mathbf{g}_\sigma]^\alpha + [f_{x'}]^i_\beta [\mathbf{h}_\sigma]^\beta = 0;$$
$$i = 1,\ldots,n; \quad \alpha = 1,\ldots,n_y; \quad \beta = 1,\ldots,n_x; \quad \phi = 1,\ldots,n_\epsilon.$$

Inspection of the previous equations shows that they are linear and homogeneous equations in $\mathbf{g}_\sigma$ and $\mathbf{h}_\sigma$. Thus, if a unique solution exists, it satisfies:

$$\mathbf{g}_\sigma = \mathbf{0}$$
$$\mathbf{h}_\sigma = \mathbf{0}$$

In other words, the coefficients associated with the perturbation parameter are zero and the first-order approximation is

$$\mathbf{g}(\mathbf{x};\sigma) - \bar{\mathbf{y}} = \mathbf{g}_\mathbf{x}(\bar{\mathbf{x}};0)(\mathbf{x}-\bar{\mathbf{x}})'$$
$$\mathbf{h}(\mathbf{x};\sigma) - \bar{\mathbf{x}} = \mathbf{h}_\mathbf{x}(\bar{\mathbf{x}};0)(\mathbf{x}-\bar{\mathbf{x}})'.$$

These equations embody certainty equivalence as defined by Simon (1956) and Theil (1957). Under certainty equivalence, the solution of the model, up to first-order, is identical to the solution of the same model under perfect foresight (or under the assumption that $\sigma = 0$). Certainty equivalence does not preclude the realization of the shock from appearing in the decision rule. What certainty equivalence precludes is that the standard deviation of it appears as an argument by itself, regardless of the realization of the shock.

The intuition for the presence of certainty equivalence is simple. Risk-aversion depends on the second derivative of the utility function (concave utility). However, Leland (1968) and Sandmo (1970) showed that precautionary behavior depends on the third derivative of the utility function. But a first-order perturbation involves the equilibrium conditions of the model (which includes first derivatives of the utility function, for example, in the Euler equation that equates marginal utilities over time) and first derivatives of these equilibrium conditions (and, therefore, second derivatives of the utility function), but not higher-order derivatives.

Certainty equivalence has several drawbacks. First, it makes it difficult to talk about the welfare effects of uncertainty. Although the dynamics of the model are still partially driven by the variance of the innovations (the realizations of the innovations depend on it), the agents in the model do not take any precautionary behavior to protect themselves from that variance, biasing any welfare computation. Second, related to the first point, the approximated solution generated under certainty equivalence cannot generate any risk premia for assets, a strongly counterfactual prediction.[h] Third, certainty equivalence prevents researchers from analyzing the consequences of changes in volatility.

---

[h] In general equilibrium, there is an intimate link between welfare computations and asset pricing. An exercise on the former is always implicitly an exercise on the latter (see Alvarez and Jermann, 2004).

Remark 13 (Perturbation and LQ approximations) Kydland and Prescott (1982)—and many papers after them—took a different route to solving DSGE models. Imagine that we have an optimal control problem that depends on n_x states $\mathbf{x}_t$ and n_u control variables $\mathbf{u}_t$. To save on notation, let us also define the column vector $\mathbf{w}_t = [\mathbf{x}_t, \mathbf{u}_t]'$ of dimension $n_w = n_x + n_u$. Then, we can write the optimal control problem as:

$$\max \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t r(\mathbf{w}_t)$$
$$\text{s.t. } \mathbf{x}_{t+1} = A(\mathbf{w}_t, \varepsilon_t)$$

where r is a return function, ε_t a vector of n_ε innovations with zero mean and finite variance, and A summarizes all the constraints and laws of motion of the economy. By appropriately enlarging the state space, this notation can accommodate the innovations having an impact on the period return function and some variables being both controls and states.

In the case where the return function r is quadratic, ie,

$$r(\mathbf{w}_t) = B_0 + B_1 \mathbf{w}_t + \mathbf{w}_t' Q \mathbf{w}_t$$

(where B_0 is a constant, B_1 a row vector $1 \times n_w$, and B_2 is an $n_w \times n_w$ matrix) and the function A is linear:

$$\mathbf{x}_{t+1} = B_3 \mathbf{w}_t + B_4 \varepsilon_t$$

(where B_3 is an $n_x \times n_w$ matrix and B_4 is an $n_x \times n_\varepsilon$ matrix), we are facing a stochastic discounted linear-quadratic regulator (LQR) problem. There is a large and well-developed research area on LQR problems. This literature is summarized by Anderson et al. (1996) and Hansen and Sargent (2013). In particular, we know that the optimal decision rule in this environment is a linear function of the states and the innovations:

$$\mathbf{u}_t = F_w \mathbf{w}_t + F_\varepsilon \varepsilon_t$$

where F_w can be found by solving a Ricatti equation Anderson et al. (1996, pp. 182–183) and F_ε by solving a Sylvester equation Anderson et al. (1996, pp. 202–205). Interestingly, F_w is independent of the variance of ε_t. That is, if ε_t has a zero variance, then the optimal decision rule is simply:

$$\mathbf{u}_t = F_w \mathbf{w}_t.$$

This neat separation between the computation of F_w and of F_ε allows the researcher to deal with large problems with ease. However, it also implies certainty equivalence.

Kydland and Prescott (1982) setup the social planner's problem of their economy, which fits into an optimal regulator problem, and they were able to write a function A that was linear in $\mathbf{w}_t$, but they did not have a quadratic return function. Instead, they took a quadratic approximation to the objective function of the social planner. Most of

the literature that followed them used a Taylor series approximation of the objective function around the deterministic steady state, sometimes called the approximated LQR problem (Kydland and Prescott also employed a slightly different point of approximation that attempted to control for uncertainty; this did not make much quantitative difference). Furthermore, Kydland and Prescott worked with the value function representation of the problem. See Díaz-Giménez (1999) for an explanation of how to deal with the LQ approximation to the value function.

The result of solving the approximated LQR when the function A is linear is equivalent to the result of a first-order perturbation of the equilibrium conditions of the model. The intuition is simple. Derivatives are unique, and since both approaches search for a linear approximation to the solution of the model, they have to yield identical results.

However, approximated LQR have lost their popularity for three reasons. First, it is often hard to write the function A in a linear form. Second, it is challenging to set up a social planner's problem when the economy is not Pareto efficient. And even when it is possible to have a modified social planner's problem that incorporates additional constraints that incorporate non-optimalities (see, for instance, Benigno and Woodford, 2004), the same task is usually easier to accomplish by perturbing the equilibrium conditions of the model. Third, and perhaps most important, perturbations can easily go to higher-order terms and incorporate nonlinearities that break certainty equivalence.

4.2.5 Second-Order Perturbation

Once we have finished the first-order perturbation, we can iterate on the steps before to generate higher-order solutions. More concretely, the second-order approximations to $\mathbf{g}$ around $(\mathbf{x}; \sigma) = (\bar{\mathbf{x}}; 0)$ are:

$$[\mathbf{g}(\mathbf{x}; \sigma)]^i = [\mathbf{g}(\bar{\mathbf{x}}; 0)]^i + [\mathbf{g}_x(\bar{\mathbf{x}}; 0)]^i_a[(\mathbf{x} - \mathbf{x})]_a + [\mathbf{g}_\sigma(\bar{\mathbf{x}}; 0)]^i[\sigma]$$

$$+ \frac{1}{2}[\mathbf{g}_{xx}(\bar{\mathbf{x}}; 0)]^i_{ab}[(\mathbf{x} - \bar{\mathbf{x}})]_a[(\mathbf{x} - \bar{\mathbf{x}})]_b$$

$$+ \frac{1}{2}[\mathbf{g}_{x\sigma}(\bar{\mathbf{x}}; 0)]^i_a[(\mathbf{x} - \bar{\mathbf{x}})]_a[\sigma]$$

$$+ \frac{1}{2}[\mathbf{g}_{\sigma x}(\bar{\mathbf{x}}; 0)]^i_a[(\mathbf{x} - \bar{\mathbf{x}})]_a[\sigma]$$

$$+ \frac{1}{2}[\mathbf{g}_{\sigma\sigma}(\bar{\mathbf{x}}; 0)]^i[\sigma][\sigma]$$

where $i = 1, \ldots, n_y$, $a, b = 1, \ldots, n_x$, and $j = 1, \ldots, n_x$.

Similarly, the second-order approximations to $\mathbf{h}$ around $(\mathbf{x}; \sigma) = (\bar{\mathbf{x}}; 0)$ are:

$$[\mathbf{h}(\mathbf{x};\sigma)]^j = [\mathbf{h}(\bar{\mathbf{x}};0)]^j + [\mathbf{h}_x(\bar{\mathbf{x}};0)]^j_a[(\mathbf{x}-\mathbf{x})]_a + [\mathbf{h}_\sigma(\bar{\mathbf{x}};0)]^j[\sigma]$$

$$+ \frac{1}{2}[\mathbf{h}_{xx}(\bar{\mathbf{x}};0)]^j_{ab}[(\mathbf{x}-\bar{\mathbf{x}})]_a[(\mathbf{x}-\bar{\mathbf{x}})]_b$$

$$+ \frac{1}{2}[\mathbf{h}_{x\sigma}(\bar{\mathbf{x}};0)]^j_a[(\mathbf{x}-\bar{\mathbf{x}})]_a[\sigma]$$

$$+ \frac{1}{2}[\mathbf{h}_{\sigma x}(\bar{\mathbf{x}};0)]^j_a[(\mathbf{x}-\bar{\mathbf{x}})]_a[\sigma]$$

$$+ \frac{1}{2}[\mathbf{h}_{\sigma\sigma}(\bar{\mathbf{x}};0)]^j[\sigma][\sigma],$$

where $i=1,\ldots,n_y$, $a,b=1,\ldots,n_x$, and $j=1,\ldots,n_x$.

The unknown coefficients in these approximations are $[\mathbf{g}_{xx}]^i_{ab}$, $[\mathbf{g}_{x\sigma}]^i_a$, $[\mathbf{g}_{\sigma x}]^i_a$, $[\mathbf{g}_{\sigma\sigma}]^i$, $[\mathbf{h}_{xx}]^j_{ab}$, $[\mathbf{h}_{x\sigma}]^j_a$, $[\mathbf{h}_{\sigma x}]^j_a$, $[\mathbf{h}_{\sigma\sigma}]^j$. As before, we solve for these coefficients by taking the second derivatives of $F(\mathbf{x};\sigma)$ with respect to x and σ, making them equal to zero, and evaluating them at $(\bar{\mathbf{x}};0)$.

How do we solve the system? First, we exploit $F_{xx}(\bar{\mathbf{x}};0)$ to solve for $\mathbf{g}_{xx}(\bar{\mathbf{x}};0)$ and $h_{xx}(\bar{\mathbf{x}};0)$:

$$[F_{xx}(\bar{\mathbf{x}};0)]^i_{jk} =$$

$$\left([\mathcal{H}_{y'y'}]^i_{\alpha\gamma}[\mathbf{g}_x]^\gamma_\delta[\mathbf{h}_x]^\delta_k + [\mathcal{H}_{y'y}]^i_{\alpha\gamma}[\mathbf{g}_x]^\gamma_k + [\mathcal{H}_{y'x'}]^i_{\alpha\delta}[\mathbf{h}_x]^\delta_k + [\mathcal{H}_{y'x}]^i_{\alpha k}\right)[\mathbf{g}_x]^\alpha_\beta[\mathbf{h}_x]^\beta_j$$

$$+[\mathcal{H}_{y'}]^i_\alpha[\mathbf{g}_{xx}]^\alpha_{\beta\delta}[\mathbf{h}_x]^\delta_k[\mathbf{h}_x]^\beta_j + [\mathcal{H}_{y'}]^i_\alpha[\mathbf{g}_x]^\alpha_\beta[\mathbf{h}_{xx}]^\beta_{jk}$$

$$+\left([\mathcal{H}_{yy'}]^i_{\alpha\gamma}[\mathbf{g}_x]^\gamma_\delta[\mathbf{h}_x]^\delta_k + [\mathcal{H}_{yy}]^i_{\alpha\gamma}[\mathbf{g}_x]^\gamma_k + [\mathcal{H}_{yx'}]^i_{\alpha\delta}[\mathbf{h}_x]^\delta_k + [\mathcal{H}_{yx}]^i_{\alpha k}\right)[\mathbf{g}_x]^\alpha_j + [\mathcal{H}_y]^i_\alpha[\mathbf{g}_{xx}]^\alpha_{jk}$$

$$+\left([\mathcal{H}_{x'y'}]^i_{\beta\gamma}[\mathbf{g}_x]^\gamma_\delta[\mathbf{h}_x]^\delta_k + [\mathcal{H}_{x'y}]^i_{\beta\gamma}[\mathbf{g}_x]^\gamma_k + [\mathcal{H}_{x'x'}]^i_{\beta\delta}[\mathbf{h}_x]^\delta_k + [\mathcal{H}_{x'x}]^i_{\beta k}\right)[\mathbf{h}_x]^\beta_j + [\mathcal{H}_{x'}]^i_\beta[\mathbf{h}_{xx}]^\beta_{jk}$$

$$+[\mathcal{H}_{xy'}]^i_{j\gamma}[\mathbf{g}_x]^\gamma_\delta[\mathbf{h}_x]^\delta_k + [\mathcal{H}_{xy}]^i_{j\gamma}[\mathbf{g}_x]^\gamma_k + [\mathcal{H}_{xx'}]^i_{j\delta}[\mathbf{h}_x]^\delta_k + [\mathcal{H}_{xx}]^i_{jk} = 0;$$

$$i=1,\ldots n, \quad j,k,\beta,\delta=1,\ldots n_x; \quad \alpha,\gamma=1,\ldots n_y.$$

But we know the derivatives of $\mathcal{H}$. We also know the first derivatives of $\mathbf{g}$ and $\mathbf{h}$ evaluated at $(\bar{\mathbf{x}}, 0)$. Hence, the above expression is a system of $n \times n_x \times n_x$ linear equations in the $n \times n_x \times n_x$ unknown elements of $\mathbf{g}_{xx}$ and $\mathbf{h}_{xx}$. This point is crucial: linear solvers are fast and efficient. In the first-order approximation we had to solve a quadratic system to select between stable and unstable solutions. But once we are already in the stable manifold, there are no further additional solutions that we need to rule out. These quadratic terms involve the endogenous state vector x_1. Those terms capture nonlinear behavior and induce nonsymmetries. We will discuss those in more detail in our worked-out example below.

The coefficients in $\mathbf{g}_{\sigma\sigma}$ and $\mathbf{h}_{\sigma\sigma}$ come from solving the system of n linear equations in the n unknowns:

$$
\begin{aligned}
[F_{\sigma\sigma}(\bar{\mathbf{x}};0)]^i = {}& [\mathcal{H}_{y'}]^i_\alpha [\mathbf{g}_x]^\alpha_\beta [\mathbf{h}_{\sigma\sigma}]^\beta \\
& + [\mathcal{H}_{y'y'}]^i_{\alpha\gamma}[\mathbf{g}_x]^\gamma_\delta[\eta]^\delta_\xi[\mathbf{g}_x]^\alpha_\beta[\eta]^\beta_\phi[I]^\phi_\xi + [\mathcal{H}_{y'x'}]^i_{\alpha\delta}[\eta]^\delta_\xi[\mathbf{g}_x]^\alpha_\beta[\eta]^\beta_\phi[I]^\phi_\xi \\
& + [\mathcal{H}_{y'}]^i_\alpha[\mathbf{g}_{xx}]^\alpha_{\beta\delta}[\eta]^\delta_\xi[\eta]^\beta_\phi[I]^\phi_\xi + [\mathcal{H}_{y'}]^i_\alpha[\mathbf{g}_{\sigma\sigma}]^\alpha \\
& + [\mathcal{H}_y]^i_\alpha[\mathbf{g}_{\sigma\sigma}]^\alpha + [\mathcal{H}_{x'}]^i_\beta[\mathbf{h}_{\sigma\sigma}]^\beta \\
& + [\mathcal{H}_{x'y'}]^i_{\beta\gamma}[\mathbf{g}_x]^\gamma_\delta[\eta]^\delta_\xi[\eta]^\beta_\phi[I]^\phi_\xi + [\mathcal{H}_{x'x'}]^i_{\beta\delta}[\eta]^\delta_\xi[\eta]^\beta_\phi[I]^\phi_\xi = 0; \\
& = 1,\ldots,n; \, \alpha,\gamma = 1,\ldots,n_y; \, \beta,\, \delta = 1,\ldots,n_x; \, \phi,\, \xi = 1,\ldots,n_\epsilon.
\end{aligned}
$$

The coefficients $\mathbf{g}_{\sigma\sigma}$ and $\mathbf{h}_{\sigma\sigma}$ capture the correction for risk that breaks certainty equivalence. In addition, the cross derivatives $\mathbf{g}_{x\sigma}$ and $\mathbf{h}_{x\sigma}$ are zero when evaluated at $(\bar{\mathbf{x}};0)$. To see this, write the system $F_{\sigma x}(\bar{\mathbf{x}};0) = 0$, taking into account that all terms containing either g_σ or $\mathbf{h}_\sigma$ are zero at $(\bar{\mathbf{x}};0)$. Then, we have a homogeneous system of $n \times n_x$ equations in the $n \times n_x$ elements of $\mathbf{g}_{\sigma x}$ and $\mathbf{h}_{\sigma x}$:

$$
\begin{aligned}
[F_{\sigma x}(\bar{\mathbf{x}};0)]^i_j = {}& [\mathcal{H}_{y'}]^i_\alpha[\mathbf{g}_x]^\alpha_\beta[\mathbf{h}_{\sigma x}]^\beta_j + [\mathcal{H}_{y'}]^i_\alpha[\mathbf{g}_{\sigma x}]^\alpha_\gamma[\mathbf{h}_x]^\gamma_j + [\mathcal{H}_y]^i_\alpha[\mathbf{g}_{\sigma x}]^\alpha_j + [\mathcal{H}_{x'}]^i_\beta[\mathbf{h}_{\sigma x}]^\beta_j = 0; \\
& i = 1,\ldots n; \quad \alpha = 1,\ldots,n_y; \quad \beta,\gamma,j = 1,\ldots,n_x.
\end{aligned}
$$

Hence, the last component of the second–order perturbation is given by:[i]

$$
\mathbf{g}_{\sigma x} = \mathbf{0}
$$
$$
\mathbf{h}_{\sigma x} = \mathbf{0}.
$$

4.2.6 Higher-Order Perturbations

We can iterate the previous procedure (taking higher-order derivatives, plugging in the already found terms, and solving for the remaining ones) as many times as we want to obtain n-th order approximations. All the associated systems of equations that we would need to solve are linear, which keeps the computational complexity manageable. The only additional point to remember is that we will need to make assumptions about the higher moments of the innovations, as we will have expectational terms involving these higher moments.

If the functions $\mathbf{g}$ and $\mathbf{h}$ are analytic in a neighborhood of $\bar{\mathbf{x}}$, then the series we are building by taking higher-order approximations has an infinite number of terms and is convergent. Convergence will occur in a radius of convergence centered around $\bar{\mathbf{x}}$, (ie, the r such that for all state values with a distance with respect to $\bar{\mathbf{x}}$ smaller then r). This radius can be infinite. In that case, the series is guaranteed to converge uniformly everywhere. However, the radius can also be finite and there exist a nonremovable

---

[i] We conjecture (and we have checked up to as high an order of a perturbation as computer memory allows) that all terms involving odd derivatives of σ are zero. Unfortunately, we do not have a formal proof.

singularity on its boundary. Disappointingly, for most DSGE models, the radius of convergence is unknown (for more details and examples, see Swanson et al., 2006 and Aldrich and Kung, 2011). More research on this topic is sorely needed. Also, even when the series is convergent, there are two potential problems. First, at a j-th order approximation, we may lose the "right" shape of **g** and **h**. For example, Aruoba et al. (2006) document how the decision rules for consumption and capital of the stochastic neoclassical growth model approximated with a fifth-order perturbation are no longer globally concave, as implied by economic theory. Instead, the approximated functions present oscillating patterns. Second, the convergence to the exact solution may not be monotone: it is easy to build examples where the errors a bit away from $\bar{\mathbf{x}}$ are worse for a $j + 1$-th order approximation than for a j-th order approximation. Neither of these two problems is fatal, but the researcher needs to be aware of them and undertake the necessary tests to minimize their impact (for instance, checking the solution for different approximation orders).

Later, we will discuss how to gauge the accuracy of a solution and how to decide whether a higher-order approximation is required. For example, to deal with models with time-varying volatility, we would need at least a third-order approximation. Levintal (2015a) has argued that to approximate well models with disaster risk, we need a fifth-order approximation. The drawback of higher-order approximations is that we will run into problems of computational cost and memory use.

4.3 A Worked-Out Example

The previous derivations were somewhat abstract and the notation, even using tensors, burdensome. Consequently, it is useful to show how perturbation works in a concrete example. For that, we come back to our example of the neoclassical growth model defined by Eqs. (2)–(4), except that, to make the algebra easier, we assume $u(c) = \log c$ and $\delta = 1$.

The equilibrium conditions of the model are then:

$$\frac{1}{c_t} = \beta \mathbb{E}_t \frac{\alpha e^{z_{t+1}} k_{t+1}^{\alpha-1}}{c_{t+1}}$$

$$c_t + k_{t+1} = e^{z_t} k_t^\alpha$$

$$z_t = \rho z_{t-1} + \eta \varepsilon_t$$

While this parameterization is unrealistic for periods of time such as a quarter or a year typically employed in business cycle analysis, it has the enormous advantage of implying that the model has a closed-form solution. With $\delta = 1$, the income and the substitution effect from a productivity shock cancel each other, and consumption and investment are constant fractions of income:

$$c_t = (1 - \alpha\beta) e^{z_t} k_t^\alpha$$

$$k_{t+1} = \alpha\beta e^{z_t} k_t^\alpha$$

(these optimal decision rules can be verified by plugging them into the equilibrium conditions and checking that indeed these conditions are satisfied).

Imagine, however, that we do not know this exact solution and that we are searching a decision rule for consumption:

$$c_t = c(k_t, z_t)$$

and another one for capital:

$$k_{t+1} = k(k_t, z_t)$$

In our general notation, d would just be the stack of $c(k_t, z_t)$ and $k(k_t, z_t)$. We substitute these decision rules in the equilibrium conditions above (and, to reduce the dimensionality of the problem, we substitute out the budget constraint and the law of motion for technology) to get:

$$\frac{1}{c(k_t, z_t)} = \beta \mathbb{E}_t \frac{\alpha e^{\rho z_t + \sigma \varepsilon_{t+1}} k(k_t, z_t)^{\alpha-1}}{c(k(k_t, z_t), \rho z_t + \eta \varepsilon_{t+1})} \tag{23}$$

$$c(k_t, z_t) + k(k_t, z_t) = e^{z_t} k_t^{\alpha} \tag{24}$$

The decision rules are approximated by perturbation solutions on the two state variables plus the perturbation parameter σ:

$$c_t = c(k_t, z_t; \sigma)$$
$$k_{t+1} = k(k_t, z_t; \sigma).$$

We introduce σ in the law of motion for technology:

$$z_t = \rho z_{t-1} + \sigma \eta \varepsilon_t.$$

In that way, if we set $\sigma = 0$, we recover a deterministic model. If $z_t = 0$ (either because $z_0 = 0$ or because t is sufficiently large such that $z_t \to 0$), we can find the steady state k by solving the system of equilibrium conditions:

$$\frac{1}{c} = \beta \frac{\alpha k^{\alpha-1}}{c}$$
$$c + k = k^{\alpha}$$

which has a unique solution $k = k(k, 0; 0) = (\alpha\beta)^{\frac{1}{1-\alpha}}$ and $c = c(k, 0; 0) = (\alpha\beta)^{\frac{\alpha}{1-\alpha}} - (\alpha\beta)^{\frac{1}{1-\alpha}}$.

The second-order expansion for the consumption decision rule is given by:

$$c_t = c + c_k(k_t - k) + c_z z_t + c_\sigma \sigma$$
$$+ \frac{1}{2} c_{kk}(k_t - k)^2 + c_{kz}(k_t - k) z_t + c_{k\sigma}(k_t - k)\sigma \tag{25}$$
$$+ \frac{1}{2} c_{zz} z_t^2 + c_{z\sigma} z_t \sigma + \frac{1}{2} c_{\sigma^2} \sigma^2$$

and for the capital decision rule:

$$k_{t+1} = k + k_k(k_t - k) + k_z z_t + k_\sigma \sigma$$

$$+ \frac{1}{2}k_{kk}(k_t - k)^2 + k_{kz}(k_t - k)z_t + k_{k\sigma}(k_t - k)\sigma \tag{26}$$

$$+ \frac{1}{2}k_{zz}z_t^2 + \frac{1}{2}k_{\sigma z}\sigma z_t + \frac{1}{2}k_{\sigma^2}\sigma^2$$

(where we have already used the symmetry of second derivatives and assumed that all terms are evaluated at $(k,0;0)$). Higher-order approximations can be written in a similar way, but, for this example, a second-order approximation is all we need.

Beyond the correction for risk $\frac{1}{2}c_{\sigma^2}\sigma^2$ and $\frac{1}{2}k_{\sigma^2}\sigma^2$ that we discussed above, the additional terms in Eqs. (25) and (26) introduce dynamics that cannot be captured by a first-order perturbation. In the linear solution, the terms $c_z z_t$ and $k_\sigma \sigma$ imply that the effects of positive and negative shocks are mirrors of each other. That is why, for instance, researchers using linearized models only report impulse response functions to a positive or a negative shock: the other impulse response functions are the same but inverted. In comparison, in the second-order perturbation, the terms $\frac{1}{2}c_{zz}z_t^2$ and $\frac{1}{2}k_{zz}z_t^2$ mean that positive and negative shocks have divergent effects: z_t^2 is always positive and the impulse response functions are asymmetric. The terms $c_{kz}(k_t - k)z_t$ and $k_{kz}(k_t - k)z_t$ cause the effect of a shock to also depend on how much capital the economy has at period t, a mechanism missed in the first-order approximation since z_t enters linearly. This might be of importance in many applications. For example, the effects of a financial shock may depend on the household asset level.

To find the unknown coefficients in Eqs. (25) and (26), we come back to the equilibrium conditions (23) and (24), we substitute the decision rules with the approximated decision rules $c(k_t, z_t; \sigma)$ and $k(k_t, z_t; \sigma)$, and we rearrange terms to get:

$$F(k_t, z_t; \sigma) = \mathbb{E}_t \begin{bmatrix} \frac{1}{c(k_t, z_t; \sigma)} - \beta \frac{\alpha e^{\rho z_t + \sigma \eta \varepsilon_{t+1}} k(k_t, z_t; \sigma)^{\alpha-1}}{c(k(k_t, z_t; \sigma), \rho z_t + \sigma \eta \varepsilon_{t+1}; \sigma)} \\ c(k_t, z_t; \sigma) + k(k_t, z_t; \sigma) - e^{z_t} k_t^\alpha \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \end{bmatrix}$$

More compactly:

$$F(k_t, z_t; \sigma) = \mathcal{H}(c(k_t, z_t; \sigma), c(k(k_t, z_t; \sigma), z_{t+1}; \sigma), k_t, k(k_t, z_t; \sigma), z_t; \sigma)$$

We will use $\mathcal{H}_i$ to represent the partial derivative of $\mathcal{H}$ with respect to the i component and drop the evaluation at the steady state of the functions when we do not need it.

We start with the first-order terms. We take derivatives of $F(k_t, z_t; \sigma)$ with respect to k_t, z_t, and σ and we equate them to zero:

$$F_k = \mathcal{H}_1 c_k + \mathcal{H}_2 c_k k_k + \mathcal{H}_3 + \mathcal{H}_4 k_k = \mathbf{0}$$
$$F_z = \mathcal{H}_1 c_z + \mathcal{H}_2 (c_k k_z + c_k \rho) + \mathcal{H}_4 k_z + \mathcal{H}_5 = \mathbf{0}$$
$$F_\sigma = \mathcal{H}_1 c_\sigma + \mathcal{H}_2 (c_k k_\sigma + c_\sigma) + \mathcal{H}_4 k_\sigma + \mathcal{H}_6 = \mathbf{0}$$

Note that:

$$F_k = \mathcal{H}_1 c_k + \mathcal{H}_2 c_k k_k + \mathcal{H}_3 + \mathcal{H}_4 k_k = \mathbf{0}$$
$$F_z = \mathcal{H}_1 c_z + \mathcal{H}_2 (c_k k_z + c_k \rho) + \mathcal{H}_4 k_z + \mathcal{H}_5 = \mathbf{0}$$

is a quadratic system of four equations in four unknowns: c_k, c_z, k_k, and k_z (the operator F has two dimensions). As we mentioned above, the system can be solved recursively. The first two equations:

$$F_k = \mathcal{H}_1 c_k + \mathcal{H}_2 c_k k_k + \mathcal{H}_3 + \mathcal{H}_4 k_k = \mathbf{0}$$

only involve c_k and k_k (the terms affecting the deterministic variables).

Remark 14 (Quadratic problem, again) The first two equations:

$$F_k = \mathcal{H}_1 c_k + \mathcal{H}_2 c_k k_k + \mathcal{H}_3 + \mathcal{H}_4 k_k = \mathbf{0}$$

can easily be written in the form of a quadratic matrix system as follows. First, we write the two equations as:

$$\begin{pmatrix} \mathcal{H}_1^1 \\ \mathcal{H}_1^2 \end{pmatrix} c_k + \begin{pmatrix} \mathcal{H}_2^1 \\ \mathcal{H}_2^2 \end{pmatrix} c_k k_k + \begin{pmatrix} \mathcal{H}_3^1 \\ \mathcal{H}_3^2 \end{pmatrix} + \begin{pmatrix} \mathcal{H}_4^1 \\ \mathcal{H}_4^2 \end{pmatrix} k_k = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

where $\mathcal{H}_i^j$ is the j-th dimension of $\mathcal{H}_i$. But $\mathcal{H}_2^2 = 0$ and $\mathcal{H}_3^1 = 0$, then

$$\begin{pmatrix} \mathcal{H}_1^1 \\ \mathcal{H}_1^2 \end{pmatrix} c_k + \begin{pmatrix} \mathcal{H}_2^1 \\ 0 \end{pmatrix} c_k k_k + \begin{pmatrix} 0 \\ \mathcal{H}_3^2 \end{pmatrix} + \begin{pmatrix} \mathcal{H}_4^1 \\ \mathcal{H}_4^2 \end{pmatrix} k_k = \begin{pmatrix} 0 \\ 0 \end{pmatrix}.$$

We can use the second equation to eliminate c_k from the first equation. Then, rearranging the terms and calling $P = k_k$ we have the equation:

$$AP^2 - BP - C = 0$$

that we presented in the previous subsection. Note that, in this example, instead of a complicated matrix equation, we have a much simpler quadratic scalar equation.

Our quadratic system will have two solutions. One solution will imply that $k_k > 1$ and the other solution $k_k < 1$. The first solution is unstable. Remember that the first elements of the decision rule are

$$k_{t+1} = k + k_k (k_t - k) + \dots$$

If $k_k > 1$, a deviation of k_t with respect to k will imply an even bigger deviation of k_{t+1} with respect to k, leading to explosive behavior. In comparison, when $k_k < 1$, deviations

of k_t with respect to k will, in the absence of additional shocks, dissipate over time. Once we know c_k and k_k, we can come back to

$$F_z = \mathcal{H}_1 c_z + \mathcal{H}_2(c_k k_z + c_k \rho) + \mathcal{H}_4 k_z + \mathcal{H}_5 = 0$$

and solve for c_z and k_z. As emphasized in Remark 12, this system is linear.

Finally, as in the general case, the last two equations

$$F_\sigma = \mathcal{H}_1 c_\sigma + \mathcal{H}_2(c_k k_\sigma + c_\sigma) + \mathcal{H}_4 k_\sigma + \mathcal{H}_6 = 0$$

form a linear, and homogeneous system in c_σ and k_σ. Hence, $c_\sigma = k_\sigma = 0$ and we obtain the certainty equivalence of first-order approximations.

To find the second-order approximation, we take second derivatives of $F(k_t, z_t; \sigma)$ around k, 0, and 0:

$$F_{kk} = 0$$
$$F_{kz} = 0$$
$$F_{k\sigma} = 0$$
$$F_{zz} = 0$$
$$F_{z\sigma} = 0$$
$$F_{\sigma\sigma} = 0$$

(where we have already eliminated symmetric second derivatives). We substitute the coefficients that we already know from the first-order approximation and we get a linear system of 12 equations in 12 unknowns. Again, we get that all cross-terms on $k\sigma$ and $z\sigma$ are zero.

Imposing the results concerning the coefficients that are equal to zero, we can rewrite Eqs. (25) and (26) up to second-order as:

$$c_t = c + c_k(k_t - k) + c_z z_t$$
$$+ \frac{1}{2}c_{kk}(k_t - k)^2 + c_{kz}(k_t - k)z_t + \frac{1}{2}c_{zz}z_t^2 + \frac{1}{2}c_{\sigma^2}\sigma^2 \tag{27}$$

and

$$k_{t+1} = k + k_k(k_t - k) + k_z z_t$$
$$+ \frac{1}{2}k_{kk}(k_t - k)^2 + k_{kz}(k_t - k)z_t + \frac{1}{2}k_{zz}z_t^2 + \frac{1}{2}k_{\sigma^2}\sigma^2. \tag{28}$$

Since even with this simple neoclassical growth model the previous systems of equations are too involved to be written explicitly, we illustrate the procedure numerically. In Table 1, we summarize the parameter values for the four parameters of the model. We do not pretend to be selecting a realistic calibration (our choice of $\delta = 1$ precludes any attempt at matching observed data). Instead, we pick standard parameter values in the

Table 1 Calibration

Parameter	Value
β	0.99
α	0.33
ρ	0.95
η	0.01

literature. The discount factor, β, is 0.99, the elasticity of output with respect to capital, α, is 0.33, the persistence of the autoregressive process, ρ, is 0.95, and the standard deviation of the innovation, η, is 0.01. With this calibration, the steady state is $c = 0.388$ and $k = 0.188$.

The first-order components of the solution are (already selecting the stable solution):

$$c_k = 0.680 \quad c_z = 0.388$$
$$k_k = 0.330 \quad k_z = 0.188$$

and the second-order components:

$$c_{kk} = -2.420 \quad c_{kz} = 0.680 \quad c_{zz} = 0.388 \quad c_{\sigma\sigma} = 0$$
$$k_{kk} = -1.174 \quad k_{kz} = 0.330 \quad k_{zz} = 0.188 \quad k_{\sigma\sigma} = 0$$

In addition, recall that we have the theoretical results: $c_\sigma = k_\sigma = c_{k\sigma} = k_{k\sigma} = c_{z\sigma} = k_{z\sigma} = 0$. Thus, we get our second-order approximated solutions for the consumption decision rule:

$$c_t = 0.388 + 0.680(k_t - 0.188) + 0.388 z_t$$
$$- 1.210(k_t - 0.188)^2 + 0.680(k_t - 0.188)z_t + 0.194 z_t^2$$

and for the capital decision rule:

$$k_{t+1} = 0.188 + 0.330\,(k_t - 0.188) + 0.188 z_t$$
$$- 0.587(k_t - 0.188)^2 + 0.330\,(k_t - 0.188)\,z_t + 0.094 z_t^2.$$

In this case, the correction for risk is zero. This should not be a surprise. In the neoclassical growth model, risk is production risk driven by technology shocks. This production risk is brought about by capital: the more capital the representative household accumulates, the more it exposes itself to production risk. At the same time, the only asset available for net saving in this economy is capital. Thus, any increment in risk (ie, a rise in the standard deviation of the technology shock) generates two counterbalancing mechanisms: a desire to accumulate more capital to buffer future negative shocks and a desire to accumulate less capital to avoid the additional production risk. For low values of risk aversion, both mechanisms nearly cancel each other (with a log utility function, they perfectly compensate each other: in the exact solution, the standard deviation of the innovation to the

shock does not appear, only the realization of z_t). For higher values of risk aversion or for models with different assets (for instance, a model where the representative household can save in the form of an international bond whose payments are not perfectly correlated with the productivity shock within the country), the correction for risk can be quite different from zero.

The next step is to compare the exact and the approximated decision rules. With our calibration, the exact solution is given by:

$$c_t = 0.673 e^{z_t} k_t^{0.33}$$
$$k_{t+1} = 0.327 e^{z_t} k_t^{0.33}.$$

To gauge how close these two solutions are, we plot in Fig. 1 the exact decision rule for capital (continuous line in the top and bottom panels), the first-order approximation (discontinuous line in the top panel), and the second-order approximation (discontinuous line in the bottom panel). In both panels, we plot the decision rule for capital when $z_t = 0$ and for values of capital that are $\pm 25\%$ of the value of capital in the steady state. The first-order approximation is nearly identical to the exact solution close to the steady state. Only farther away, do both solutions diverge. At the start of the grid (with $k = 0.1412$),
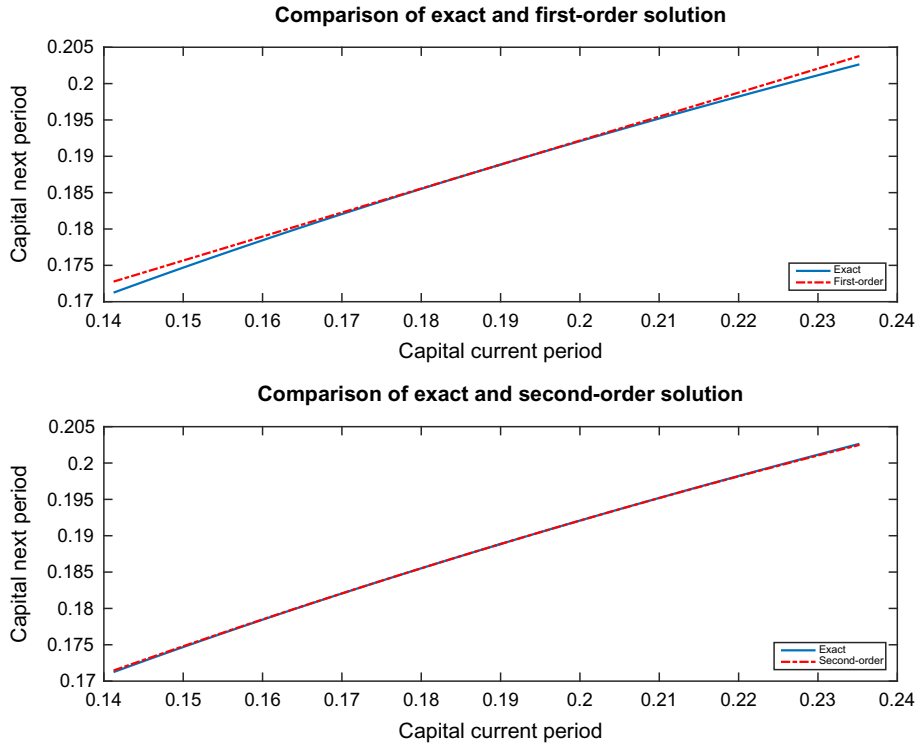
Fig. 1 Comparison of exact and perturbation solution.

the exact decision rule and the first–order approximation diverge by nearly 1%. The second-order approximation, in comparison, is more accurate along the whole range of values of capital. Even at $k = 0.1412$, the difference between both solutions is only 0.13%. This result shows the often good global properties of perturbation solutions.

We will revisit below how to assess the accuracy of a solution. Suffice it to say at this moment that whether 0.13% is too large or accurate enough is application dependent. For instance, for the computation of business cycle moments, we often need less accuracy than for welfare evaluations. The reason is that while errors in the approximation of a moment of the model, such as the mean or variance of consumption, tend to cancel each other, welfare is a nonlinear function of the allocation and small errors in computing an allocation can translate into large errors in computing welfare.

4.4 Pruning

Although the higher-order perturbations that we described are intuitive and straightforward to compute, they often generate explosive sample paths even when the corresponding linear approximation is stable. These explosive sample paths arise because the higher-order terms induce additional fixed points for the system, around which the approximated solution is unstable (see Kim et al., 2008 and Den Haan and De Wind, 2012). A simple example clarifies this point. Imagine that we have an approximated decision rule for capital (where, for simplicity, we have eliminated the persistence on the productivity process z_t) that has the form:

$$k_{t+1} = a_0 + a_1 k_t + a_2 k_t^2 + \ldots + b_1 \varepsilon_t + \ldots$$

If we substitute recursively, we find:

$$k_{t+1} = a_1 k_t + a_2 \left(a_1 k_{t-1} + a_2 k_{t-1}^2 \right)^2 + \ldots + b_1 \varepsilon_t + \ldots,$$

an expression that involves terms in k_{t-1}^3 and k_{t-1}^4. If the support of ε_t is not bounded, sooner or later, we will have, in a simulation, an innovation large enough such that k_{t+1} is far away from its steady-state value. As the simulation progresses over time, that value of k_{t+1} will be raised to cubic and higher-order powers, and trigger an explosive path. The presence of this explosive behavior complicates any model evaluation because no unconditional moments would exist based on this approximation. It also means that any unconditional moment-matching estimation methods, such as the generalized method of moments (GMM) or the simulated method of moments (SMM), are inapplicable in this context as they rely on finite moments from stationary and ergodic probability distributions.

For second-order approximations, Kim et al. (2008) propose pruning the approximation. Loosely speaking, pruning means to eliminate, in the recursions, all terms that are of a higher order than the order of the solution (ie, if we are dealing with a second-order

perturbation, all terms involving states or innovations raised to powers higher than 2). Kim et al. (2008) prove that the pruned approximation does not explode.

Andreasen et al. (2013) extend Kim et al. (2008)'s approach by showing how to apply pruning to an approximation of any arbitrary order by exploiting what the authors refer to as the pruned state-space system. Under general technical conditions, Andreasen et al. (2013) show that first and second unconditional moments for a pruned state-space system exist. Then, they provide closed-form expressions for first and second unconditional moments and impulse response functions. This is important because these expressions let researchers avoid the use of numerical simulations to compute these moments. These numerical simulations have often been shown to be unreliable, in particular, when solving for the generalized impulse response functions of DSGE models (for the definition of generalized impulse response functions, see Koop et al., 1996). Andreasen et al. (2013) also derive conditions for the existence of higher unconditional moments, such as skewness and kurtosis.

4.5 Change of Variables

In Remark 4, we discussed the possibility of performing the perturbation of a DSGE model in logs of the variables of interest, instead of doing it in levels. In a creative contribution, Judd (2003) argues that loglinearization is a particular case of the more general idea of a change of variables and shows how this technique could be efficiently implemented. In this subsection, we explain Judd's contribution by following Fernández-Villaverde and Rubio-Ramírez (2006).

The point of departure is to note that if we have a Taylor expansion of a variable x around a point a:

$$d(x) \simeq d(a) + \frac{\partial d(a)}{\partial a}(x-a) + H.O.T.,$$

(where $H.O.T.$ stands for higher-order terms), we can rewrite the expansion in terms of a transformed variable $Y(x)$:

$$g(y) = h(d(X(y))) = g(b) + \frac{\partial g(b)}{\partial b}(Y(x) - b) + H.O.T.$$

where $b = Y(a)$ and $X(y)$ is the inverse of $Y(x)$. Since with a perturbation we find a Taylor series approximation of the unknown function d that solves the operator $\mathcal{H}(\cdot)$ as a function of the states x, the change of variables means we can find an alternative Taylor series in terms of $Y(x)$.

Why do we want to perform this change of variables? The famous British meteorologist Eric Eady (1915–1966) remarked once that: "It is not the process of linearization that limits insight. It is the nature of the state that we choose to linearize about."

By picking the right change of variables, we can reshape a highly nonlinear problem into a much more linear one and, therefore, significantly increase the accuracy of the perturbation.[j]

4.5.1 A Simple Example

Imagine that our aim is to approximate the decision rule for capital in our workhorse stochastic neoclassical growth model with a first-order perturbation (the same ideas would apply if we are trying to approximate other decision rules, expectations, value functions, etc.). Remember that we derived that such an approximation had the form:

$$k_{t+1} = k + a_1(k_t - k) + b_1 z_t$$

where a and b are the coefficients that we find by taking derivatives of $F(k_t, z_t; \sigma)$ and k is the steady-state value of capital. In this section, it is more convenient to rewrite the decision rule as:

$$(k_{t+1} - k) = a_1(k_t - k) + b_1 z_t.$$

Analogously a loglinear approximation of the policy function will take the form:

$$\log k_{t+1} - \log k = a_2(\log k_t - \log k) + b_2 z$$

or in equivalent notation:

$$\hat{k}_{t+1} = a_2 \hat{k}_t + b_2 z_t$$

where $\hat{x} = \log x - \log x_0$ is the percentage deviation of the variable x with respect to its steady state.

How do we go from one approximation to the second one? First, we write the linear system in levels as:

$$k_{t+1} = d(k_t, z_t; \sigma) = d(k, 0; 0) + d_1(k, 0; 0)(k_t - k) + d_2(k, 0; 0)z_t$$

where $d(k, 0; 0) = k$, $d_1(k, 0; 0) = a_1$, $d_2(k, 0; 0) = b_1$. Second, we propose the changes of variables $h = \log d$, where $Y(x) = \log x$ and $Y(x) = \log x$. Third, we apply Judd (2003)'s formulae for this example:

$$\log k_{t+1} - \log k = d_1(k, 0, 0)(\log k_t - \log k) + \frac{1}{k} d_2(k, 0, 0)z$$

Finally, by equating coefficients, we obtain a simple closed-form relation between the parameters of both representations: $a_2 = a_1$ and $b_2 = \frac{1}{k} b_1$.

---

[j] An idea related to the change of variables is the use of gauges, where the perturbation is undertaken not in terms of powers of the perturbation parameter, σ, but of a series of gauge functions $\{\delta_n(\sigma)\}_{n=1}^{\infty}$ such that: $\lim_{n\to\infty} \frac{\delta_{n+1}(\sigma)}{\delta_n(\sigma)} = 0$. See Judd (1998) for details.

Three points are important. First, moving from a_1 and b_1 to a_2 and b_2 is an operation that only involves k, a value that we already know from the computation of the first-order perturbation in levels. Therefore, once the researcher has access to the linear solution, obtaining the loglinear one is immediate.[k] Second, we have not used any assumption on the utility or production functions except that they satisfy the general technical conditions of the stochastic neoclassical growth model. Third, the change of variables can be applied to a perturbation of an arbitrary order. We only presented the case for a first-order approximation to keep the exposition succinct.

4.5.2 A More General Case

We can now present a more general case of change of variables. The first-order solution of a model is:

$$d(x) \simeq d(a) + \frac{\partial d(a)}{\partial a}(x - a).$$

If we expand $g(y) = h(d(X(y)))$ around $b = Y(a)$, where $X(y)$ is the inverse of $Y(x)$, we can write:

$$g(y) = h(d(X(y))) = g(b) + g_\alpha(b)(Y^\alpha(x) - b^\alpha)$$

where $g_\alpha = h_A d_i^A X_\alpha^i$ comes from the application of the chain rule.

Following Judd (2003), we use this approach to encompass any power function approximation of the form:

$$k_{t+1}(k, z; \gamma, \zeta, \varphi)^\gamma - k^\gamma = a_3\left(k_t^\zeta - k^\zeta\right) + b_3 z^\varphi$$

where we impose $\varphi \geq 1$ to ensure that we have real values for the power z^φ. Power functions are attractive because, with only three free parameters (γ, ζ, φ), we can capture many nonlinear structures and nest the log transformation as the limit case when the coefficients γ and ζ tend to zero and $\varphi = 1$. The changes of variables for this family of functions are given by $h = d^\gamma$, $Y = x^\zeta$, and $X = y^{\frac{1}{\zeta}}$. Following the same reasoning as before, we derive:

$$k_{t+1}(k, z; \gamma, \zeta, \varphi)^\gamma - k^\gamma = \frac{\gamma}{\zeta}k^{\gamma-\zeta}a_1\left(k_t^\zeta - k^\zeta\right) + \frac{\gamma}{\varphi}k^{\gamma-1}b_1 z^\varphi.$$

The relation between the new and the old coefficients is again easy to compute: $a_3 = \frac{\gamma}{\zeta}k^{\gamma-\zeta}a_1$ and $b_3 = \frac{\gamma}{\varphi}k^{\gamma-1}b_1.$

---

[k] A heuristic argument that delivers the same result takes: $(k_{t+1} - k) = a_1(k_t - k) + b_1 z_t$ and divides on both sides by k: $\frac{k_{t+1} - k}{k} = a_1\frac{k_t - k}{k} + \frac{1}{k}b_1 z$. Noticing that $\frac{x_t - x}{x} \simeq \log x_t - \log x$, we get back the same relation as the one above. Our argument in the main text is more general and does not depend on an additional approximation.

A slightly more restrictive case is to impose that $\gamma = \zeta$ and $z = 1$. Then, we get a power function with only one free parameter γ:

$$k_{t+1}(k, z; \gamma)^{\gamma} - k^{\gamma} = a_4\left(k_t^{\zeta} - k^{\zeta}\right) + b_4 z$$

or, by defining $\tilde{k}_t = k_t^{\gamma} - k^{\gamma}$, we get:

$$\tilde{k}_{t+1} = a_4 \tilde{k}_t + b_4 z$$

with $a_4 = a_1$ and $b_4 = k^{\gamma - 1} b_1$. This representation has the enormous advantage of being a linear system, which makes it suitable for analytic study and, as we will see in Section 10, for estimation with a Kalman filter.

4.5.3 The Optimal Change of Variables

The previous subsection showed how to go from a first-order approximation to the solution of a DSGE model to a more general representation indexed by some parameters. The remaining question is how to select the optimal value of these parameters.[1]

Fernández-Villaverde and Rubio-Ramírez (2006) argue that a reasonable criterion (and part of the motivation for the change of variables) is to select these parameters to improve the accuracy of the solution of the model. More concretely, the authors propose to minimize the Euler error function with respect to some metric. Since we have not introduced the measures of accuracy of the solution to a DSGE model, we will skip the details of how to do so. Suffice it to say that Fernández-Villaverde and Rubio-Ramírez (2006) find that the optimal change of variables improves the average accuracy of the solution by a factor of around three. This improvement makes a first-order approximation competitive in terms of accuracy with much more involved methods. Fernández-Villaverde and Rubio-Ramírez (2006) also report that the optimal parameter values depend on the standard deviation of the exogenous shocks to the economy. This is a significant result: the change of variables corrects by the level of uncertainty existing in the economy and breaks certainty equivalence.

Remark 15 (Loglinearization v. lognormal–loglinear approximation) A different solution technique, called lognormal–loglinear approximation, is popular in finance. Its relation with standard loglinearization (as a particular case of first-order perturbation with a change of variables in logs) often causes confusion among researchers and students. Thus, once we have understood the change of variables technique, it is worthwhile to dedicate this remark to clarifying the similarities and differences between the first-order

---

[1] We do not even need to find the optimal value of these parameters. It may be the case that a direct but not optimal choice of parameter values already delivers substantial improvements in accuracy at a very low computational cost. When one is maximizing, for example, a likelihood function, being at the true maximum matters. When one is finding parameters that improve accuracy, optimality is desirable but not essential, and it can be traded off against computational cost.

perturbation in logs and the lognormal–loglinear approximation. The best way to illustrate this point is with a concrete example. Imagine that we have a household with utility function

$$\max \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \log C_t$$

and budget constraint:

$$W_{t+1} = R_{t+1}(W_t - C_t)$$

where W_t is total wealth and W_0 is given. Then, the optimality conditions are:

$$1 = \beta \mathbb{E}_t \frac{C_t}{C_{t+1}} R_{t+1}$$

$$W_{t+1} = R_{t+1}(W_t - C_t)$$

with steady state $R = \dfrac{1}{\beta}$ and $W = R(W - C)$.

Under a standard first-order perturbation in logs (loglinearization) around the previous steady state, and after some algebra:

$$\mathbb{E}_t \Delta \hat{c}_{t+1} = \mathbb{E}_t \hat{r}_{t+1}$$

$$\hat{w}_{t+1} = \hat{r}_{t+1} + \frac{1}{\rho} \hat{w}_t + \left(1 - \frac{1}{\rho}\right) \hat{c}_t$$

where, for a variable X_t,

$$\hat{x}_t = x_t - x = \log X_t - \log X$$

and $\rho = \dfrac{W - C}{W}$. Subtracting $\hat{w}_t$ from the second equation:

$$\Delta \hat{w}_{t+1} = \hat{r}_{t+1} + \left(1 - \frac{1}{\rho}\right)(\hat{c}_t - \hat{w}_t)$$

If we want to express these two equations in logs, instead of log-deviations (and using the fact that $r = -\log\beta$):

$$\mathbb{E}_t \Delta c_{t+1} = \log\beta + \mathbb{E}_t r_{t+1} \tag{29}$$

$$\Delta w_{t+1} = r_{t+1} + k + \left(1 - \frac{1}{\rho}\right)(c_t - w_t) \tag{30}$$

where

$$k = -r - \left(1 - \frac{1}{\rho}\right)(c - w).$$

In comparison, a lognormal–loglinearization still uses the approximation of the budget constraint (30), but it assumes that $\dfrac{C_t}{C_{t+1}}R_{t+1}$ is distributed as a lognormal random variable. Since, for an arbitrary variable:

$$\log \mathbb{E}_t X_t = \mathbb{E}_t \log X_t + \frac{1}{2} Var_t \log X_t,$$

we can go back to the Euler equation

$$1 = \beta \mathbb{E}_t \frac{C_t}{C_{t+1}} R_{t+1}$$

and rewrite it as:

$$0 = \log \beta + \log \mathbb{E}_t \frac{C_t}{C_{t+1}} R_{t+1}$$

$$= \log \beta + \mathbb{E}_t \log \frac{C_t}{C_{t+1}} R_{t+1} + \frac{1}{2} Var_t \log \frac{C_t}{C_{t+1}} R_{t+1}$$

or, rearranging terms:

$$\mathbb{E}_t \Delta c_{t+1} = \log \beta + \mathbb{E}_t r_{t+1} + \frac{1}{2} \left[Var_t \Delta c_{t+1} + Var_t r_{t+1} - 2cov_t(\Delta c_{t+1}, r_{t+1}) \right] \qquad (31)$$

More in general, in a lognormal–loglinearization, we approximate the nonexpectational equations with a standard loglinearization and we develop the expectational ones (or at least the ones with returns on them) using a lognormal assumption. In particular, we do not approximate the Euler equation. Once we have assumed that $\dfrac{C_t}{C_{t+1}}R_{t+1}$ is lognormal, all the results are exact.

If we compare the two equations for the first difference of consumption, (29) and (31), we see that the lognormal–loglinear approximation introduces an additional term

$$\frac{1}{2} \left[Var_t \Delta c_{t+1} + Var_t r_{t+1} - 2cov_t(\Delta c_{t+1}, r_{t+1}) \right]$$

that breaks certainty equivalence. This novel feature has important advantages. For example, for a pricing kernel M_t and an asset i, we have the pricing equation:

$$1 = \mathbb{E}_t M_{t+1} R_{i,t+1}.$$

Then:

$$0 = \mathbb{E}_t \log M_{t+1} R_{i,t+1} + \frac{1}{2} Var_t \log M_{t+1} R_{i,t+1}$$

or:

$$\mathbb{E}_t r_{i,t+1} = -\mathbb{E}_t m_{t+1} - \frac{1}{2} Var_t m_{t+1} - \frac{1}{2} Var_t r_{i,t+1} - cov_t \left(m_{t+1}, r_{i,t+1} \right)$$

If we look at the same expression for the risk-free bond:

$$1 = \mathbb{E}_t M_{t+1} R_{f,t+1}$$

we get:

$$r_{f,t+1} = -\mathbb{E}_t m_{t+1} - \frac{1}{2} Var_t m_{t+1}$$

and we can find that the excess return is:

$$\mathbb{E}_t r_{i,t+1} - r_{f,t+1} = -\frac{1}{2} Var_t r_{i,t+1} - cov_t \left(m_{t+1}, r_{i,t+1} \right),$$

an expression that it is easy to interpret.

On the other hand, this expression also embodies several problems. First, it is often unclear to what extent, in a general equilibrium economy, $\frac{C_t}{C_{t+1}} R_{t+1}$ is close to lognormality. Second, in lognormal–loglinear approximation, we are mixing two approaches, a lognormal assumption with a loglinearization. This is not necessarily coherent from the perspective of perturbation theory and we may lack theoretical foundations for the approach (including an absence of convergence theorems). Third, in the loglinearization, we can compute all the coefficients by solving a quadratic matrix system. In the lognormal–loglinear approximation, we need to compute second moments and, in many applications, how to do so may not be straightforward. Finally, it is not obvious how to get higher-order approximations with the lognormal–loglinear approximation, while perturbation theory can easily handle higher-order solutions.

4.6 Perturbing the Value Function

In some applications, it is necessary to perturb the value function of a DSGE model, for example, when we are dealing with recursive preferences or when we want to evaluate welfare. Furthermore, a perturbed value function can be an outstanding initial guess for value function iteration, making it possible to deal with high-dimensional problems that could be otherwise too slow to converge. Given the importance of perturbing the value function, this section illustrates in some detail how to do so.

Since all that we learned in the general case subsection will still apply by just changing the operator $\mathcal{H}$ from the equilibrium conditions to the Bellman operator, we can go directly to a concrete application. Consider a value function problem (following the same notation as above).

$$V(k_t, z_t) = \max_{c_t} \left[(1-\beta)\log c_t + \beta \mathbb{E}_t V(k_{t+1}, z_{t+1}) \right]$$
$$\text{s.t. } c_t + k_{t+1} = e^{z_t} k_t^\alpha + (1-\delta)k_t$$
$$z_t = \rho z_{t-1} + \eta \varepsilon_t, \varepsilon_t \sim N(0,1)$$

where we have "normalized" $\log c_t$ by $(1-\beta)$ to make the value function and the utility function have the same order of magnitude (thanks to normalization, $V_{ss} = \log c$, where V_{ss} is the steady-state value function and c is the steady-state consumption).

We can rewrite the problem in terms of a perturbation parameter σ:

$$V(k_t, z_t; \sigma) = \max_{c_t} \left[\log c_t + \beta \mathbb{E}_t V \left(e^{z_t} k_t^\alpha + (1-\delta)k_t - c_t, \rho z_t + \sigma \eta \varepsilon_{t+1}; \sigma \right) \right].$$

Note that we have made explicit the dependencies in the next period states from the current period state. The perturbation solution of this problem is a value function $V(k_t, z_t; \sigma)$ and a policy function for consumption $c(k_t, z_t; \sigma)$. For example, the second-order Taylor approximation of the value function around the deterministic steady state $(k, 0; 0)$ is:

$$V(k_t, z_t; \sigma) = V_{ss} + V_{1,ss}(k_t - k) + V_{2,ss}z_t + V_{3,ss}\sigma$$
$$+ \frac{1}{2}V_{11,ss}(k_t - k)^2 + V_{12,ss}(k_t - k)z_t + V_{13,ss}(k_t - k)\sigma$$
$$+ \frac{1}{2}V_{22,ss}z_t^2 + V_{23,ss}z_t\sigma + \frac{1}{2}V_{33,ss}\sigma^2$$

where:

$$V_{ss} = V(k, 0; 0)$$
$$V_{i,ss} = V_i(k, 0; 0) \text{ for } i = \{1,2,3\}$$
$$V_{ij,ss} = V_{ij}(k, 0; 0) \text{ for } i, j = \{1,2,3\}$$

By certainty equivalence:

$$V_{3,ss} = V_{13,ss} = V_{23,ss} = 0$$

and then:

$$V(k_t, z_t; 1) = V_{ss} + V_{1,ss}(k_t - k) + V_{2,ss}z_t$$
$$+ \frac{1}{2}V_{11,ss}(k_t - k)^2 + \frac{1}{2}V_{22,ss}z_{tt}^2 + V_{12,ss}(k_t - k)z + \frac{1}{2}V_{33,ss}\sigma^2$$

Note that $V_{33,ss} \neq s0$, a difference from the LQ approximation to the utility function that we discussed in Remark 13.

Similarly, the policy function for consumption can be expanded as:

$$c_t = c(k_t, z_t; \sigma) = c_{ss} + c_{1,ss}(k_t - k) + c_{2,ss}z_t + c_{3,ss}\sigma$$

where $c_{i,ss} = c_1(k, 0; 0)$ for $i = \{1,2,3\}$. Since the first derivatives of the consumption function only depend on the first and second derivatives of the value function, we must

have that $c_{3,ss} = 0$ (remember that precautionary consumption depends on the third derivative of the value function; Kimball, 1990).

To find the linear components of our approximation to the value function, we take derivatives of the value function with respect to controls (c_t), states (k_t, z_t), and the perturbation parameter σ and solve the associated system of equations when $\sigma = 0$. We can find the quadratic components of the value function by taking second derivatives, plugging in the known components from the previous step, and solving the system when $\sigma = 0$.

We are ready now to show some of the advantages of perturbing the value function. First, we have an evaluation of the welfare cost of business cycle fluctuations readily available. At the deterministic steady state $k_t = k$ and $z_t = 0$, we have:

$$V(k, 0; \sigma) = V_{ss} + \frac{1}{2}V_{33,ss}\sigma^2.$$

Hence $\frac{1}{2}V_{33,ss}\sigma^2$ is a measure of the welfare cost of the business cycle: it is the difference, up to second-order, between the value function evaluated at the steady-state value of the state variables $(k, 0)$ and the steady-state value function (where not only are we at the steady state, but where we know that in future periods we will be at that point as well). Note that this last quantity is not necessarily negative. Indeed, it may well be positive in many models, such as in a stochastic neoclassical growth model with leisure choice. For an explanation and quantitative evidence, see Cho et al. (2015).[m]

It is easier to interpret $V_{33,ss}$ if we can transform it into consumption units. To do so, we compute the decrease in consumption τ that will make the household indifferent between consuming $(1 - \tau)c$ units per period with certainty or c_t units with uncertainty. That is, τ satisfies:

$$\log(1 - \tau)c = \log c + \frac{1}{2}V_{33,ss}\sigma^2$$

where we have used $V_{ss} = \log c$. Then,

$$\tau = 1 - e^{\frac{1}{2}V_{33,ss}\sigma^2}.$$

---

[m] In his classical calculation about the welfare cost of the business cycle, Lucas Jr. (1987) assumed an endowment economy, where the representative household faces the same consumption process as the one observed for the US economy. Thus, for any utility function with risk aversion, the welfare cost of the business cycle must be positive (although Lucas' point, of course, was that it was rather small). When consumption and labor supply are endogenous, agents can take advantage of uncertainty to increase their welfare. A direct utility function that is concave in allocations can generate a convex indirect utility function on prices and those prices change in general equilibrium as a consequence of the agents' responses to uncertainty.

We close this section with a numerical application. To do so, we pick the same calibration as in Table 1. We get:

$$V(k_t, z_t; 1) = -0.54000 + 0.026(k_t - 0.188) + 0.250z_t - 0.069(k_t - 0.188)^2 \quad (32)$$

(where, for this calibration, $V_{kz} = V_{z^2} = V_{\sigma^2} = 0$) and:

$$c(k_t, z_t; \chi) = 0.388 + 0.680(k_t - 0.188) + 0.388z_t,$$

which is the same approximation to the consumption decision rule we found when we tackled the equilibrium conditions of the model. For this calibration, the welfare cost of the business cycle is zero.[n]

We can also use Eq. (32) as an initial guess for value function iteration. Thanks to it, instead of having to iterate hundreds of times, as if we were starting from a blind initial guess, value function iteration can converge after only a few dozen interactions.

Finally, a mixed strategy is to stack both the equilibrium conditions of the model and the value function evaluated at the optimal decision rules:

$$V(k_t, z_t) = (1 - \beta) \log c_t + \beta \mathbb{E}_t V(k_{t+1}, z_{t+1}).$$

in the operator $\mathcal{H}$. This strategy delivers an approximation to the value function and the decision rules with a trivial cost.[o]

5. PROJECTION

Projection methods (also known as weighted residual methods) handle DSGE models by building a function indexed by some coefficients that approximately solves the operator $\mathcal{H}$. The coefficients are selected to minimize a residual function that evaluates how far away the solution is from generating a zero in $\mathcal{H}$. More concretely, projection methods solve:

$$\mathcal{H}(d) = \mathbf{0}$$

---

[n] Recall that the exact consumption decision rule is $c_t = 0.673 e^{z_t} k_t^{0.33}$. Since the utility function is log, the period utility from this decision rule is $\log c_t = z_t + \log 0.673 + 0.33 \log k_t$. The unconditional mean of z_t is 0 and the capital decision rule is certainty equivalent in logs. Thus, there is no (unconditional) welfare cost of changing the variance of z_t.

[o] We could also stack derivatives of the value function, such as:

$$(1 - \beta)c_t^{-1} - \beta \mathbb{E}_t V_{1,t+1} = 0$$

and find the perturbation approximation to the derivative of the value function (which can be of interest in itself or employed in finding higher-order approximations of the value function).

by specifying a linear combination:

$$d^j(\mathbf{x}|\theta) = \sum_{i=0}^{j} \theta_i \Psi_i(\mathbf{x}) \tag{33}$$

of basis function $\Psi_i(\mathbf{x})$ given coefficients $\theta = \{\theta_0, \dots, \theta_j\}$. Then, we define a residual function:

$$R(\mathbf{x}|\theta) = \mathcal{H}\big(d^j(\mathbf{x}|\theta)\big)$$

and we select the values of the coefficients θ that minimize the residual given some metric. This last step is known as "projecting" $\mathcal{H}$ against that basis to find the components of θ (and hence the name of the method).

Inspection of Eq. (33) reveals that to build the function $d^j(\mathbf{x}|\theta)$, we need to pick a basis $\{\Psi_i(\mathbf{x})\}_{i=0}^{\infty}$ and decide which inner product we will use to "project" $\mathcal{H}$ against that basis to compute θ. Different choices of bases and of the projection algorithm will imply different projection methods. These alternative projections are often called in the literature by their own particular names, which can be sometimes bewildering.

Projection theory, which has been applied in *ad hoc* ways by economists over the years, was popularized as a rigorous approach in economics by Judd (1992) and Gaspar and Judd (1997) and, as in the case of perturbation, it has been authoritatively presented by Judd (1998).[P]

Remark 16 (Linear v. nonlinear combinations) Instead of linear combinations of basis functions, we could deal with more general nonlinear combinations:

$$d^j(\mathbf{x}|\theta) = f\left(\{\Psi_i(\mathbf{x})\}_{i=0}^{j}|\theta\right)$$

for a known function f. However, the theory for nonlinear combinations is less well developed, and we can already capture a full range of nonlinearities in d^j with the appropriate choice of basis functions Ψ_i. In any case, it is more pedagogical to start with the linear combination case. Most of the ideas in the next pages carry over the case of nonlinear combinations. The fact that we are working with linear combinations of basis functions also means that, in general, we will have the same number of coefficients θ as the number of basis functions Ψ_i times the dimensionality of d^j.

[P] Projection theory is more modern than perturbation. Nevertheless, projection methods have been used for many decades in the natural sciences and engineering. Spectral methods go back, at least, to Lanczos (1938). Alexander Hrennikoff and Richard Courant developed the finite elements method in the 1940s, although the method was christened by Clough (1960), who made pioneering contributions while working at Boeing. See Clough and Wilson (1999) for a history of the early research on finite elements.

5.1 A Basic Projection Algorithm

Conceptually, projection is easier to present than perturbation (although its computational implementation is harder). We can start directly by outlining a projection algorithm:

Algorithm 1 (Projection Algorithm)

1. Define $j + 1$ known linearly independent functions $\psi_i : \Omega \to \mathbb{R}$ where $j < \infty$. We call the $\psi_0, \psi_1, \ldots, \psi_j$ the *basis functions*. These basis functions depend on the vector of state variables x.
2. Define a vector of coefficients $\theta^l = [\theta^l_0, \theta^l_1, \ldots, \theta^l_j]$ for $l = 1, \ldots, m$ (where recall that m is the dimension that the function d of interest maps into). Stack all coefficients on a $m \times (j+1)$ matrix $\theta = [\theta^1; \theta^2; \ldots; \theta^l]$.
3. Define a combination of the basis functions and the θ's:

$$d^{l,j}\left(\cdot \,|\theta^l \right) = \sum_{i=0}^{j} \theta^l_i \psi_i(\cdot)$$

for $l = 1, \ldots, m$. Then:

$$d^j(\cdot \,|\theta) = \left[d^{1,j}\left(\cdot \,|\theta^l \right); d^{2,j}\left(\cdot \,|\theta^l \right); \ldots; d^{m,j}\left(\cdot \,|\theta^l \right) \right].$$

4. Plug $d^j (\cdot|\theta)$ into the operator $H(\cdot)$ to find the *residual equation*:

$$R(\cdot \,|\theta) = \mathcal{H}\left(d^j(\cdot \,|\theta) \right).$$

5. Find the value of $\theta\,\hat{}$ that makes the residual equation as close to **0** as possible given some objective function $\rho : J^2 \times J^2 \to \mathbb{R}$:

$$\hat{\theta} = \arg\min_{\theta \in \mathbb{R}^{m \times (j+1)}} \rho\left(R(\cdot \,|\theta), \mathbf{0} \right).$$

To ease notation, we have made two simplifications on the previous algorithm. First, we assumed that, along each dimension of d, we used the same basis functions ψ_i and the same number $j + 1$ of them. Nothing forces us to do so. At the mere cost of cumbersome notation, we could have different basis functions for each dimension and a different number of them (ie, different j's). While the former is not too common in practice, the latter is standard, since some variables' influence on the function d can be harder to approximate than others'.[q]

We specify a metric function ρ to gauge how close the residual function is to zero over the domain of the state variables. For example, in Fig. 2, we plot two different residual

---

[q] For the nonlinear combination case, $f\left(\{\Psi_i(\mathbf{x})\}_{i=0}^{j} |\theta \right)$, we would just write the residual function:

$$R(\cdot \,|\theta) = \mathcal{H}\left(f\left(\{\Psi_i(\mathbf{x})\}_{i=0}^{j} |\theta \right) \right)$$

and find the θ's that minimize a given metric. Besides the possible computational complexities of dealing with arbitrary functions $f\left(\{\Psi_i(\mathbf{x})\}_{i=0}^{j} |\theta \right)$, the conceptual steps are the same.
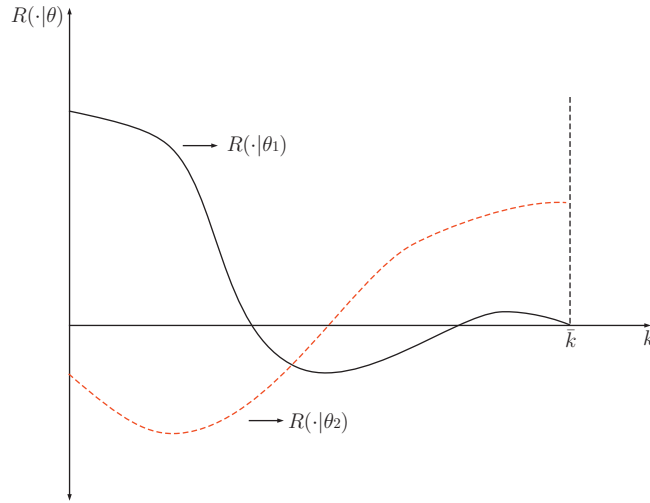
Fig. 2 Residual functions.

functions for a problem with only one state variable k_t (think, for instance, of a deterministic neoclassical growth model) that belongs to the interval $[0, \bar{k}]$, one for coefficients θ_1 (continuous line) and one for coefficients θ_2 (discontinuous line). $R(\cdot|\theta_1)$ has large values for low values of k_t, but has small values for high levels of k_t. $R(\cdot|\theta_2)$ has larger values on average, but it never gets as large as $R(\cdot|\theta_1)$. Which of the two residual functions is closer to zero over the interval? Obviously, different choices of ρ will yield different answers. We will discuss below how to select a good ρ.

A small example illustrates the previous steps. Remember that we had, for the stochastic neoclassical growth model, the system built by the Euler equation and the resource constraint of the economy:

$$\mathcal{H}(d) = \begin{cases} u'\left(d^1\left(k_t, z_t\right)\right) \\ -\beta \mathbb{E}_t\left[u'\left(d^1\left(d^2(k_t, z_t), z_{t+1}\right)\right)\left(\alpha e^{\rho z_t + \sigma\varepsilon_{t+1}}\left(d^2\left(k_t, z_t\right)\right)^{\alpha-1} + 1 - \delta\right)\right] = \mathbf{0}, \\ d^1\left(k_t, z_t\right) + d^2\left(k_t, z_t\right) - e^{z_t}k_t^\alpha - (1-\delta)k_t \end{cases}$$

for all k_t and z_t and where:

$$c_t = d^1\left(k_t, z_t\right)$$
$$k_{t+1} = d^2\left(k_t, z_t\right)$$

and we have already recursively substituted k_{t+1} in the decision rule of consumption evaluated at $t + 1$. Then, we can define

$$c_t = d^{1,j}\left(k_t, z_t|\theta^1\right) = \sum_{i=0}^{j} \theta_i^1 \psi_i(k_t, z_t)$$

and

$$k_{t+1} = d^{2,j}\left(k_t, z_t | \theta^2\right) = \sum_{i=0}^{j} \theta_i^2 \psi_i(k_t, z_t)$$

for some $\psi_0(k_t, z_t), \psi_1(k_t, z_t), \ldots, \psi_j(k_t, z_t)$. Below we will discuss which basis functions we can select for this role.

The next step is to write the residual function:

$$R(k_t, z_t | \theta) = \left\{ \begin{array}{l} u'\left(\sum_{i=0}^{j} \theta_i^1 \psi_i(k_t, z_t)\right) \\ -\beta \mathbb{E}_t \left[\begin{array}{l} u'\left(\sum_{i=0}^{j} \theta_i^1 \psi_i\left(\sum_{i=0}^{j} \theta_i^2 \psi_n(k_t, z_t), \rho z_t + \sigma \varepsilon_{t+1}\right)\right) \\ \times \left(\alpha e^{\rho z_t + \sigma \varepsilon_{t+1}} \left(\sum_{i=0}^{j} \theta_i^2 \psi_i(k_t, z_t)\right)^{\alpha-1} + 1 - \delta\right) \end{array} \right] \\ \sum_{i=0}^{j} \theta_i^1 \psi_i(k_t, z_t) + \sum_{i=0}^{j} \theta_i^2 \psi_i(k_t, z_t) - e^{z_t} k_t^{\alpha} - (1-\delta)k_t \end{array} \right. ,$$

for all k_t and z_t, $\theta = [\theta^1; \theta^2]$.

The final step is to find $\hat{\theta} = \arg\min_{\theta \in \mathbb{R}^{m \times (j+1)}} \rho(R(\cdot | \theta), \mathbf{0})$. Again, we will discuss these choices below in detail, but just for concreteness, let us imagine that we pick $m \times (j+1)$ points (k_l, z_l) and select the metric function to be zero at each of these $m \times (j+1)$ points and one everywhere else. Such a metric is trivially minimized if we make the residual function equal to zero exactly on those points. This is equivalent to solving the system of $m \times (j+1)$ equations:

$$R(k_l, z_l | \theta) = \mathbf{0}, \text{ for } l = 1, \ldots, m \times (j+1)$$

with $m \times (j+1)$ unknowns (we avoid here the discussion about the existence and uniqueness of such a solution).

Remark 17 (Relation to econometrics) Many readers will be familiar with the use of the word "projection" in econometrics. This is not a coincidence. A common way to present linear regression is to think about the problem of searching for the unknown conditional expectation function:

$$\mathbb{E}(Y|X)$$

for some variables Y and X. Given that this conditional expectation is unknown, we can approximate it with the first two monomials on X, 1 (a constant) and X (a linear function), and associated coefficients θ_0 and θ_1:

$$\mathbb{E}(Y|X) \simeq \theta_0 + \theta_1 X.$$

These two monomials are the first two elements of a basis composed by the monomials (and also of the Chebyshev polynomials, a basis of choice later in this section). The residual function is then:

$$R(Y, X|\theta_0, \theta_1) = Y - \theta_0 - \theta_1 X.$$

The most common metric in statistical work is to minimize the square of this residual:

$$R(Y, X|\theta_0, \theta_1)^2$$

by plugging in the observed series $\{Y, X\}_{t=1:T}$. The difference, thus, between ordinary least squares and the projection algorithm is that while in the former we use observed data, in the latter we use the operator $\mathcal{H}(d)$ imposed by economic theory. This link is even clearer when we study the econometrics of semi-nonparametric methods, such as sieves (Chen, 2007), which look for flexible basis functions indexed by a low number of coefficients and that, nevertheless, impose fewer restrictions than a linear regression.

Remark 18 (Comparison with other methods) From our short description of projection methods, we can already see that other algorithms in economics are particular cases of it. Think, for example, about the parameterized expectations approach (Marcet and Lorenzoni, 1999). This approach consists of four steps.

First, the conditional expectations that appear in the equilibrium conditions of the model are written as a flexible function of the state variables of the model and some coefficients. Second, the coefficients are initialized at an arbitrary value. Third, the values of the coefficients are updated by running a nonlinear regression that minimizes the distance between the conditional expectations forecasted by the function guessed in step 1 and the actual realization of the model along a sufficiently long simulation. Step 3 is repeated until the coefficient values used to simulate the model and the coefficient values that come out of the nonlinear regression are close enough.

Step 1 is the same as in any other projection method: the function of interest (in this case the conditional expectation) is approximated by a flexible combination of basis functions. Often the parameterized expectations approach relies on monomials to do so (or functions of the monomials), which, as we will argue below, is rarely an optimal choice. But this is not an inherent property of the approach. Christiano and Fisher (2000) propose to use functions of Chebyshev polynomials, which will yield better results. More important is the iterative procedure outlined by steps 2–4. Finding the fixed point of the values of the coefficients by simulation and a quadratic distances is rarely the best option. Even if, under certain technical conditions (Marcet and Marshall, 1994) the algorithm converges, such convergence can be slow and fragile. In the main text, we will explain that a collocation approach can achieve the same goal much more efficiently and without having to resort to simulation (although there may be concrete cases where simulation is a superior strategy).

Value function iteration and policy function iteration can also be understood as particular forms of projection, where the basis functions are linear functions (or higher-order

interpolating functions such as splines). Since in this chapter we are not dealing with these methods, we skip further details.

5.2 Choice of Basis and Metric Functions

The previous subsection highlighted the two issues ahead of us: how to decide which basis $\psi_0, \psi_1, \ldots, \psi_j$ to select and which metric function ρ to use. Different choices in each of these issues will result in slightly different projection methods, each with its weaknesses and strengths.

Regarding the first issue, we can pick a global basis (ie, basis functions that are nonzero and smooth for most of the domain of the state variable Ω) or a local basis (ie, basis functions that are zero for most of the domain of the state variable, and nonzero and smooth for only a small portion of the domain Ω). Projection methods with a global basis are often known as spectral methods. Projection methods with a local basis are also known as finite elements methods.

5.3 Spectral Bases

Spectral techniques were introduced in economics by Judd (1992). The main advantage of this class of global basis functions is their simplicity: building and working with the approximation will be straightforward. The main disadvantage of spectral bases is that they have a hard time dealing with local behavior. Think, for instance, about Fig. 3, which plots the decision rule $k_{t+1} = d(k_t)$ that determines capital tomorrow given capital today for some model that implies a nonmonotone, local behavior represented by the hump in the middle of the capital range (perhaps due to a complicated incentive constraint). The change in the coefficients θ required to capture that local shape of d would leak into the approximation for the whole domain Ω. Similar local behavior appears when we deal with occasionally binding constraints, kinks, or singularities.
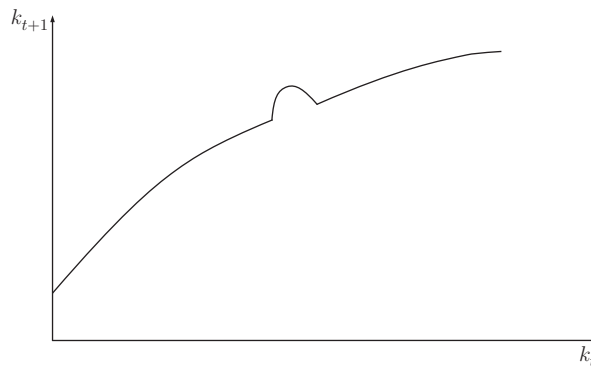
Fig. 3 Decision rule for capital.
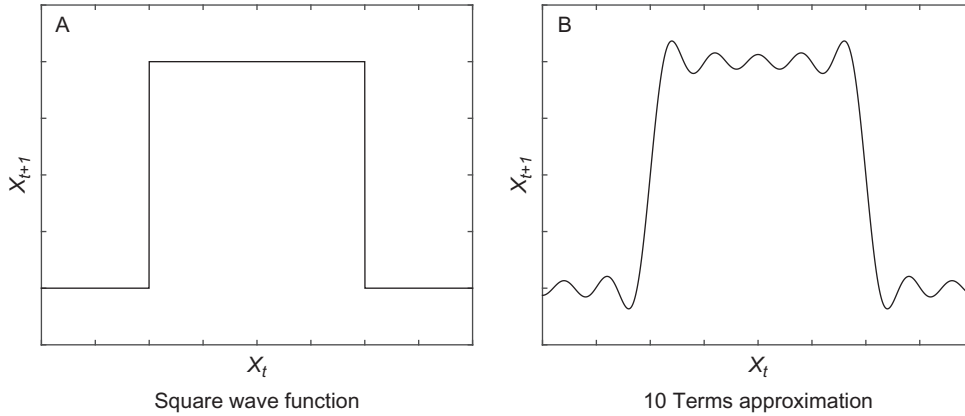
Fig. 4 Gibbs phenomenon.

A well-known example of this problem is the Gibbs phenomenon. Imagine that we are trying to approximate a piecewise continuously differentiable periodic function with a jump discontinuity, such as a square wave function (Fig. 4, panel A):

$$f(x) = \begin{cases} \dfrac{\pi}{4}, & \text{if } x \in [2j\pi, 2(j+1)\pi] \text{ and for } \forall j \in \mathbb{N} \\ -\dfrac{\pi}{4}, & \text{otherwise.} \end{cases}$$

Given that the function is periodic, a sensible choice for a basis is a trigonometric series $\sin(x)$, $\sin(2x)$, $\sin(3x)$, ... The optimal approximation is:

$$\sin(x) + \frac{1}{3}\sin(3x) + \frac{1}{5}\sin(5x) + \dots$$

The approximation behaves poorly at a jump discontinuity. As shown in Fig. 4, panel B, even after using 10 terms, the approximation shows large fluctuations around all the discontinuity points $2j\pi$ and $2(j+1)\pi$. These fluctuations will exist even if we keep adding many more terms to the approximation. In fact, the rate of convergence to the true solution as $n \to \infty$ is only $O(n)$.

5.3.1 Unidimensional Bases

We will introduce in this subsection some of the most common spectral bases. First, we will deal with the unidimensional case where there is only one state variable. This will allow us to present most of the relevant information in a succinct fashion. It would be important to remember, however, that our exposition of unidimensional bases cannot be exhaustive (for instance, in the interest of space, we will skip splines) and that the researcher may find herself tackling a problem that requires a specific basis. One of the great advantages of projection methods is their flexibility to accommodate

unexpected requirements. In the next subsection, we will deal with the case of an arbitrary number of state variables and we will discuss how to address the biggest challenge of projection methods: the curse of dimensionality.

5.3.1.1 Monomials

A first basis is the monomials 1, x, x^2, x^3, ... Monomials are simple and intuitive. Furthermore, even if this basis is not composed by orthogonal functions, if J_1 is the space of bounded measurable functions on a compact set, the Stone–Weierstrass theorem tells us that we can uniformly approximate any continuous function defined on a closed interval with linear combinations of these monomials.

Rudin (1976, p. 162) provides a formal statement of the theorem:

Theorem 1 (Stone–Weierstrass) *Let $\mathcal{A}$ be an algebra of real continuous functions on a compact set K. If $\mathcal{A}$ separates points on K and if $\mathcal{A}$ vanishes at no point of K, then the uniform closure $\mathcal{B}$ of $\mathcal{A}$ consists of all real continuous functions on K.*

A consequence of this theorem is that if we have a real function f that is continuous on K, we can find another function $h \in \mathcal{B}$ such that for $\varepsilon > 0$:

$$|f(x) - h(x)| < \varepsilon,$$

for all $x \in K$.

Unfortunately, monomials suffer from two severe problems. First, monomials are (nearly) multicollinear. Fig. 5 plots the graphs of x^{10} (continuous line) and x^{11} (discontinuous line) for $x \in [0.5, 1.5]$. Both functions have a very similar shape. As we add higher monomials, the new components of the solution do not allow the distance between the exact function we want to approximate and the computed approximation to diminish sufficiently fast.[r]

Second, monomials vary considerably in size, leading to scaling problems and the accumulation of numerical errors. We can also see this point in Fig. 5: x^{11} goes from $4.8828e^{-04}$ to 86.4976 just by moving x from 0.5 to 1.5.

The challenges presented by the use of monomials motivate the search for an orthogonal basis in a natural inner product that has a bounded variation in range. Orthogonality will imply that when we add more one element of the basis (ie, when we go from order j

---

[r] A sharp case of this problem is when $\mathcal{H}(\cdot)$ is linear. In that situation, the solution of the projection involves the inversion of matrices. When the basis functions are similar, the condition numbers of these matrices (the ratio of the largest and smallest absolute eigenvalues) are too high. Just the first six monomials can generate condition numbers of 10^{10}. In fact, the matrix of the least squares problem of fitting a polynomial of degree 6 to a function (the *Hilbert Matrix*) is a popular test of numerical accuracy since it maximizes rounding errors. The problem of the multicollinearity of monomials is also well appreciated in econometrics.
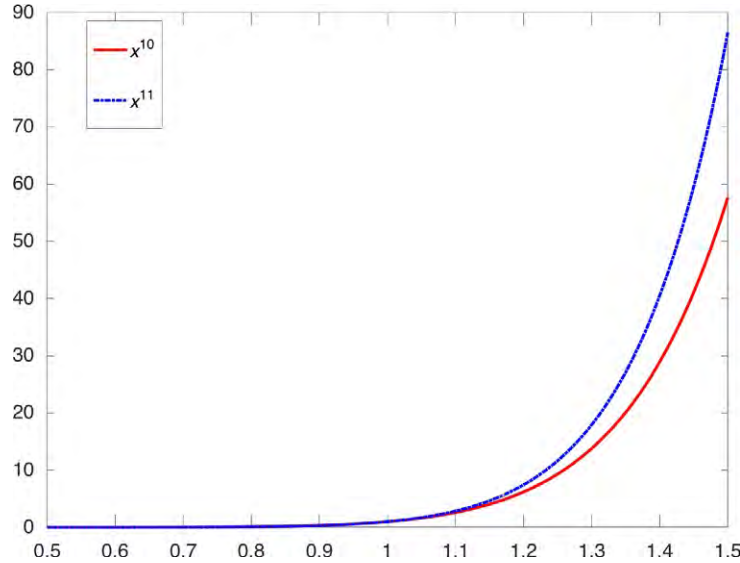
Fig. 5 Graphs of x^{10} and x^{11}.

to order $j + 1$), the newest element brings a sufficiently different behavior so as to capture features of the unknown function d not well approximated by the previous elements of the basis.

5.3.1.2 Trigonometric Series

A second basis is a trigonometric series

$$1/(2\pi)^{0.5}, \cos x/(2\pi)^{0.5}, \sin x/(2\pi)^{0.5}, \ldots,$$
$$\cos kx/(2\pi)^{0.5}, \sin kx/(2\pi)^{0.5}, \ldots$$

Trigonometric series are well suited to approximate periodic functions (recall our example before of the square wave function). Trigonometric series are, therefore, quite popular in the natural sciences and engineering, where periodic problems are common. Furthermore, they are easy to manipulate as we have plenty of results involving the transformation of trigonometric functions and we can bring to the table the powerful tools of Fourier analysis. Sadly, economic problems are rarely periodic (except in the frequency analysis of time series) and periodic approximations to nonperiodic functions are highly inefficient.

5.3.1.3 Orthogonal Polynomials of the Jacobi Type

We motivated before the need to use a basis of orthogonal functions. Orthogonal polynomials of the Jacobi (also known as hypergeometric) type are a flexible class of polynomials well suited for our needs.

The Jacobi polynomial of degree n, $P_n^{\alpha,\beta}(x)$ for $\alpha, \beta > -1$, is defined by the orthogonality condition:

$$\int_{-1}^{1} (1-x)^{\alpha}(1+x)^{\beta} P_n^{\alpha,\beta}(x) P_m^{\alpha,\beta}(x) dx = 0 \text{ for } m \neq n$$

One advantage of this class of polynomials is that we have a large number of alternative expressions for them. The orthogonality condition implies, with the normalizations:

$$P_n^{\alpha,\beta}(1) = \binom{n+\alpha}{n},$$

that the general n term is given by:

$$2^{-n} \sum_{k=0}^{n} \binom{n+\alpha}{k}\binom{n+\beta}{n-k}(x-1)^{n-k}(x+1)^{k}$$

Recursively:

$$2(n+1)(n+\alpha+\beta+1)(2n+\alpha+\beta)P_{n+1} = \left(\begin{array}{c} (2n+\alpha+\beta+1)(\alpha^2-\beta^2) \\ +(2n+\alpha+\beta)(2n+\alpha+\beta+1)(2n+\alpha+\beta+2)x \end{array} \right) P_n$$
$$-2(n+\alpha)(n+\beta)(2n+\alpha+\beta+2)P_{n-1}$$

Two important cases of Jacobi polynomials are the Legendre polynomials, where $\alpha = \beta = -\frac{1}{2}$, and the Chebyshev polynomials, where $\alpha = \beta = 0$. There is a generalization of Legendre and Chebyshev polynomials, still within the Jacobi family, known as the Gegenbauer polynomials, which set $\alpha = \beta = v - \frac{1}{2}$ for a parameter v.

Boyd and Petschek (2014) compare the performance of Gegenbauer, Legendre, and Chebyshev polynomials. Their table 1 is particularly informative. We read it as suggesting that, except for some exceptions that we find of less relevance in the solution of DSGE models, Chebyshev polynomials are the most convenient of the three classes of polynomials. Thus, from now on, we focus on Chebyshev polynomials.

5.3.1.4 Chebyshev Polynomials

Chebyshev polynomials are one of the most common tools of applied mathematics. See, for example, Boyd (2000) and Fornberg (1996) for references and background material. The popularity of Chebyshev polynomials is easily explained if we consider some of their advantages.

First, numerous simple closed-form expressions for the Chebyshev polynomials are available. Thus, the researcher can easily move from one representation to another

according to her convenience. Second, the change between the coefficients of a Chebyshev expansion of a function and the values of the function at the Chebyshev nodes is quickly performed by the cosine transform. Third, Chebyshev polynomials are more robust than their alternatives for interpolation. Fourth, Chebyshev polynomials are smooth and bounded between $[-1, 1]$. Finally, several theorems bound the errors for Chebyshev polynomials' interpolations.

The most common definition of the Chebyshev polynomials is recursive, with $T_0(x) = 1$, $T_1(x) = x$, and the general $n + 1$-th order polynomial given by:

$$T_{n+1}(x) = 2x T_n(x) - T_{n-1}(x)$$

Applying this recursive definition, the first few polynomials are 1, x, $2x^2 - 1$, $4x^3 - 3x$, $8x^4 - 8x^2 + 1,\ldots$ Thus, the approximation of a function with Chebyshev polynomials is not different from an approximation with monomials (and, thus, we can rely on appropriate versions of the Stone–Weierstrass theorem), except that the orthogonality properties of how Chebyshev polynomials group the monomials make the approximation better conditioned.

Fig. 6 plots the Chebyshev polynomials of order 0–5. The first two polynomials coincide with the first two monomials, a constant and the 45-degree line. The Chebyshev polynomial of order two is a parabola. Higher-order Chebyshev polynomials accumulate several waves. Fig. 6 shows that the Chebyshev polynomials of order n has n zeros, given by

$$x_k = \cos\left(\frac{2k-1}{2n}\pi\right), k = 1,\ldots,n.$$

This property will be useful when we describe collocation in a few pages. Also, these zeros are quadratically clustered toward ± 1.

Other explicit and equivalent definitions for the Chebyshev polynomials include

$$T_n(x) = \cos(n \arccos x)$$

$$= \frac{1}{2}\left(z^n + \frac{1}{z^n}\right) \text{ where } \frac{1}{2}\left(z + \frac{1}{z}\right) = x$$

$$= \frac{1}{2}\left(\left(x + (x^2 - 1)^{0.5}\right)^n + \left(x - (x^2 - 1)^{0.5}\right)^n\right)$$

$$= \frac{1}{2}\sum_{k=0}^{[n/2]}(-1)^k \frac{(n-k-1)!}{k!(n-2k)!}(2x)^{n-2k}$$

$$= \frac{(-1)^n \pi^{0.5}}{2^n \Gamma\left(n + \frac{1}{2}\right)}(1 - x^2)^{0.5}\frac{d^n}{dx^n}\left((1 - x^2)^{n-\frac{1}{2}}\right).$$
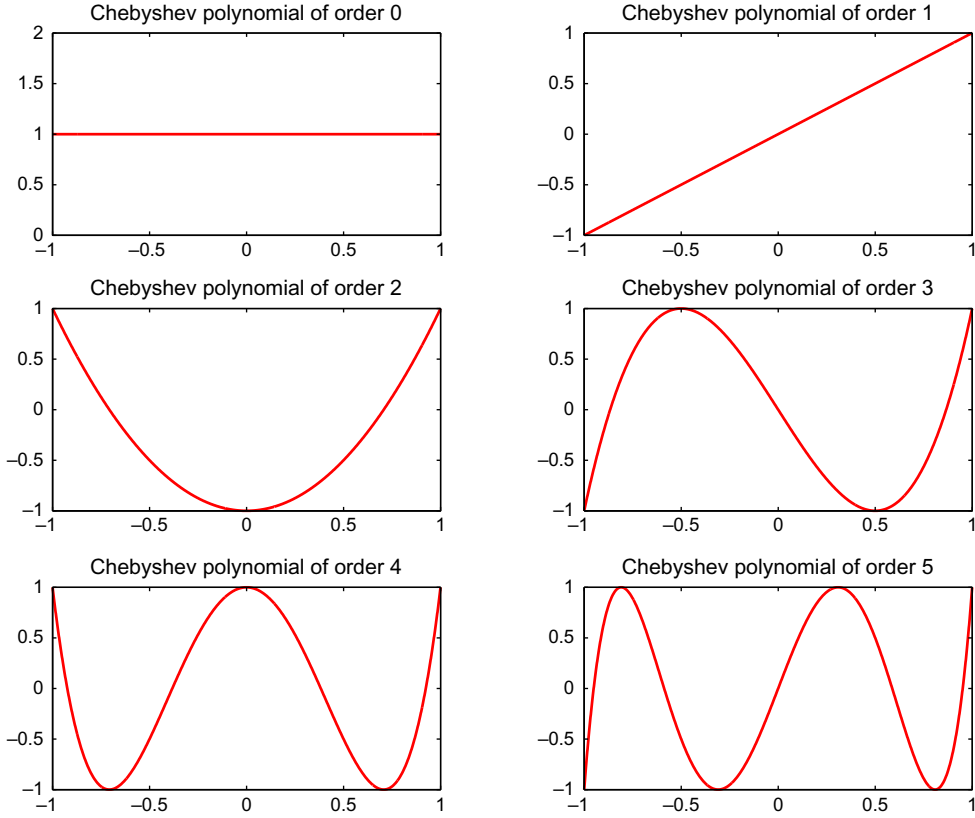
Fig. 6 First six Chebyshev polynomials.

Perhaps the most interesting of these definitions is the first one, since it tells us that Chebyshev polynomials are a trigonometric series in disguise (Boyd, 2000).

A few additional facts about Chebyshev polynomials deserve to be highlighted. First, the $n + 1$ extrema of the polynomial $T_n(x_k)$ $(n > 0)$ are given by:

$$x_k = \cos\left(\frac{k}{n}\pi\right), \quad k = 0, \dots, n. \tag{34}$$

All these extrema are either –1 or 1. Furthermore, two of the extrema are at the endpoints of the domain: $T_n(-1) = (-1)^n$ and $T_n(1) = 1$. Second, the domain of the Chebyshev polynomials is $[-1, 1]$. Since the domain of a state variable x in a DSGE model would be, in general, different from $[-1, 1]$, we can use a linear translation from $[a, b]$ into $[-1, 1]$:

$$2\frac{x - a}{b - a} - 1.$$

Third, the Chebyshev polynomials are orthogonal with respect to the weight function:

$$w(x) = \frac{1}{(1-x^2)^{0.5}}.$$

We conclude the presentation of Chebyshev polynomials with two remarkable results, which we will use below. The first result, due to Erdös and Turán (1937),[s] tells us that if an approximating function is exact at the roots of the n_1^{th} order Chebyshev polynomial, then, as $n_1 \to \infty$, the approximation error becomes arbitrarily small. The *Chebyshev interpolation theorem* will motivate, in a few pages, the use of orthogonal collocation where we pick as collocation points the zeros of a Chebyshev polynomial (there are also related, less used, results when the extrema of the polynomials are chosen instead of the zeros).

Theorem 2 (Chebyshev interpolation theorem) *If $d(x) \in \mathcal{C}[a,b]$, if $\{\phi_i(x), i=0,...\}$ is a system of polynomials (where $\phi_i(x)$ is of exact degree i) orthogonal to with respect to $w(x)$ on $[a,b]$ and if $p_j = \sum_{i=0}^{j} \theta_i \phi_i(x)$ interpolates $f(x)$ in the zeros of $\phi_{n+1}(x)$, then:*

$$\lim_{j \to \infty} \left(\|d - p_j\|_2 \right)^2 = \lim_{n \to \infty} \int_a^b w(x) \left(d(x) - p_j \right)^2 dx = 0$$

We stated a version of the theorem that shows $\mathcal{L}_2$ convergence (a natural norm in economics), but the result holds for $\mathcal{L}_p$ convergence for any $p > 1$. Even if we called this result the *Chebyshev interpolation theorem*, its statement is more general, as it will apply to other polynomials that satisfy an orthogonality condition. The reason we used *Chebyshev* in the theorem's name is that the results are even stronger if the function $d(x)$ satisfies a Dini–Lipschitz condition and the polynomials $\phi_i(x)$ are Chebyshev to uniform convergence, a much more reassuring finding.[t]

But the previous result requires that $j \to \infty$, which is impossible in real applications. The second result will give a sense of how big is the error we are accepting by truncating the approximation of $d(\cdot)$ after a finite (and often relatively low) j.

---

[s] We reproduce the statement of the theorem, with only minor notational changes, from Mason and Handscomb (2003), chapter 3, where the interested reader can find related results and all the relevant details. This class of theorems is usually derived in the context of interpolating functions.

[t] A function f satisfies a Dini–Lipschitz condition if

$$\lim_{\delta \to 0^+} \omega(\delta) \log \delta = 0$$

where $\omega(\delta)$ is a modulus of continuity of f with respect to δ such that:

$$|f(x + \delta) - f(x)| \leq \omega(\delta).$$

Theorem 3 (Chebyshev truncation theorem, Boyd (Boyd (2000), p. 47)) *The error in approximating* d *is bounded by the sum of the absolute values of all the neglected coefficients. In other words, if we have*

$$d^j(\,\cdot\,|\theta) = \sum_{i=0}^{j} \theta_i \psi_i(\,\cdot\,)$$

then

$$\left| d(x) - d^j(x|\theta) \right| \le \sum_{i=j+1}^{\infty} |\theta_i|$$

for any $x \in [-1,1]$ *and any* j.

We can make the last result even stronger. Under certain technical conditions, we will have a geometric convergence of the Chebyshev approximation to the exact unknown function.[u] And when we have geometric convergence,

$$\left| d(x) - d^j(x|\theta) \right| \sim O(\theta_j)$$

that is, the truncation error created by stopping at the polynomial j is of the same order of magnitude as the coefficient θ_j of the last polynomial. This result also provides us with a simple numerical test: we can check the coefficient θ_j from our approximation: if θ_j is not close enough to zero, we probably need to increase j. We will revisit the evaluation of the accuracy of an approximation in Section 7.

Remark 19 (Change of variables) We mentioned above that, since a state variable x_t in a DSGE model would have, in general, a domain different from $[-1,1]$, we can use a linear translation from $[a,b]$ into $[-1,1]$:

$$2\frac{x_t - a}{b - a} - 1.$$

This transformation points to a more general idea: the change of variables as a way to improve the accuracy of an approximation (see also Section 4.5 for the application of the same idea in perturbation). Imagine that we are solving the stochastic neoclassical growth model. Instead of searching for

$$c_t = d^1(k_t, z_t)$$

and

$$k_{t+1} = d^2(k_t, z_t),$$

---

[u] Convergence of the coefficients is geometric if $\lim_{j \to \infty} \log\left(|\theta_j|\right)/j = $ constant. If the lim is infinity, convergence is supergeometric; if the lim is zero, convergence is subgeometric.

we could, instead, search for

$$\log c_t = d^1\left(\log k_t,\ z_t\right)$$

and

$$\log k_{t+1} = d^2\left(\log k_t,\ z_t\right),$$

by defining

$$\log c_t = d^{1,j}\left(\log k_t,\ z_t | \theta^1\right) = \sum_{i=0}^{j} \theta_i^1 \psi_i(\log k_t,\ z_t)$$

and

$$\log k_{t+1} = d^{2,j}\left(\log k_t,\ z_t | \theta^2\right) = \sum_{i=0}^{j} \theta_i^2 \psi_i(\log k_t,\ z_t).$$

In fact, even in the basic projection example above, we already have a taste of this idea, as we used z_t as a state variable, despite the fact that it appears in the production function as e^{z_t}. An alternative yet equivalent reparameterization writes $A_t = e^{z_t}$ and $z_t = \log A_t$. The researcher can use her a priori knowledge of the model (or preliminary computational results) to search for an appropriate change of variables in her problem. We have changed both state and control variables, but nothing forced us to do so: we could have just changed one variable but not the other or employed different changes of variables.

Remark 20 (Boyd's moral principle)
All of the conveniences of Chebyshev polynomials we just presented are not just theoretical. Decades of real-life applications have repeatedly shown how well Chebyshev polynomials work in a wide variety of applications. In the case of DSGE models, the outstanding performance of Chebyshev polynomial has been shown by Aruoba et al. (2006) and Caldara et al. (2012). Boyd (2000, p. 10), only half-jokingly, has summarized these decades of experience in what he has named his Moral Principle 1:
1. When in doubt, use Chebyshev polynomials unless the solution is spatially periodic, in which case an ordinary Fourier series is better.
2. Unless you are sure another set of basis functions is better, use Chebyshev polynomials.
3. Unless you are really, really sure another set of basis functions is better, use Chebyshev polynomials.

5.3.2 Multidimensional Bases
All of the previous discussion presented unidimensional basis functions. This was useful to introduce the topic. However, most problems in economics are multidimensional: nearly

all DSGE models involve several state variables. How do we generalize our basis functions?

The answer to this question is surprisingly important. Projection methods suffer from an acute curse of dimensionality. While solving DSGE models with one or two state variables and projection methods is relatively straightforward, solving DSGE models with 20 state variables and projection methods is a challenging task due to the curse of dimensionality. The key to tackling this class of problems is to intelligently select the multidimensional basis.

5.3.2.1 Discrete State Variables

The idea that the state variables are continuous was implicit in our previous discussion. However, there are many DSGE models where either some state variable is discrete (ie, the government can be in default or not, as in Bocola (2015), or monetary policy can be either active or passive in the sense of Leeper, 1991) or where we can discretize one continuous state variable without losing much accuracy. The best example of the latter is the discretization of exogenous stochastic processes for productivity or preference shocks. Such discretization can be done with the procedures proposed by Tauchen (1986) or Kopecky and Suen (2010), who find a finite state Markov chain that generates the same population moments as the continuous process. Experience suggests that, in most applications, a Markov chain with 5 or 7 states suffices to capture nearly all the implications of the stochastic process for quantitative analysis.

A problem with discrete state variables can be thought of as one where we search for a different decision rule for each value of that state variable. For instance, in the stochastic neoclassical growth model with state variables k_t and z_t, we can discretize the productivity level z_t into a Markov chain with n points

$$z_t \in \{z_1, .., z_n\}$$

and transition matrix:

$$P_{z,z'} = \begin{Bmatrix} p_{11} & \cdots & p_{1n} \\ \vdots & \vdots & \vdots \\ p_{n1} & \cdots & p_{nn} \end{Bmatrix} \tag{35}$$

where entry p_{ij} is the probability that the chain will move from position i in the current period to position j in the next period.

Remark 21 (Discretization methods) Tauchen (1986) procedure to discretize an AR(1) stochastic process

$$z_t = \rho z_{t-1} + \epsilon_t$$

with stationary distribution $N(0, \sigma_z^2)$, where $\sigma_z = \dfrac{\sigma_\epsilon}{\sqrt{1-\rho^2}}$, works as follows:

Algorithm 2 (AR(1) Discretization)

1. Set n, the number of potential realizations of the process z.

2. Set the upper ($\bar{z}$) and lower ($\underline{z}$) bounds for the process. An intuitive way to set the bounds is to pick m such that:

$$\bar{z} = m\sigma_z$$
$$\underline{z} = -m\sigma_z$$

The latter alternative is appealing given the symmetry of the normal distribution around 0. Usual values of m are between 2 and 3.

3. Set $\{z_i\}_{i=1}^n$ such that:

$$z_i = \underline{z} + \frac{\bar{z} - \underline{z}}{n-1}(i-1)$$

and construct the midpoints $\{\tilde{z}_i\}_{i=1}^{n-1}$, which are given by:

$$\tilde{z}_i = \frac{z_{i+1} + z_i}{2}$$

4. The transition probability $p_{ij} \in P_{z,z'}$ (the probability of going to state z_j conditional on being on state z_i), is computed according to:

$$p_{ij} = \Phi\left(\frac{\tilde{z}_j - \rho z_i}{\sigma}\right) - \Phi\left(\frac{\tilde{z}_{j-1} - \rho z_i}{\sigma}\right) \quad j = 2, 3, \ldots, n-1$$

$$p_{i1} = \Phi\left(\frac{\tilde{z}_1 - \rho z_i}{\sigma}\right)$$

$$p_{in} = 1 - \Phi\left(\frac{\tilde{z}_{n-1} - \rho z_i}{\sigma}\right)$$

where $\Phi(\cdot)$ denotes a CDF of a $N(0,1)$.

To illustrate Tauchen's procedure, let us assume we have a stochastic process:

$$z_t = 0.95 z_{t-1} + \epsilon_t$$

with $N(0, 0.007^2)$ (this is a standard quarterly calibration for the productivity process for the US economy; using data after 1984 the standard deviation is around 0.0035) and we want to approximate it with a 5-point Markov chain and $m = 3$. Tauchen's procedure gives us:

$$z_t \in \{-0.0673, -0.03360, 0.0336, 0.0673\} \tag{36}$$

and transition matrix:

$$
P_{z,z'} = \begin{Bmatrix}
0.9727 & 0.0273 & 0 & 0 & 0 \\
0.0041 & 0.9806 & 0.0153 & 0 & 0 \\
0 & 0.0082 & 0.9837 & 0.0082 & 0 \\
0 & 0 & 0.0153 & 0.9806 & 0.0041 \\
0 & 0 & 0 & 0.0273 & 0.9727
\end{Bmatrix}
\tag{37}
$$

Note how the entries in the diagonal are close to 1 (the persistence of the continuous stochastic process is high) and that the probability of moving two or more positions is zero. It would take at least 4 quarters for the Markov chain to travel from z_1 to z_5 (and vice versa).

Tauchen's procedure can be extended to VAR processes instead of an AR process. This is convenient because we can always rewrite a general ARMA(p,q) process as a VAR(1) (and a VAR(p) as a VAR(1)) by changing the definition of the state variables. Furthermore, open source implementations of the procedure exist for all major programming languages.

Kopecky and Suen (2010) show that an alternative procedure proposed by Rouwenhorst (1995) is superior to Tauchen's method when ρ, the persistence of the stochastic process, is close to 1. The steps of Rouwenhorst (1995)'s procedure are:

Algorithm 3 (Alternative AR(1) Discretization)
1. Set n, the number of potential realizations of the process z.
2. Set the upper ($\bar{z}$) and lower ($\underline{z}$) bounds for the process. Let $\underline{z} = -\lambda$ and $\bar{z} = \lambda$. λ can be set to be $\lambda = \sqrt{n-1}\sigma_z$.
3. Set $\{z_i\}_{i=1}^n$ such that:

$$
z_i = \underline{z} + \frac{\bar{z} - \underline{z}}{n-1}(i-1)
$$

4. When $n = 2$, let P_2 be given by:

$$
P_2 = \begin{bmatrix} p & 1-p \\ 1-q & q \end{bmatrix}
$$

p, q can be set to be $p = q = \dfrac{1+\rho}{2}$.

5. For $n \geq 3$, construct recursively the transition matrix:

$$
P_n = p\begin{bmatrix} P_{n-1} & \mathbf{0} \\ \mathbf{0}' & 0 \end{bmatrix} + (1-p)\begin{bmatrix} \mathbf{0} & P_{n-1} \\ 0 & \mathbf{0}' \end{bmatrix} + (1-q)\begin{bmatrix} \mathbf{0}' & 0 \\ P_{n-1} & \mathbf{0} \end{bmatrix} + q\begin{bmatrix} 0 & \mathbf{0}' \\ \mathbf{0} & P_{n-1} \end{bmatrix}
$$

where $\mathbf{0}$ is an $(n-1) \times 1$ column vector of zeros. Divide all but the top and bottom rows by 2 so that the sum of the elements of each row is equal to 1. The final outcome is $P_{z,z'}$.

Once productivity has been discretized, we can search for

$$c(k, z_m) = d^{c,m,j}(k|\theta^{m,c}) = \sum_{i=0}^{j} \theta_i^{m,c} \psi_i(k)$$

$$k(k, z_m) = d^{k,m,j}(k|\theta^{m,k}) = \sum_{i=0}^{j} \theta_i^{m,k} \psi_i(k)$$

where $m = 1,\ldots,n$. That is, we search for decision rules for capital and consumption when productivity is z_1 today, decision rules for capital and consumption when productivity is z_2 today, and so on, for a total of $2 \times n$ decision rules. Since n is usually a small number (we mentioned above 5 or 7), the complexity of the problem is not exploding.

Note that since we substitute these decision rules in the Euler equation:

$$u'(c_t) = \beta \mathbb{E}_t \left[u'(c_{t+1}) \left(\alpha e^{z_{t+1}} k_{t+1}^{\alpha-1} + 1 - \delta \right) \right]. \tag{38}$$

to get:

$$u'\left(d^{c,m,j}(k|\theta^{m,c}) \right) =$$
$$\beta \sum_{l=0}^{n} p_{ml} \left[u'\left(d^{c,l,j}\left(d^{k,m,j}(k|\theta^{m,k}) | \theta^{l,c} \right) \right) \left(\alpha e^{z_{t+1}} \left(d^{k,m,j}(k|\theta^{m,k}) \right)_{t+1}^{\alpha-1} + 1 - \delta \right) \right]$$

we are still taking account of the fact that productivity can change in the next period (and hence, consumption and capital accumulation will be determined by the decision rule for the next period level of productivity). Also, since now the stochastic process is discrete, we can substitute the integral on the right-hand side of Eq. (38) for the much simpler sum operator with the probabilities from the transition matrix (35). Otherwise, we would need to use a quadrature method to evaluate the integral (see Judd, 1998 for the relevant formulae and the proposal in Judd et al., 2011a).

Thus, discretization of state variables such as the productivity shock is more often than not an excellent strategy to deal with multidimensional problems: simple, transparent, and not too burdensome computationally. Furthermore, we can discretize some of the state variables and apply the methods in the next paragraphs to deal with the remaining continuous state variables. In computation, mixing of strategies is often welcomed.

5.3.2.2 Tensors
Tensors build multidimensional basis functions by finding the Kronecker product of all unidimensional basis functions.[v] Imagine, for example, that we have two state variables,

---

[v] One should not confuse the tensors presented here with the tensor notation used for perturbation methods. While both situations deal with closely related mathematical objects, the key when we were dealing with perturbation was the convenience that tensor notation offered.

physical capital k_t and human capital h_t. We have three Chebyshev polynomials for each of these two state variables:

$$\psi_0^k(k_t), \psi_1^k(k_t), \text{ and } \psi_2^k(k_t)$$

and

$$\psi_0^h(h_t), \psi_1^h(h_t), \text{ and } \psi_2^h(h_t).$$

Then, the tensor is given by:

$$\psi_0^k(k_t)\psi_0^h(h_t), \psi_0^k(k_t)\psi_1^h(h_t), \psi_0^k(k_t)\psi_2^h(h_t),$$
$$\psi_1^k(k_t)\psi_0^h(h_t), \psi_1^k(k_t)\psi_1^h(h_t), \psi_1^k(k_t)\psi_2^h(h_t),$$
$$\psi_2^k(k_t)\psi_2^h(h_t), \psi_2^k(k_t)\psi_1^h(h_t), \text{ and } \psi_2^k(k_t)\psi_2^h(h_t).$$

More formally, imagine that we want to approximate a function of n state variables $d : [-1, 1]^n \to \mathbb{R}$ with Chebyshev polynomial of degree j. We build the sum:

$$d^j(\cdot \mid \theta) = \sum_{i_1=0}^{j} \cdots \sum_{i_n=0}^{j} \theta_{i_1,\ldots,i_n} \psi_{i_1}^1(\cdot) * \ldots * \psi_{i_n}^n(\cdot)$$

where $\psi_{i_\kappa}^\kappa$ is the Chebyshev polynomials of degree i_κ on the state variable κ and θ is the vector of coefficients $\theta_{i_1,\ldots,i_n}$. To make the presentation concise, we have made three simplifying assumptions. First, we are dealing with the case that d is one dimensional. Second, we are using the same number of Chebyshev polynomials for each state variable. Three, the functions $\psi_{i_\kappa}^\kappa$ could be different from the Chebyshev polynomials and belong to any basis we want (there can even be a different basis for each state variable). Eliminating these simplifications is straightforward, but notationally cumbersome.

There are two main advantages of a tensor basis. First, it is trivial to build. Second, if the one-dimensional basis is orthogonal, then the tensor basis is orthogonal in the product norm. The main disadvantage is the exponential growth in the number of coefficients $\theta_{i_1,\ldots,i_n}$: $(j+1)^n$. In the example above, even using only three Chebyshev polynomials (ie, $j = 2$) for each of these two state variables, we end up having to solve for nine coefficients. This curse of dimensionality is acute: with five state variables and three Chebyshev polynomials, we end up with 243 coefficients. With ten Chebyshev polynomials, we end up with 100,000 coefficients.

5.3.2.3 Complete Polynomials

In practice, it is infeasible to use tensors when we are dealing with models with more than 3 continuous state variables and a moderate j. A solution is to eliminate some elements of the tensor in a way that avoids much numerical degradation. In particular, Gaspar and Judd (1997) propose using the complete polynomials:

$$\mathcal{P}_\kappa^n \equiv \left\{ \psi_{i_1}^1 * \ldots * \psi_{i_n}^n \text{ with } |\mathbf{i}| \leq \kappa \right\}$$

where

$$|\mathbf{i}| = \sum_{l=1}^{n} i_l, \, 0 \leq i_1, \ldots, i_n.$$

Complete polynomials, instead of employing all the elements of the tensor, keep only those such that the sum of the order of the basis functions is less than a prefixed κ. The intuition is that the elements of the tensor $\psi_{i_1}^1 * \ldots * \psi_{i_n}^n, |\mathbf{i}| > \kappa$ add little additional information to the basis: most of the flexibility required to capture the behavior of d is already in the complete polynomials. For instance, if we are dealing with three state variables and Chebyshev polynomials $j = 4$, we can keep the complete polynomials of order 6:

$$\mathcal{P}_6^3 \equiv \left\{ \psi_{i_1}^1 * \ldots * \psi_{i_n}^n \text{ with } |\mathbf{i}| \leq 6 \right\}.$$

Complete polynomials eliminate many coefficients: in our example, instead of $(4 + 1)^3 = 125$ coefficients of the tensor, when $\kappa = 6$ we only need to approximate 87 coefficients. Unfortunately, we still need too many coefficients. In Section 5.7, we will present an alternative: Smolyak's algorithm. However, since the method requires the introduction of a fair amount of new notation and the presentation of the notion of interpolating polynomials, we postpone the discussion and, instead, start analyzing the finite element methods.

5.4 Finite Elements

Finite elements techniques, based on local basis functions, were popularized in economics by McGrattan (1996) (see, also, Hughes (2000), for more background, and Brenner and Scott (2008), for all the mathematical details that we are forced to skip in a handbook chapter). The main advantage of this class of basis functions is they can easily capture local behavior and achieve a tremendous level of accuracy even in the most challenging problems. That is why finite element methods are often used in mission-critical design in industry, such as in aerospace or nuclear power plant engineering. The main disadvantage of finite elements methods is that they are hard to code and expensive to compute. Therefore, we should choose this strategy when accuracy is more important than speed of computation or when we are dealing with complicated, irregular problems.

Finite elements start by bounding the domain Ω of the state variables. Some of the bounds would be natural (ie, $k_t > 0$). Other bounds are not ($k_t < \bar{k}$) and we need some care in picking them. For example, we can guess a $\bar{k}$ sufficiently large such that, in the simulations of the model, k_t never reaches $\bar{k}$. This needs, however, to be verified and some iterative fine-tuning may be required.[w]

[w] Even if the simulation rarely reaches $\bar{k}$, it may be useful to repeat the computation with a slightly higher bound $\omega\bar{k}$, with $\omega > 1$, to check that we still do not get to $\bar{k}$. In some rare cases, the first simulation might not have reached $\bar{k}$ because the approximation of the function $d(\cdot)$ precluded traveling into that region.

The second step in the finite elements method is to partition Ω into small, nonintersecting elements. These small sections are called elements (hence the name, "finite elements"). The boundaries of the elements are called nodes. The researcher enjoys a fantastic laxity in selecting the partition. One natural partition is to divide Ω into equal elements: simple and direct. But elements can be of unequal size. More concretely, we can have small elements in the areas of Ω where the economy will spend most of the time, while just a few large elements will cover areas of Ω infrequently visited (these areas can be guessed based on the theoretical properties of the model, or they can be verified by an iterative procedure of element partition; we will come back to this point below). Or we can have small elements in the areas of Ω where the function $d(\cdot)$ we are looking for changes quickly in shape, while we reserve large elements for areas of Ω where the function d is close to linear. Thanks to this flexibility in the element partition, we can handle kinks or constraints, which are harder to tackle with spectral methods (or next to impossible to do with perturbation, as they violate differentiability conditions).[x]

An illustration of such capability appears in Fig. 7, where we plot the domain Ω of a dynamic model of a firm with two state variables, bonds b_t on the x-axis (values to the right denote positive bond holdings by the firm and values to the left negative bond holdings), and capital k_t on the y-axis. The domain Ω does not include an area in the lower left corner, of combinations of negative bond holdings (ie, debt) and low capital. This area is excluded because of a financial constraint: firms cannot take large amounts of debt when
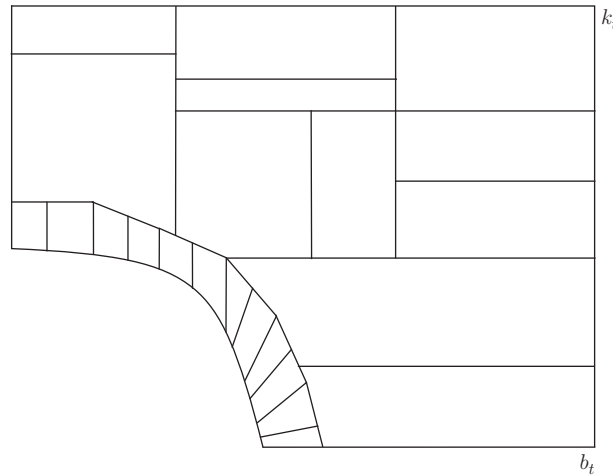

Fig. 7 Two-dimensional element grid.

[x] This flexibility in the definition of the elements is a main reason why finite elements methods are appreciated in industry, where applications often do not conform to the regularity technical conditions required by perturbation or spectral techniques.

they do not have enough capital to use as collateral (the concrete details of this financial constraint or why the shape of the restricted area is the one we draw are immaterial for the argument). In Fig. 7, the researcher has divided the domain Ω into unequal elements: there are many of them, of small size, close to the lower left corner boundary. One can suspect that the decision rule for the firm for b_t and k_t may change rapidly close to the frontier or, simply, the researcher wants to ensure the accuracy of the solution in that area. Farther away from the frontier, elements become larger. But even in those other regions, the researcher can partition the domain Ω with very different elements, some smaller (high levels of debt and k_t), some larger (high levels of b_t and k_t), depending on what the researcher knows about the shape of the decision rule.

There is a whole area of research concentrated on the optimal generation of an element grid that we do not have space to review. The interested reader can check Thompson et al. (1985). For a concrete application of unequal finite elements to the stochastic neoclassical growth model to reduce computational time, see Fernández-Villaverde and Rubio-Ramírez (2004).

The third step in the finite elements method is to choose a basis for the policy functions in each element. Since the elements of the partition of Ω are usually small, a linear basis is often good enough. For instance, letting $\{k_0, k_1, \ldots, k_j\}$ be the nodes of a partition of Ω into elements, we can define the tent functions for $i \in \{1, j-1\}$

$$
\psi_i(k) = \begin{cases} \dfrac{k - k_{i-1}}{k_i - k_{i-1}}, & \text{if } x \in [k_{i-1}, k_i] \\[2mm] \dfrac{k_{i+1} - k}{k_{i+1} - k_i}, & \text{if } k \in [k_i, k_{i+1}] \\[2mm] 0 & \text{elsewhere} \end{cases}
$$

and the corresponding adjustments for the first function:

$$
\psi_0(k) = \begin{cases} \dfrac{k_0 - k}{k_1 - k_0}, & \text{if } x \in [k_0, k_1] \\[2mm] 0 & \text{elsewhere} \end{cases}
$$

and the last one

$$
\psi_j(k) = \begin{cases} \dfrac{k - k_{j-1}}{k_j - k_{j-1}}, & \text{if } k \in [k_i, k_{i+1}] \\[2mm] 0 & \text{elsewhere.} \end{cases}
$$

We plot examples of these tent functions in Fig. 8.

We can extend this basis to higher dimensions by either discretizing some of the state variables (as we did when we talked about spectral bases) or by building tensors of them. Below, we will also see how to use Smolyak's algorithm with finite elements.
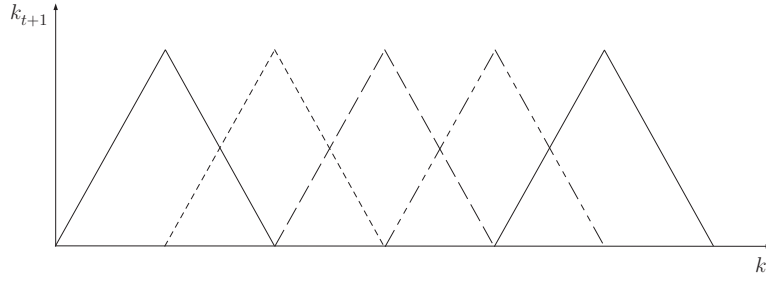
Fig. 8 Five basis functions.

The fourth step in the finite elements method is the same as for any other projection method: we build

$$d^{n,j}(\,\cdot\,|\theta^n) = \sum_{i=0}^{j} \theta_i^n \psi_i(\,\cdot\,)$$

and we plug them into the operator $\mathcal{H}$. Then, we find the unknown coefficients as we would do with Chebyshev polynomials.

By construction, the different parts of the approximating function will be pasted together to ensure continuity. For example, in our Fig. 8, there are two basis functions in the element defined by the nodes k_i and k_{i+1}

$$\psi_i(k) = \frac{k_{i+1} - k}{k_{i+1} - k_i}$$

$$\psi_{i+1}(k) = \frac{k - k_i}{k_{i+1} - k_i}$$

and their linear combination (ie, the value of $d^{n,j}(\,\cdot\,|\theta^n)$ in that element) is:

$$\hat{d}\left(k|k_{i+1}, k_i, \theta_{i+1}^n, \theta_i^n\right) = \theta_i^n \frac{k_{i+1} - k}{k_{i+1} - k_i} + \theta_{i+1}^n \frac{k - k_i}{k_{i+1} - k_i} = \frac{\left(\theta_{i+1}^n - \theta_i^n\right)k + \theta_i^n k_{i+1} - \theta_{i+1}^n k_i}{k_{i+1} - k_i},$$

which is a linear function, with positive or negative slope depending on the sign of $\theta_{i+1}^n - \theta_i^n$. Also note that the value of $d^{n,j}(\,\cdot\,|\theta^n)$ in the previous element is the linear function:

$$\hat{d}\left(k|k_i, k_{i-1}, \theta_i^n, \theta_{i-1}^n\right) = \frac{\left(\theta_i^n - \theta_{i-1}^n\right)k + \theta_{i-1}^n k_i - \theta_i^n k_{i-1}}{k_i - k_{i-1}}.$$

When we evaluate both linear functions at k_i

$$\hat{d}\left(k_i|k_i, k_{i-1}, \theta_i^n, \theta_{i-1}^n\right) = \theta_i^n$$
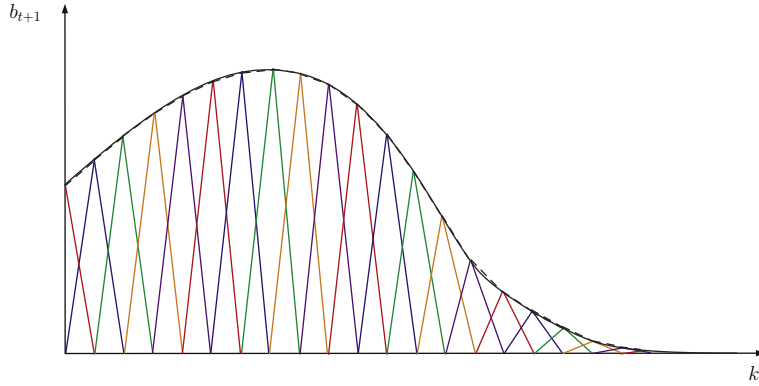
Fig. 9 Finite element approximation.

and

$$\hat{d}\left(k_i | k_{i+1}, k_i, \theta_{i+1}^n, \theta_i^n\right) = \theta_i^n$$

that is, both functions have the same value equal to the coefficient θ_i^n, which ensures continuity (although, with only tent functions, we cannot deliver differentiability).

The previous derivation also shows why finite elements are a smart strategy. Imagine that our metric ρ is such that we want to make the residual function equal to zero in the nodes of the elements (below we will present a metric like this one). With our tent functions, this amounts to picking, at each k_i, the coefficient θ_i^n such that the approximating and exact function coincide:

$$d^{n,j}(\cdot | \theta^n) = d^n(\cdot).$$

This implies that the value of d^n outside k_i are irrelevant for our choice of θ_i^n. An example of such piecewise linear approximation to a decision rule for the level of debt tomorrow, b_{t+1}, given capital today, k_t, in a model of financial frictions, is drawn in Fig. 9. The discontinuous line is the approximated decision rule and the continuous line the exact one. The tent functions are multiplied by the coefficients to make the approximation and the exact solution equal at the node points. We can appreciate an already high level of accuracy. As the elements become smaller and smaller, the approximation will become even more accurate (ie, smooth functions are locally linear).

This is a stark example of a more general point: the large system of nonlinear equations that we will need to solve in a finite element method will be sparse, a property that can be suitably exploited by modern nonlinear solvers.

Remark 22 (Finite elements method refinements) An advantage of the finite elements method is that we can refine the solution that we obtain as much as we desire (with only the constraints of computational time and memory). The literature distinguishes among

three different refinements. First, we have the *h-refinement*. This scheme subdivides each element into smaller elements to improve resolution uniformly over the domain. That is, once we have obtained a first solution, we check whether this solution achieves the desired level of accuracy. If it does not, we go back to our partition, and we subdivide the elements. We can iterate in this procedure as often as we need. Second, we have *r-refinement*: This scheme subdivides each element only in those regions where there are high nonlinearities. Third, we have the *p-refinement*: This scheme increases the order of the approximation in each element, that is, it adds more basis functions (for example, several Chebyshev polynomials). If the order of the expansion is high enough, we generate a hybrid of finite and spectral methods known as spectral elements. This approach has gained much popularity in the natural sciences and engineering. See, for example, Solín et al. (2004).

Sometimes, *h-refinements* and *p-refinements* are mixed in what is known as the hp-finite element method, which delivers exponential convergence to the exact solution. Although difficult to code and computationally expensive, an hp-finite element method is, perhaps, the most powerful solution technique available for DSGE models, as it can tackle even the most challenging problems.[y]

The three refinements can be automatically implemented: we can code the finite element algorithm to identify the regions of Ω where, according to some goal of interest (for example, how tightly a Euler equation is satisfied), we refine the approximation without further input from the researcher. See Demkowicz (2007).

5.5 Objective Functions

Our second choice is to select a metric function ρ to determine how we "project." The most common answer to this question is given by a *weighted residual*: we select θ to get the residual close to 0 in the weighted integral sense. Since we did not impose much structure on the operator $\mathcal{H}$ and therefore, on the residual function $R(\cdot|\theta)$, we will deal with the simplest case where $R(\cdot|\theta)$ is unidimensional. More general cases can be dealt with at the cost of heavier notation. Given some weight functions $\phi_i : \Omega \to \mathbb{R}$, we define the metric:

$$\rho(R(\cdot|\theta),0) = \begin{cases} 0 & \text{if } \int_\Omega \phi_i(\mathbf{x})R(\cdot|\theta)d\mathbf{x} = 0, i = 1,..,j+1 \\ 1 & \text{otherwise} \end{cases}$$

---

[y] An additional, new refinement is the extended finite element method (x-fem), which adds to the basis discontinuous functions that can help in capturing irregularities in the solution. We are not aware of applications of the x-fem in economics.

Hence, the problem is to choose the θ that solves the system of integral equations:

$$\int_{\Omega} \phi_i(\mathbf{x}) R(\cdot|\theta) d\mathbf{x} = 0, \, i = 1, .., j + 1. \tag{39}$$

Note that, for the system to have a solution, we need $j + 1$ weight functions. Thanks to the combination of approximating the function d by basis functions ψ_i and the definition of weight functions ϕ_i, we have transformed a rather intractable functional equation problem into a standard nonlinear equations system. The solution of this system can be found using standard methods, such as a Newton algorithm for small problems or a Levenberg–Marquardt method for bigger ones.

However, the system (39) may have no solution or it may have multiple ones. We know very little about the theoretical properties of projection methods in economic applications. The literature in applied mathematics was developed for the natural sciences and engineering and many of the technical conditions required for existence and convergence theorems to work do not easily travel across disciplines. In fact, some care must be put into ensuring that the solution of the system (39) satisfies the transversality conditions of the DSGE model (ie, we are picking the stable manifold). This can usually be achieved with the right choice of an initial guess θ_0 or by adding boundary conditions to the solver.

As was the case with the bases, we will have plenty of choices for our weight functions. Instead of reviewing all possible alternatives, we will focus on the most popular ones in economics.

5.5.1 Weight Function I: Least Squares

Least squares use as weight functions the derivatives of the residual function:

$$\phi_i(\mathbf{x}) = \frac{\partial R(\mathbf{x}|\theta)}{\partial \theta_{i-1}}$$

for all $i \in 1, .., j + 1$. This choice is motivated by the variational problem:

$$\min_{\theta} \int_{\Omega} R^2(\cdot|\theta) d\mathbf{x}$$

with first–order condition:

$$\int_{\Omega} \frac{\partial R(\mathbf{x}|\theta)}{\partial \theta_{i-1}} R(\cdot|\theta) d\mathbf{x} = 0, \, i = 1, .., j + 1.$$

This variational problem is mathematically equivalent to a standard regression problem in econometrics.

While least squares are intuitive and there are algorithms that exploit some of their structure to increase speed and decrease memory requirements, they require the computation of the derivative of the residual, which can be costly. Also, least squares problems are often ill–conditioned and complicated to solve numerically.

5.5.2 Weight Function II: Subdomain

The subdomain approach divides the domain Ω into $1,..,j+1$ subdomains Ω_i and define the $j+1$ step functions:

$$\phi_i(\mathbf{x}) = \begin{cases} 1 & \text{if } \mathbf{x} \in \Omega_i \\ 0 & \text{otherwise} \end{cases}$$

This choice is equivalent to solving the system:

$$\int_{\Omega_i} R(\cdot|\theta)d\mathbf{x} = 0, \, i = 1,..,j+1.$$

The researcher has plenty of flexibility to pick her subdomains as to satisfy her criteria of interest.

5.5.3 Weight Function III: Collocation

This method is also known as pseudospectral or the method of selected points. It defines the weight function as:

$$\phi_i(\mathbf{x}) = \delta(\mathbf{x} - \mathbf{x}_i)$$

where δ is the Dirac delta function and $\mathbf{x}_i$ are the $j+1$ collocation points selected by the researcher.

This method implies that the residual function is zero at the n collocation points. Thus, instead of having to compute complicated integrals, we only need to solve the system:

$$R(\mathbf{x}_i|\theta) = 0, \, i = 1,..,j+1.$$

This is attractive when the operator $\mathcal{H}$ generates large nonlinearities.

A systematic way to pick collocation points is to use the zeros of the $(j+1)$-th-order Chebyshev polynomial in each dimension of the state variable (or the corresponding polynomials, if we are using different approximation orders along each dimension). This approach is known as orthogonal collocation. The Chebyshev interpolation theorem tells us that, with this choice of collocation points, we can achieve $\mathcal{L}_p$ convergence and sometimes even uniform convergence to the unknown function d. Another possibility is to pick, as collocation points, the extrema of the j-th-order Chebyshev polynomial in each dimension. Experience shows a surprisingly good performance of orthogonal collocation methods and it is one of our recommended approaches.

5.5.4 Weight Function IV: Galerkin or Rayleigh–Ritz

The last weight function we consider is the Galerkin (also called Rayleigh–Ritz when it satisfies some additional properties of less importance for economists). This approach takes as the weight function the basis functions used in the approximation:

$$\phi_i(\mathbf{x}) = \psi_{i-1}(\mathbf{x}).$$

Then we have:

$$\int_{\Omega} \psi_i(\mathbf{x}) R(\cdot|\theta) d\mathbf{x} = 0, \, i = 1, .., j+1.$$

The interpretation is that the residual has to be orthogonal to each of the basis functions.

The Galerkin approach is highly accurate and robust, but difficult to code. If the basis functions are complete over J_1 (they are indeed a basis of the space), then the Galerkin solution will converge pointwise to the true solution as n goes to infinity:

$$\lim_{j \to \infty} d^j(\cdot|\theta) = d(\cdot)$$

Also, practical experience suggests that a Galerkin approximation of order j is as accurate as a pseudospectral $j + 1$ or $j + 2$ expansion.

In the next two remarks, we provide some hints for a faster and more robust solution of the system of nonlinear equations:

$$\int_{\Omega} \phi_i(\mathbf{x}) R(\cdot|\theta) d\mathbf{x} = 0, \, i = 1, .., j+1, \tag{40}$$

a task that can be difficult if the number of coefficients is large and the researcher does not have a good initial guess θ_0 for the solver.

Remark 23 (Transformations of the problem) A bottleneck for the solution of (39) can be the presence of strong nonlinearities. Fortunately, it is often the case that simple changes in the problem can reduce these nonlinearities. For example, Judd (1992) proposes that if we have an Euler equation:

$$\frac{1}{c_t} = \beta \mathbb{E}_t \left\{ \frac{1}{c_{t+1}} R_{t+1} \right\}$$

where R_{t+1} is the gross return rate of capital, we can take its inverse:

$$\beta c_t = \left(\mathbb{E}_t \left\{ \frac{1}{c_{t+1}} R_{t+1} \right\} \right)^{-1},$$

which now is linear on the left-hand side and much closer to linear on the right-hand side. Thus, instead of computing the residual for some state variable x_t

$$R(\cdot|\theta) = \frac{1}{c(x_t|\theta)} - \beta \mathbb{E}_t \left\{ \frac{1}{c(x_t|\theta)} R_{t+1}(x_t|\theta) \right\},$$

we compute:

$$\tilde{R}(\cdot|\theta) = \beta c(x_t|\theta) - \left(\mathbb{E}_t \left\{ \frac{1}{c(x_t|\theta)} R_{t+1}(x_t|\theta) \right\} \right)^{-1}.$$

Similar algebraic manipulations are possible in many DSGE models.

Remark 24 (Multistep schemes) The system (39) can involve a large number of coefficients. A natural strategy is to solve first a smaller system and to use that solution as an input for a larger system. This strategy, called a multistep scheme, often delivers excellent results, in particular when dealing with orthogonal bases such as Chebyshev polynomials.

More concretely, instead of solving the system for an approximation with $j + 1$ basis functions, we can start by solving the system with only $j' + 1 \ll j + 1$ basis functions and use the solution to this first problem as a guess for the more complicated problem. For example, if we are searching for a solution with 10 Chebyshev polynomials and m dimensions, we first find the approximation with only 3 Chebyshev polynomials. Therefore, instead of solving a system of $10 \times m$ equations, we solve a system of $3 \times m$. Once we have the solution θ^3, we build the initial guess for the problem with 10 Chebyshev polynomials as:

$$\theta_0 = \left[\theta^3, \mathbf{0}_{1 \times m}, \ldots, \mathbf{0}_{1 \times m}\right],$$

that is, we use θ^3 for the first coefficients and zero for the additional new coefficients. Since the additional polynomials are orthogonal to the previous ones, the final values of the coefficients associated with the three first polynomials will change little with the addition of 7 more polynomials: the initial guess θ^3 is, thus, most splendid. Also, given the fast convergence of Chebyshev polynomials, the coefficients associated with higher-order polynomials will be close to zero. Therefore, our initial guess for those coefficients is also informative.

The researcher can use as many steps as she needs. By judiciously coding the projection solver, the researcher can write the program as depending on an abstract number of Chebyshev polynomials. Then, she can call the solver inside a loop and iteratively increase the level of approximation from j' to j as slow or as fast as required.

5.6 A Worked-Out Example

We present now a worked-out example of how to implement a projection method in a DSGE model. In particular, we will use Chebyshev polynomials and orthogonal collocation to solve the stochastic neoclassical growth model with endogenous labor supply.

In this economy, there is a representative household, whose preferences over consumption, c_t, and leisure, $1 - l_t$, are representable by the utility function:

$$\mathbb{E}_0 \sum_{t=1}^{\infty} \beta^{t-1} \frac{\left(c_t^\tau (1 - l_t)^{1-\tau}\right)^{1-\eta}}{1 - \eta}$$

where $\beta \in (0, 1)$ is the discount factor, η controls the elasticity of intertemporal substitution and risk aversion, τ controls labor supply, and $\mathbb{E}_0$ is the conditional expectation operator.

There is one good in the economy, produced according to the aggregate production function:

$$y_t = e^{z_t} k_t^\alpha l_t^{1-\alpha}$$

where k_t is the aggregate capital stock, l_t is aggregate labor, and z_t is a stochastic process for technology:

$$z_t = \rho z_{t-1} + \epsilon_t$$

with $|\rho| < 1$ and $\epsilon_t \sim N(0, \sigma^2)$. Capital evolves according to:

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

and the economy must satisfy the resource constraint $y_t = c_t + i_t$.

Since both welfare theorems hold in this economy, we solve directly for the social planner's problem:

$$V(k_t, z_t) = \max_{c_t, l_t} \frac{\left(c_t^\tau (1-l_t)^{1-\tau}\right)^{1-\eta}}{1-\eta} + \beta \mathbb{E}_t V(k_{t+1}, z_{t+1})$$
$$\text{s.t. } k_{t+1} = e^{z_t} k_t^\alpha l_t^{1-\alpha} + (1-\delta)k_t - c_t$$
$$z_t = \rho z_{t-1} + \epsilon_t$$

given some initial conditions k_0 and z_0. Tackling the social planner's problem is only done for convenience, and we could also solve for the competitive equilibrium. In fact, one key advantage of projection methods is that they easily handle non–Pareto efficient economies.

We calibrate the model with standard parameter values to match US quarterly data (see Table 2). The only exception is η, for which we pick a value of 5, in the higher range of empirical estimates. Such high–risk aversion induces, through precautionary behavior, more curvature in the decision rules. This curvature would present a more challenging test bed for the projection method.

We discretize z_t into a 5-point Markov chain $\{z_1, \ldots, z_5\}$ using Tauchen's procedure and covering ± 3 unconditional standard deviations of z_t (this is the same Markov chain as

Table 2 Calibration

Parameter	Value
β	0.991
η	5.000
τ	0.357
α	0.300
δ	.0196
ρ	0.950
σ	0.007

the example in Remark 21, see (36) and (37) for the concrete values of the discretization). We will use p_{mn} to denote the generic entry of the transition matrix $P_{z,z'}$ generated by Tauchen's procedure for z_m today moving to z_n next period.

Then, we approximate the value function $V^j(k_t)$ and the decision rule for labor, $l^j(k_t)$, for $j = 1,\ldots,5$ using 11 Chebyshev polynomials as:

$$V^j\left(k_t|\theta^{V,j}\right) = \sum_{i=0}^{10} \theta_i^{V,j} T_i(k_t) \tag{41}$$

$$l^j\left(k_t|\theta^{l,j}\right) = \sum_{i=0}^{10} \theta_i^{l,k} T_i(k_t) \tag{42}$$

Once we have the decision rule for labor, we can find output:

$$y^j(k_t) = e^{z_t} k_t^\alpha \left(l^j\left(k_t|\theta^{l,j}\right)\right)^{1-\alpha},$$

With output, from the first-order condition that relates the marginal utility consumption and the marginal productivity of labor, we can find consumption:

$$c^j(k_t) = \frac{\tau}{1-\tau}(1-\alpha)e^{z_t} k_t^\alpha \left(l^j\left(k_t|\theta^{l,j}\right)\right)^{-\alpha}\left(1 - l^j\left(k_t|\theta^{l,j}\right)\right) \tag{43}$$

and, from the resource constraint, capital next period:

$$k^j(k_t) = e^{z_t} k_t^\alpha \left(l^j\left(k_t|\theta^{l,j}\right)\right)^{1-\alpha} + (1-\delta)k_t - c^j(k_t) \tag{44}$$

Our notations $y^j(k_t)$, $c^j(k_t)$, and $k^j(k_t)$ emphasize the exact dependence of these three variables on capital and the productivity level: once we have approximated $l^j\left(k_t|\theta^{l,j}\right)$, simple algebra with the equilibrium conditions allows us to avoid further approximation.

We decided to approximate the value function and the decision rule for labor and use them to derive the other variables of interest to illustrate how flexible projection methods are. We could, as well, have decided to approximate the decision rules for consumption and capital and find labor and the value function using the equilibrium conditions. The researcher should pick the approximating functions that are more convenient, either for algebraic reasons or her particular goals.

To solve for the unknown coefficients θ^V and θ^l, we plug the functions (41), (42), (43), and (44) into the Bellman equation to get:

$$\sum_{i=0}^{10} \theta_i^{V,j} T_i(k_t) = \frac{\left(\left(c^j(k_t)\right)^\theta \left(1 - \sum_{i=0}^{10}\theta_i^l T_i(k_t)\right)^{1-\theta}\right)^{1-\tau}}{1-\tau} + \beta\sum_{m=1}^{5} p_{jm} \sum_{i=0}^{10} \theta_i^{V,j} T_i\left(k^j(k_t)\right) \tag{45}$$

where, since we are already using the optimal decision rules, we can drop the max operator. Also, we have substituted the expectation by the sum operator and the transition

probabilities p_{jm}. We plug the same functions (41), (42), (43), and (44) into the Euler equation to get:

$$\frac{\left(c_t^\theta \left(1 - \sum_{i=0}^{10} \theta_i^{l,k} T_i(k_t)\right)^{1-\theta}\right)^{1-\tau}}{c_t} = \beta \mathbb{E}_t \sum_{m=1}^{5} p_{jm} \sum_{i=0}^{10} \theta_i^{V,j} T_i\big(k^j(k_t)\big)\prime, \tag{46}$$

where $T_i(k^j(k_t))\prime$ is the derivative of the Chebyshev polynomial with respect to its argument.

The residual equation groups Eqs. (45) and (46):

$$R(k_t, z_j | \theta) = \begin{cases} \sum_{i=0}^{10} \theta_i^{V,j} T_i(k_t) - \dfrac{\left(\left(c^j(k_t)\right)^\theta \left(1 - \sum_{i=0}^{10} \theta_i^l T_i(k_t)\right)^{1-\theta}\right)^{1-\tau}}{1-\tau} \\[2mm] \qquad\qquad -\beta \sum_{m=1}^{5} p_{jm} \sum_{i=0}^{10} \theta_i^{V,j} T_i\big(k^j(k_t)\big) \\[4mm] \dfrac{\left(c_t^\theta \left(1 - \sum_{i=0}^{10} \theta_i^{l,k} T_i(k_t)\right)^{1-\theta}\right)^{1-\tau}}{c_t} - \beta \mathbb{E}_t \sum_{m=1}^{5} p_{jm} \sum_{i=0}^{10} \theta_i^{V,j} T_i\big(k^j(k_t)\big)\prime \end{cases}$$

where θ stacks $\theta^{V,j}$ and $\theta^{l,k}$. Given that we use 11 Chebyshev polynomials for the value function and another 11 for the decision rule for labor for each of the 5 levels of z_j, θ has 110 elements ($110 = 11 \times 2 \times 5$). If we evaluate the residual function at each of the 11 zeros of the Chebyshev of order 11 for capital and the 5 levels of z_j, we will have the 110 equations required to solve for those 110 coefficients. A Newton solver can easily deal with this system (although, as explained in Remark 24, using a multistep approach simplifies the computation: we used 3 Chebyshev polynomials in the first step and 11 Chebyshev polynomials in the second one).

We plot the main components of the solution in Fig. 10. The top left panel draws the value function, with one line for each of the five values of productivity and capital on the *x*-axis. As predicted by theory, the value function is increasing and concave in both state variables, k_t and z_t. We follow the same convention for the decision rules for consumption (top right panel), labor supply (bottom left panel), and capital next period, k_{t+1} (bottom right panel). The most noticeable pattern is the near linearity of the capital decision rule. Once the researcher has found the value function and all the decision rules, she can easily simulate the model, compute impulse response functions, and evaluate welfare.

The accuracy of the solution is impressive, with Euler equation errors below –13 in the $\log_{10}$ scale. Section 7 discusses how to interpret these errors. Suffice it to say here that, for practical purposes, the solution plotted in Fig. 10 can be used instead of the exact solution of the stochastic neoclassical growth model with a discrete productivity level.
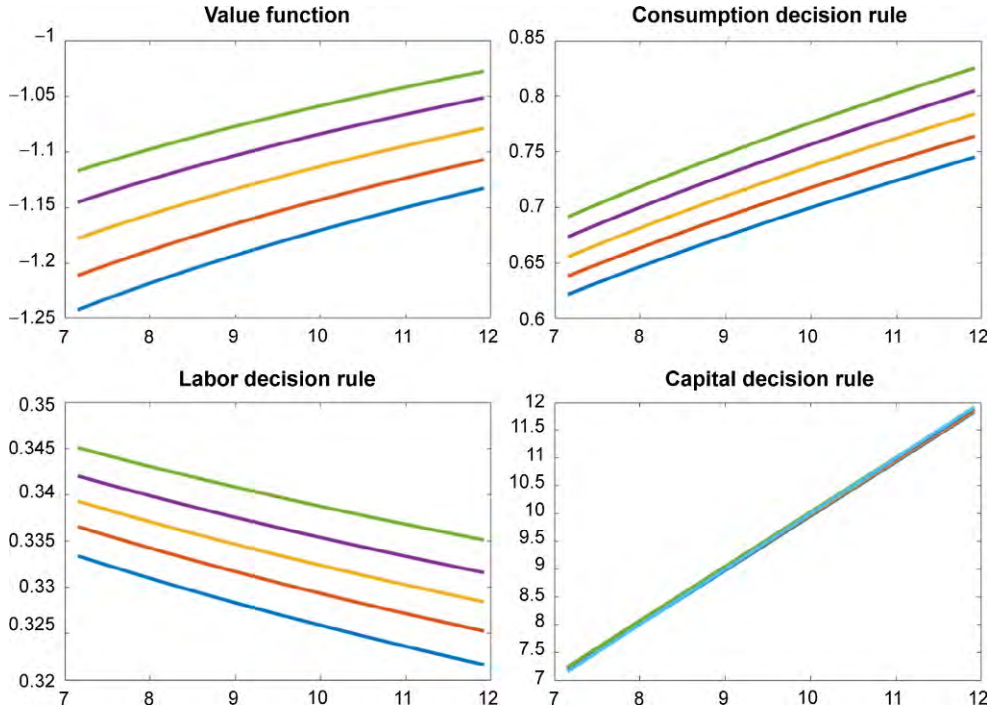
Fig. 10 Solution, stochastic neoclassical growth model.

5.7 Smolyak's Algorithm

An alternative to complete polynomials that can handle the curse of dimensionality better than other methods is Smolyak's algorithm. See Smolyak (1963), Delvos (1982), Barthelmann et al. (2000), and, especially, Bungartz and Griebel (2004) for a summary of the literature. Krüger and Kubler (2004) and Malin et al. (2011) introduced the algorithm in economics as a solution method for DSGE models. Subsequently, Smolyak's algorithm has been applied by many researchers. For example, Fernández-Villaverde et al. (2015a) rely on Smolyak's algorithm to solve a New Keynesian model with a ZLB (a model with 5 state variables), Fernández-Villaverde and Levintal (2016) exploit it to solve a New Keynesian model with big disasters risk (a model with 12 state variables), and Gordon (2011) uses it to solve a model with heterogeneous agents. Malin et al. (2011) can accurately compute a model with 20 continuous state variables and a considerable deal of curvature in the production and utility functions. In the next pages, we closely follow the explanations in Krüger and Kubler (2004) and Malin et al. (2011) and invite the reader to check those papers for further details.[z]

---

[z] There is also a promising line of research based on the use of ergodic sets to solve highly dimensional models (Judd et al., 2011b; Maliar et al., 2011; and Maliar and Maliar, 2015). Maliar and Maliar (2014) cover the material better than we could.

As before, we want to approximate a function (decision rule, value function, expectation, etc.) on n state variables, $d : [-1,1]^n \to \mathbb{R}$ (the generalization to the case $d : [-1,1]^n \to \mathbb{R}^m$ is straightforward, but tedious). The idea of Smolyak's algorithm is to find a grid of points $\mathbb{G}(q,n) \in [-1,1]^n$ where $q > n$ and an approximating function $d(x|\theta, q, n) : [-1,1]^n \to \mathbb{R}$ indexed by some coefficients θ such that, at the points $x_i \in \mathbb{G}(q,n)$, the unknown function $d(\cdot)$ and $d(\cdot|\theta, q, n)$ are equal:

$$d(x_i) = d(x_i|\theta, q, n)$$

and, at the points $x_i \notin \mathbb{G}(q,n)$, $d(\cdot|\theta, q, n)$ is close to the unknown function $d(\cdot)$. In other words, at the points $x_i \in \mathbb{G}(q, n)$, the operator $\mathcal{H}(\cdot)$ would be exactly satisfied and, at other points, the residual function would be close to zero. The integer q indexes the size of the grid and, with it, the precision of the approximation.

The challenge is to judiciously select grid points $\mathbb{G}(q,n)$ in such a way that the number of coefficients θ does not explode with n. Smolyak's algorithm is (almost) optimal for that task within the set of polynomial approximations (Barthelmann et al., 2000). Also, the method is universal, that is, almost optimal for many different function spaces.

5.7.1 Implementing Smolyak's Algorithm

Our search of a grid of points $\mathbb{G}(q,n)$ and a function $d(x|\theta,q,n)$ will proceed in several steps.

5.7.1.1 First Step: Transform the Domain of the State Variables

For any state variable $\tilde{x}_l$, $l = 1,\ldots,n$ that has a domain $[a,b]$, we use a linear translation from $[a,b]$ into $[-1,1]$:

$$x_l = 2\frac{\tilde{x}_l - a}{b - a} - 1.$$

5.7.1.2 Second Step: Setting the Order of the Polynomial

We define $m_1 = 1$ and $m_i = 2^{i-1} + 1$, $i = 2,\ldots,$ where $m_i - 1$ will be the order of the polynomial that we will use to approximate $d(\cdot)$.

5.7.1.3 Third Step: Building the Gauss–Lobotto Nodes

We build the sets:

$$\mathcal{G}^i = \{\zeta_1^i, \ldots, \zeta_{m_i}^i\} \subset [-1,1]$$

that contain the Gauss–Lobotto nodes (also known as the Clenshaw–Curtis points), that is, the extrema of the Chebyshev polynomials:

$$\zeta_j^i = -\cos\left(\frac{j-1}{m_i-1}\pi\right), j = 1,\ldots,m_i$$

with the initial set $\mathcal{G}^1 = \{0\}$ (with a change of notation, this formula for the extrema is the same as the one in Eq. (34)). For instance, the first three sets are given by:

$$\mathcal{G}^1 = \{0\}, \text{ where } i = 1, m_1 = 1.$$

$$\mathcal{G}^2 = \{-1, 0, 1\}, \text{ where } i = 2, m_3 = 3.$$

$$\mathcal{G}^3 = \left\{-1, -\cos\left(\frac{\pi}{4}\right), 0, -\cos\left(\frac{3\pi}{4}\right), 1\right\}, \text{ where } i = 3, m_5 = 5.$$

Since, in the construction of the sets, we impose that $m_i = 2^{i-1} + 1$, we generate sets that are nested, that is, $\mathcal{G}^i \subset \mathcal{G}^{i+1}, \forall i = 1,2,\ldots$. This result is crucial for the success of the algorithm.

5.7.1.4 Fourth Step: Building a Sparse Grid

For any integer q bigger than the number of state variables n, $q > n$, we define a sparse grid as the union of the Cartesian products:

$$\mathbb{G}(q,n) = \bigcup_{q-n+1 \leq |\mathbf{i}| \leq q} (\mathcal{G}^{i_1} \times \ldots \times \mathcal{G}^{i_n}),$$

where $|\mathbf{i}| = \sum_{l=1}^{n} i_l$.

To illustrate how this sparse grid works, imagine that we are dealing with a DSGE model with two continuous state variables. If we pick $q = 2 + 1 = 3$, we have the sparse grid

$$\mathbb{G}(3,2) = \bigcup_{2 \leq |\mathbf{i}| \leq 3} (\mathcal{G}^{i_1} \times \mathcal{G}^{i_2})$$

$$= (\mathcal{G}^1 \times \mathcal{G}^1) \cup (\mathcal{G}^1 \times \mathcal{G}^2) \cup (\mathcal{G}^2 \times \mathcal{G}^1)$$

$$= \{(-1,0), (0,1), (0,0), (0,-1), (1,0)\}$$

We plot this grid in the top left panel of Fig. 11, which reproduces fig. 1 in Krüger and Kubler (2004).

If we pick $q = 2 + 2 = 4$, we have the sparse grid

$$\mathbb{G}(4,2) = \bigcup_{3 \leq |\mathbf{i}| \leq 4} (\mathcal{G}^{i_1} \times \mathcal{G}^{i_2})$$

$$= (\mathcal{G}^1 \times \mathcal{G}^2) \cup (\mathcal{G}^1 \times \mathcal{G}^3) \cup (\mathcal{G}^2 \times \mathcal{G}^2) \cup (\mathcal{G}^3 \times \mathcal{G}^1)$$

$$= \left\{ \begin{array}{l} (-1,1), (-1,0), (-1,-1), \left(-\cos\left(\frac{\pi}{4}\right),0\right), \\ (0,1), \left(0,-\cos\left(\frac{3\pi}{4}\right)\right), (0,0), \left(0,-\cos\left(\frac{\pi}{4}\right)\right), \\ (0,-1), \left(-\cos\left(\frac{3\pi}{4}\right),0\right), (1,1), (1,0), (1,-1) \end{array} \right\}$$
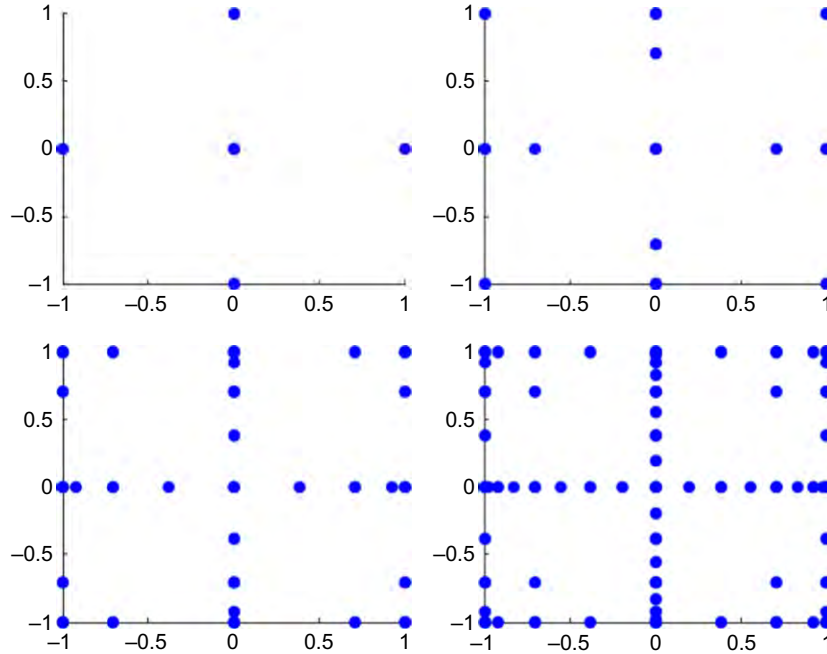
Fig. 11 Four sparse grids

We plot this grid in the top right panel of Fig. 11. Note that the sparse grids have a hierarchical structure, where $\mathbb{G}(3,2) \in \mathbb{G}(4,2)$ or, more generally, $\mathbb{G}(q,n) \in \mathbb{G}(q+1,n)$.

Following the same strategy, we can build $\mathbb{G}(5,2)$, plotted in the bottom left panel of Fig. 11, and $\mathbb{G}(6,2)$, plotted in the bottom right panel of Fig. 11 (in the interest of concision, we skip the explicit enumeration of the points of these two additional grids). In Fig. 12, we plot a grid for a problem with 3 state variables, $\mathbb{G}(5,3)$.

The sparse grid has two important properties. First, the grid points cluster around the corners of the domain of the Chebyshev polynomials and the central cross. Second, the number of points in a sparse grid when $q = n + 2$ is given by $1 + 4n + 2n(n-1)$. The cardinality of this grid grows polynomially on n^2. Similar formulae hold for other $q > n$. For example, the cardinality of the grid grows polynomially on n^3 when $q = n + 3$. In fact, the computational burden of the method notably increases as we keep n fixed and a rise q. Fortunately, experience suggests that $q = n + 2$ and $q = n + 3$ are usually enough to deliver the desired accuracy in DSGE models.

The nestedness of the sets of the Gauss–Lobotto nodes plays a central role in controlling the cardinality of $\mathbb{G}(q,n)$. In comparison, the number of points in a rect-angular grid is 5^n, an integer that grows exponentially on n. If $n = 2$, this would correspond, in the top right panel of Fig. 11, to having all possible tensors of
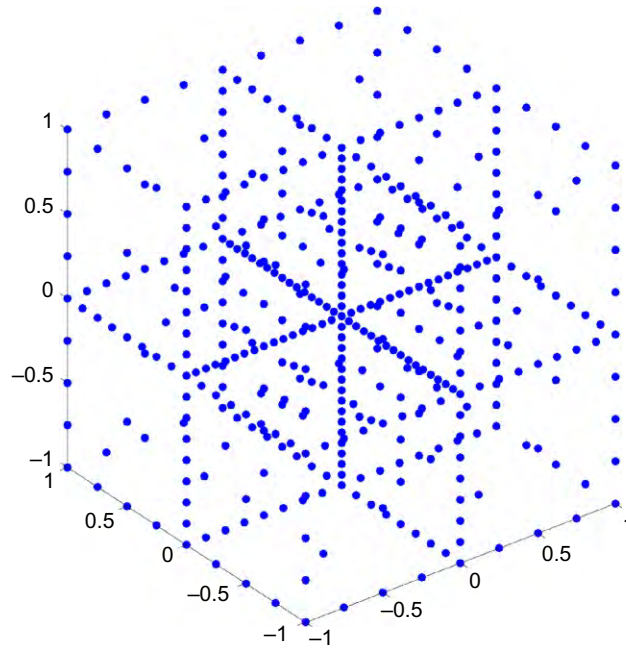
Fig. 12 A sparse grid, 3 state variables.

Table 3 Size of the grid for $q = n + 2$

n	$\mathbb{G}(q,n)$	5^n
2	13	25
3	25	125
4	41	625
5	61	3125
12	313	244,140,625

$\left\{ -1, -\cos\left(\frac{\pi}{4}\right), 0, -\cos\left(\frac{3\pi}{4}\right), 1 \right\}$ and $\left\{ -1, -\cos\left(\frac{\pi}{4}\right), 0, -\cos\left(\frac{3\pi}{4}\right), 1 \right\}$ covering the whole of the $[-1,1]^2$ square. Instead of keeping these 25 points, Smolyak's algorithm eliminates 12 of them and only keeps 13. To illustrate how dramatic is the difference between polynomial and exponential growth, Table 3 shows the cardinality of both grids as we move from 2 state variables to 12.

5.7.1.5 Fifth Step: Building Tensor Products

We use the Chebyshev polynomials $\psi_i(x_i) = T_{i-1}(x_i)$ to build the tensor-product multivariate polynomial:

$$p^{|\mathbf{i}|}(x|\theta) = \sum_{l_1=1}^{m_{i_1}} \cdots \sum_{l_n=1}^{m_{i_n}} \theta_{l_1 \ldots l_n} \psi_{l_1}(x_1) \ldots \psi_{l_n}(x_n)$$

where $|\mathbf{i}| = \sum_{l=1}^{n} i_l$, $x_i \in [-1,1]$, $x = \{x_1, \ldots, x_n\}$, and θ stacks all the coefficients $\theta_{l_1 \ldots l_n}$. So, for example, for a DSGE model with two continuous state variables and $q = 3$, we will have:

$$p^{\mathbf{1,1}}(x|\theta) = \sum_{l_1=1}^{m_1} \sum_{l_n=1}^{m_1} \theta_{l_1 l_2} \psi_{l_1}(x_1) \psi_{l_2}(x_2) = \theta_{11}$$

$$p^{\mathbf{1,2}}(x|\theta) = \sum_{l_1=1}^{m_1} \sum_{l_n=1}^{m_2} \theta_{l_1 l_2} \psi_{l_1}(x_1) \psi_{l_2}(x_2) = \theta_{11} + \theta_{12} T_1(x_2) + \theta_{13} T_2(x_2)$$

$$p^{\mathbf{2,1}}(x|\theta) = \sum_{l_1=1}^{m_2} \sum_{l_n=1}^{m_1} \theta_{l_1 l_2} \psi_{l_1}(x_1) \psi_{l_2}(x_2) = \theta_{11} + \theta_{21} T_1(x_1) + \theta_{31} T_2(x_1)$$

where we have already used $T_0(x_i) = 1$. Therefore, for $x = \{x_1, x_2\}$:

$$p^{|2|}(x|\theta) = p^{\mathbf{1,1}}(x|\theta)$$
$$p^{|3|}(x|\theta) = p^{\mathbf{1,2}}(x|\theta) + p^{\mathbf{2,1}}(x|\theta).$$

Most conveniently, for an arbitrary grid with points $k_1, \ldots, k_n > 1$ along each dimension, these coefficients are given by:

$$\theta_{l_1 \ldots l_n} = \frac{2^n}{(k_1-1)\ldots(k_n-1)} \frac{1}{c_{l_1} \ldots c_{l_n}} \sum_{j_1=1}^{k_1} \cdots \sum_{j_n=1}^{k_n} \frac{1}{c_{j_1} \ldots c_{j_n}} \psi_{l_1}(\zeta_1) \ldots \psi_{l_d}(\zeta_n) d(\zeta_1, \ldots, \zeta_n) \quad (47)$$

where $c_j = 1$ for all j, except for the cases $c_1 = c_{k_d} = 2$, and $\zeta_k \in \mathcal{G}^i$ are the Gauss–Lobotto nodes. This approximation is exact in the Gauss–Lobotto nodes and interpolates among them.

There is nothing special about the use of Chebyshev polynomials as the basis functions $\psi_j(x)$ and we could rely, if required, on other basis functions. For instance, one can implement a finite element method with the Smolyak algorithm by partitioning Ω into elements and defining local basis functions as in Nobile et al. (2008). We use Chebyshev polynomials just because they have been popular in the applications of the Smolyak algorithm in macroeconomics.

5.7.1.6 Sixth Step: Building the Interpolating Function in n Dimensions
The Smolyak function that interpolates on $\mathbb{G}(q, n)$ is:

$$d(x|\theta, q, n) = \sum_{\max(n, q-n+1) \le |\mathbf{i}| \le q} (-1)^{q-|\mathbf{i}|} \binom{n-1}{q-|\mathbf{i}|} p^{|\mathbf{i}|}(x|\theta),$$

which is nothing more than the weighted sum of the tensors. In our previous example, a DSGE model with two continuous state variables and $q = 3$, we will have the sparse grid:

$$\mathbb{G}(3,2) = \{(-1,0),(0,1),(0,0),(0,-1),(1,0)\}$$

(this sparse grid was drawn in the top left panel of Fig. 11) and:

$$d(x|\theta,q,n) = \sum_{2 \leq |\mathbf{i}| \leq 3} (-1)^{3-|\mathbf{i}|} \binom{1}{3-|\mathbf{i}|} p^{|\mathbf{i}|}(x|\theta)$$

$$= (-1)\binom{1}{1} p^{|2|}(x|\theta) + (-1)^0 \binom{1}{0} p^{|3|}(x|\theta)$$

$$= p^{1,2}(x|\theta) + p^{2,1}(x|\theta) - p^{1,1}(x|\theta)$$

$$= \theta_{11} + \theta_{21} T_1(x_1) + \theta_{31} T_2(x_1) + \theta_{12} T_1(x_2) + \theta_{13} T_2(x_2).$$

Each of the coefficients in this approximation is given by the formula in Eq. (47):

$$\theta_{21} = \frac{1}{2}(d(1,0) - d(-1,0))$$

$$\theta_{12} = \frac{1}{2}(d(0,1) - d(0,-1))$$

$$\theta_{31} = \frac{1}{4}(d(1,0) + d(-1,0)) - \frac{1}{2}d(0,0)$$

$$\theta_{13} = \frac{1}{4}(d(0,1) + d(0,-1)) - \frac{1}{2}d(0,0)$$

except the constant term:

$$\theta_{11} = \frac{1}{4}(d(0,1) + d(0,-1) + d(1,0) + d(-1,0)),$$

which instead ensures that the interpolating function satisfies $d(0,0) = d(x|\theta,q,n)$. It is easy to check that we indeed satisfy the condition that the approximating function equates the unknown function at the points of the sparse grid. For example, at $(-1,0)$:

$$d((-1,0)|\theta,q,n) = \theta_{11} + \theta_{21} T_1(-1) + \theta_{31} T_2(-1) + \theta_{12} T_1(0) + \theta_{13} T_2(0)$$

$$= \theta_{11} - \theta_{21} + \theta_{31} - \theta_{13}$$

$$= \frac{1}{4}(d(0,1) + d(0,-1) + d(1,0) + d(-1,0))$$

$$- \frac{1}{2}(d(1,0) - d(-1,0))$$

$$+ \frac{1}{4}(d(1,0) + d(-1,0)) - \frac{1}{2}d(0,0)$$

$$- \frac{1}{4}(d(0,1) + d(0,-1)) + \frac{1}{2}d(0,0)$$

$$= d(-1,0).$$

An interesting property of this construction of $d(x|\theta,q,n)$ is that the cardinality of $\mathbb{G}(q,n)$ and the number of coefficients on θ coincide. In our previous example, $\mathbb{G}(3,2)=5$ and $\theta=\{\theta_{11},\theta_{21},\theta_{31},\theta_{12},\theta_{13}\}$. A second relevant property is that $d(x|\theta,q,n)$ exactly replicates any polynomial function built with monomials of degree less than or equal to $q-n$.

5.7.1.7 Seventh Step: Solving for the Polynomial Coefficients

We plug $d(x|\theta,q,n)$ into the operator $\mathcal{H}(\cdot)$ for all $x_i \in \mathbb{G}(q,n)$. At this point the operator needs to be exactly zero:

$$\mathcal{H}(d(x_i|\theta,q,n))=0$$

and we solve for the unknown coefficients on θ. In our previous example, we had $\mathbb{G}(3,2)=\{(-1,0),(0,1),(0,0),(0,-1),(1,0)\}$ and, therefore:

$$d((-1,0)|\theta,q,n)=\theta_{11}+\theta_{21}T_1(-1)+\theta_{31}T_2(-1)+\theta_{12}T_1(0)+\theta_{13}T_2(0)=\theta_{11}-\theta_{21}+\theta_{31}-\theta_{13}$$

$$d((0,1)|\theta,q,n)=\theta_{11}+\theta_{21}T_1(0)+\theta_{31}T_2(0)+\theta_{12}T_1(1)+\theta_{13}T_2(1)=\theta_{11}-\theta_{31}+\theta_{12}+\theta_{13}$$

$$d((0,0)|\theta,q,n)=\theta_{11}+\theta_{21}T_1(0)+\theta_{31}T_2(0)+\theta_{12}T_1(0)+\theta_{13}T_2(0)=\theta_{11}-\theta_{31}-\theta_{13}$$

$$d((0,-1)|\theta,q,n)=\theta_{11}+\theta_{21}T_1(0)+\theta_{31}T_2(0)+\theta_{12}T_1(-1)+\theta_{13}T_2(-1)=\theta_{11}-\theta_{31}-\theta_{12}+\theta_{13}$$

$$d((1,0)|\theta,q,n)=\theta_{11}+\theta_{21}T_1(1)+\theta_{31}T_2(1)+\theta_{12}T_1(0)+\theta_{13}T_2(0)=\theta_{11}+\theta_{21}+\theta_{31}-\theta_{13}$$

The system of equations:

$$\mathcal{H}(d(x_i|\theta,q,n))=0, \, x_i \in \mathbb{G}(q,n)$$

can be solved with a standard nonlinear solver. Krüger and Kubler (2004) and Malin et al. (2011) suggest a time–iteration method that starts, as an initial guess, from the first-order perturbation of the model. This choice is, nevertheless, not essential to the method.

5.7.2 Extensions

Recently, Judd et al. (2014b) have proposed an important improvement of Smolyak's algorithm. More concretely, the authors first present a more efficient implementation of Smolyak's algorithm that uses disjoint-set generators that are equivalent to the sets $\mathcal{G}^i$. Second, the authors use a Lagrange interpolation scheme. Third, the authors build an anisotropic grid, which allows having a different number of grid points and basis functions for different state variables. This may be important to capture the fact that, often, it is harder to approximate the decision rules of agents along some dimensions than along others. Finally, the authors argue that it is much more efficient to employ a derivative-free fixed-point iteration method instead of the time-iteration scheme proposed by Krüger and Kubler (2004) and Malin et al. (2011).

In comparison, Brumm and Scheidegger (2015) keep a time-iteration procedure, but they embed on it an adaptive sparse grid. This grid is refined locally in an automatic fashion, which allows the capture of steep gradients and some nondifferentiabilities. The authors provide a fully hybrid parallel implementation of the method, which takes advantage of the fast improvements in massively parallel processing.

6. COMPARISON OF PERTURBATION AND PROJECTION METHODS

After our description of perturbation and projection methods, we can offer some brief comments on their relative strengths and weaknesses.

Perturbation methods have one great advantage: their computational efficiency. We can compute, using a standard laptop computer, a third-order approximation to DSGE models with dozens of state variables in a few seconds. Perturbation methods have one great disadvantage: they only provide a local solution. The Taylor series expansion is accurate around the point at which we perform the perturbation and deteriorates as we move away from that point. Although perturbation methods often yield good global results (see Aruoba et al., 2006; Caldara et al., 2012; and Swanson et al., 2006), such performance needs to be assessed in each concrete application and even a wide range of accuracy may not be sufficient for some quantitative experiments. Furthermore, perturbation relies on differentiability conditions that are often violated by models of interest, such as those that present kinks or occasionally binding constraints.[aa]

Projection methods are nearly the mirror image of perturbation. Projection methods have one great advantage: Chebyshev and finite elements produce solutions that are of high accuracy over the whole range of state variable values. See, again, Aruoba et al. (2006) and Caldara et al. (2012). And projection methods can attack even the most complex problems with occasionally binding constraints, irregular shapes, and local behavior. But power and flexibility come at a cost: computational effort. Projection methods are harder to code, take longer to run, and suffer, as we have repeatedly pointed out, from an acute curse of dimensionality.[ab]

Thus, which method to use in real life? The answer, not unsurprisingly, is "it depends." Solution methods for DSGE models provide a menu of options. If we are dealing, for example, with a standard middle-sized New Keynesian model with

---

[aa] Researchers have proposed getting around these problems with different devices, such as the use of penalty functions. See, for example, Preston and Roca (2007). In fact, the recent experience of several central banks pushing their target interest rates below zero suggests that many constraints such as the ZLB may be closer to such a penalty function than to a traditional kink.

[ab] The real bottleneck for most research projects involving DSGE models is coding time, not running time. Moving from a few seconds of running time with perturbation to a few minutes of running time with projection is a minuscule fraction of the cost of coding a finite elements method in comparison with the cost of employing Dynare to find a perturbation.

25 state variables, perturbation methods are likely to be the best option. The New Keynesian model is sufficiently well behaved that a local approximation would be good enough for most purposes. A first-order approximation will deliver accurate estimates of the business cycle statistics such as variances and covariances, and a second- or third-order approximation is likely to generate good welfare estimates (although one should always be careful when performing welfare evaluations). If we are dealing, in contrast, with a DSGE model with financial constraints, large risk aversion, and only a few state variables, a projection method is likely to be a superior option. An experienced researcher may even want to have two different solutions to check one against the other, perhaps of a simplified version of the model, and decide which one provides her with a superior compromise between coding time, running time, and accuracy.

Remark 25 (Hybrid methods) The stark comparison between perturbation and projection methods hints at the possibility of developing hybrid methods that combine the best of both approaches. Judd (1998, section 15.6), proposes the following hybrid algorithm:

Algorithm 4 (Hybrid algorithm)
1. Use perturbation to build a basis tailored to the DSGE model we need to solve.
2. Apply a Gram–Schmidt process to build an orthogonal basis from the basis obtained in 1.
3. Employ a projection method with the basis from 2.

While this algorithm is promising (see the example provided by Judd, 1998), we are unaware of further explorations of this proposal.

More recently, Levintal (2015b) and Fernández-Villaverde and Levintal (2016) have proposed the use of Taylor-based approximations that also have the flavor of a hybrid method. The latter paper shows the high accuracy of this hybrid method in comparison with pure perturbation and projection methods when computing a DSGE model with disaster risk and a dozen state variables. Other hybrid proposals include Maliar et al. (2013).

7. ERROR ANALYSIS

A final step in every numerical solution of a DSGE model is to assess the error created by the approximation, that is, the difference between the exact and the approximated solution. This may seem challenging since the exact solution of the model is unknown. However, the literature has presented different methods to evaluate the errors.[ac] We will concentrate on the two most popular procedures to assess error: χ^2 −test proposed by

---

[ac] Here we follow much of the presentation of Aruoba et al. (2006), where the interested reader can find more details.

Den Haan and Marcet (1994) and the Euler equation error proposed by Judd (1992). Throughout this section, we will use the superscript j to index the perturbation order, the number of basis functions, or another characteristic of the solution method. For example, $c^j(k_t, z_t)$ will be the approximation to the decision rule for consumption $c(k_t, z_t)$ in a model with state variables k_t and z_t.

Remark 26 (Theoretical bounds) There are (limited) theoretical results bounding the approximation errors and their consequences. Santos and Vigo-Aguiar (1998) derive upper bounds for the error in models computed with value function iteration. Santos and Rust (2004) extend the exercise for policy function iteration. Santos and Peralta-Alva (2005) propose regularity conditions under which the error from the simulated moments of the model converge to zero as the approximated equilibrium function approaches the exact, but unknown, equilibrium function. Fernández-Villaverde et al. (2006) explore similar conditions for likelihood functions and Stachurski and Martin (2008) perform related work for the computation of densities of ergodic distributions of variables of interest. Judd et al. (2014a) have argued for the importance of constructing lower bounds on the size of approximation errors and propose a methodology to do so. Kogan and Mitra (2014) have studied the information relaxation method of Brown et al. (2010) to measure the welfare cost of using approximated decision rules. Santos and Peralta-Alva (2014) review the existing literature. But, despite all this notable work, this is an area in dire need of further investigation.

Remark 27 (Preliminary assessments) Before performing a formal error analysis, researchers should undertake several preliminary assessments. First, we need to check that the computed solution satisfies theoretical properties, such as concavity or monotonicity of the decision rules. Second, we need to check the shape and structure of decision rules, impulse response functions, and basic statistics of the model. Third, we need to check how the solution varies as we change the calibration of the model.

These steps often tell us more about the (lack of) accuracy of an approximated solution than any formal method. Obviously, the researcher should also take aggressive steps to verify that her code is correct and that she is, in fact, computing what she is supposed to compute. The use of modern, industry-tested software engineering techniques is crucial in ensuring code quality.

7.1 A χ^2 Accuracy Test

Den Haan and Marcet (1994) noted that, if some of the equilibrium conditions of the model are given by:

$$f(y_t) = \mathbb{E}_t(\phi(y_{t+1}, y_{t+2}, ..))$$

where the vector y_t contains n variables of interest at time t, $f : \mathbb{R}^n \to \mathbb{R}^m$ and $\phi : \mathbb{R}^n \times \mathbb{R}^\infty \to \mathbb{R}^m$ are known functions, then:

$$\mathbb{E}_t(u_{t+1} \otimes h(x_t)) = 0 \tag{48}$$

for any vector x_t measurable with respect to t with $u_{t+1} = \phi(y_{t+1}, y_{t+2}, ..) - f(y_t)$ and $h : \mathbb{R}^k \to \mathbb{R}^q$ being an arbitrary function.

If we simulate a series of length T from the DSGE model using a given solution method, $\{y_t^j\}_{t=1:T}$, we can find $\{u_{t+1}^j, x_t^j\}_{t=1:T}$ and compute the sample analog of (48):

$$B_T^j = \frac{1}{T} \sum_{t=1}^{T} u_{t+1}^j \otimes h(x_t^j). \tag{49}$$

The moment (49) would converge to zero as N increases almost surely if we were using the exact solution to the model. When, instead, we are using an approximation, the statistic $B(B_T^j)'(A_T^j)^{-1} B_T^j$ where A_T^j is a consistent estimate of the matrix:

$$\sum_{t=-\infty}^{\infty} \mathbb{E}_t \left[(u_{t+1} \otimes h(x_t))(u_{t+1} \otimes h(x_t))' \right]$$

converges to a χ^2 distribution with qm degrees of freedom under the null that the population moment (48) holds. Values of the test above the critical value can be interpreted as evidence against the accuracy of the solution. Since any solution method is an approximation, as T grows we will eventually reject the null. To control for this problem, Den Haan and Marcet (1990) suggest repeating the test for many simulations and report the percentage of statistics in the upper and lower critical 5% of the distribution. If the solution provides a good approximation, both percentages should be close to 5%.

This χ^2-test helps the researcher to assess how the errors of the approximated solution accumulate over time. Its main disadvantage is that rejections of accuracy may be difficult to interpret.

7.2 Euler Equation Errors

Judd (1992) proposed determining the quality of the solution method by defining normalized Euler equation errors. The idea is to measure how close the Euler equation at the core of nearly DSGE models is to be satisfied when we use the approximated solution.

The best way to understand how to implement this idea is with an example. We can go back to the stochastic neoclassical growth model that we solved in Section 5.6. This model generates an Euler equation:

$$u_c\prime(c_t, l_t) = \beta \mathbb{E}_t \{ u_c\prime(c_{t+1}, l_{t+1}) R_{t+1} \} \tag{50}$$

where

$$u_c\prime(c_t, l_t) = \frac{\left(c_t^\tau(1 - l_t)^{1-\tau}\right)^{1-\eta}}{c_t}$$

is the marginal utility of consumption and $R_{t+1} = \left(1 + \alpha e^{z_{t+1}} k_t^{\alpha-1} l_{t+1}^{1-\alpha} - \delta\right)$ is the gross return rate of capital. If we take the inverse of the marginal utility of consumption and do some algebra manipulations, we get:

$$1 - \frac{u_c'\left(\beta \mathbb{E}_t\left\{u_c'(c_{t+1}, l_{t+1}) R_{t+1}\right\}, l_t\right)^{-1}}{c_t} = 0 \tag{51}$$

If we plug into Eq. (51) the exact decision rules for consumption:

$$c_t = c(k_t, z_t),$$

labor

$$l_t = l(k_t, z_t)$$

and capital:

$$k_{t+1} = k(k_t, z_t)$$

we get:

$$1 - \frac{u_c'\left(\beta \mathbb{E}_t\left\{u_c'(c(k(k_t, z_t), z_{t+1}), l(k(k_t, z_t), z_{t+1})) R_{t+1}(k_t, z_t, z_{t+1})\right\}, l(k_t, z_t)\right)^{-1}}{c(k_t, z_t)} = 0 \tag{52}$$

where $R(k_t, z_t, z_{t+1}) = \left(1 + \alpha e^{z_{t+1}} k(k_t, z_t)^{\alpha-1} l(k(k_t, z_t), z_{t+1})^{1-\alpha} - \delta\right)$. Eq. (52) will hold exactly for any k_t and z_t.

If, instead, we plug into Eq. (52) the approximated decision rules $c^j(k_t, z_t)$, $l^j(k_t, z_t)$, and $k^j(k_t, z_t)$, we will have:

$$EEE(k_t, z_t)$$
$$= \left\{ \frac{1 -}{} \frac{u_c'\left(\beta \mathbb{E}_t\left\{u_c'\left(c^j\left(k^j\left(k_t^j, z_t\right), z_{t+1}\right), l^j(k^j(k_t, z_t), z_{t+1})\right) R_{t+1}^j(k_t, z_t, z_{t+1})\right\}, l^j(k_t, z_t)\right)^{-1}}{c^j(k_t, z_t)} \right\} \tag{53}$$

where $R^j(k_t, z_t, z_{t+1}) = \left(1 + \alpha e^{z_{t+1}} k^j(k_t, z_t)^{\alpha-1} l^j\left(k^j(k_t, z_t), z_{t+1}\right)^{1-\alpha} - \delta\right)$. Eq. (53) defines a function, $EEE(k_t, z_t)$, that we call the Euler equation error.

We highlight three points about Eq. (53). First, the error in the Euler equation depends on the value of the state variables k_t and z_t. Perturbation methods will tend to have a small Euler equation error close to the point where the perturbation is

undertaken and a larger Euler equation error farther from it. In contrast, projection methods will deliver a more uniform Euler equation error across Ω. Consequently, researchers have found it useful to summarize the Euler equation error. Proposals include the mean of the Euler equation error (either a simple average or using some estimate of the ergodic distribution of state variables[ad]) or the maximum of the Euler equation error in some region of Ω. Second, due to the algebraic transformation that we took on the Euler equation, $EEE(k_t, z_t)$ is expressed in consumption units, which have a meaningful economic interpretation as the relative optimization error incurred by the use of the approximated policy rule (Judd and Guu, 1997). For instance, if $EEE(k_t, z_t) = 0.01$, then the agent is making a \$1 mistake for each \$100 spent. In comparison, $EEE(k_t, z_t) = 1e^{-6}$ implies that the agent is making a 1 cent mistake for each 1 million spent. Third, the Euler equation error is also important because we know that, under certain conditions, the approximation error of the decision rule is of the same order of magnitude as the size of the Euler equation error. Correspondingly, the change in welfare is of the square order of the Euler equation error. Furthermore, the constants involved in these error bounds can be related to model primitives (Santos, 2000). Unfortunately, in some DSGE models it can be difficult to use algebraic transformations to achieve an expression for the Euler equation error that is interpretable as consumption units (or other natural economic unit).

Following the convention in the literature, we plot in Fig. 13, the $\log_{10}|EEE(k_t, z_t)|$ of the stochastic neoclassical growth model from Section 5.6. Taking the $\log_{10}$ eases reading: a value of –3 means \$1 mistake for each \$1000, a value of –4 a \$1 mistake for each \$10,000, and so on. Fig. 13 shows five lines, one for each value of productivity. As we hinted when we described the Chebyshev-collocation projection method, this accuracy is outstanding.

To compare this performance of Chebyshev-collocation with other solution methods, we reproduce, in Figs. 14 and 15, results from Aruoba et al. (2006). That paper uses the same stochastic neoclassical growth model with only a slightly different calibration (plus a few smaller details about how to handle z_t). Both figures display a transversal cut of the Euler equation errors when $z_t = 0$ and for values of capital between 70% and 130% of its steady-state value (23.14).

In Fig. 14, we plot the results for a first-order perturbation (in levels and in logs), a second-order perturbation, and a fifth-order perturbation. First, perturbations have smaller errors around the steady-state value of capital and deteriorate away from it. Second, there is a considerable improvement when we go from a first- to a second-order approximation. Third, a fifth-order approximation displays a great performance even 30% away from the steady state.

[ad] Using the ergodic distribution has the complication that we may not have access to it, since it is derived from the solution of the model, the object we are searching for. See Aruoba et al. (2006) for suggestions on how to handle this issue.
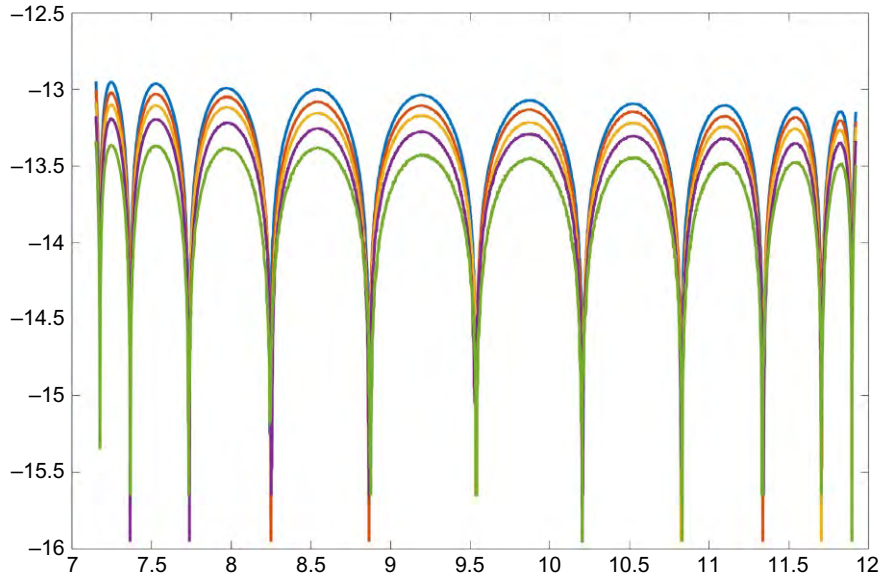
Fig. 13 Log_{10} of absolute value of Euler equation error.
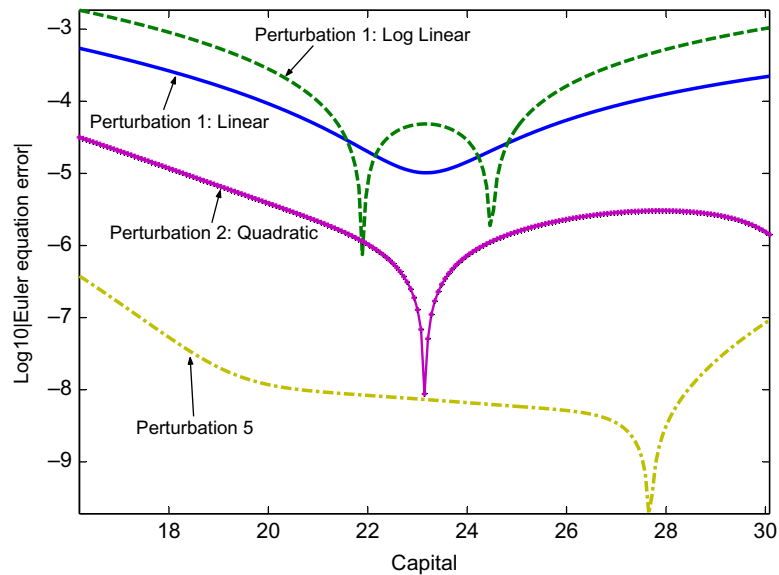
Fig. 14 Log_{10} of absolute value of Euler equation error.

In Fig. 15, we plot the results from the first-order perturbation (as a comparison with the previous graph), value function iteration (with a grid of one million points: 25,000 points for capital and 40 for the productivity level), finite elements (with 71 elements), and Chebyshev polynomials (as in Section 5.6, still with 11 polynomials). The main
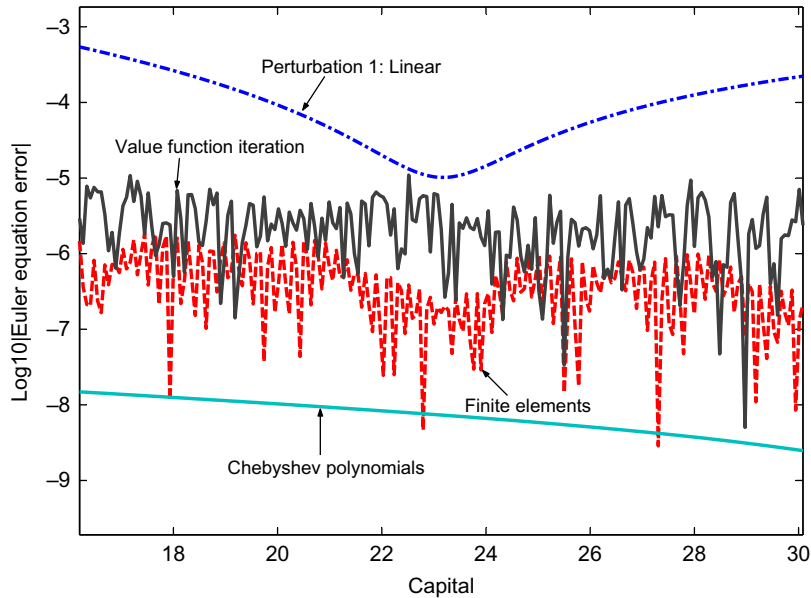
Fig. 15 Log_{10} of absolute value of Euler equation error.

lesson from this graph is that the Euler equation errors are much flatter for projection methods and value function iteration (another algorithm that delivers a global solution). The level of each of the three functions is harder to interpret, since it depends on the number of grid points (value function iteration), elements (finite elements), and Chebyshev polynomials. Nevertheless, the performance of Chebyshev is again excellent and its run time much lower than value function iteration and finite elements. This is not a surprise, since the decision rules for the stochastic neoclassical growth model are sufficiently well behaved for a spectral basis to do an extraordinary job.

Computing the Euler equation error has become standard in the literature because it often offers sharp assessments. However, Euler equation errors fail at giving a clear evaluation of how the errors of the approximated solution accumulate over time (see Santos and Peralta-Alva, 2005, for how to think about the impact of Euler equation errors on computed moments from the model). Thus, Euler equation errors should be understood as a complement to, not a substitute for, Den Haan and Marcet (1994)'s χ^2 −test.

7.3 Improving the Error

Once we have gauged the error in the solution to the DSGE, we can decide whether to improve the accuracy of the solution. Everything else equal, more accuracy is better than less accuracy. But, in real-life applications, everything else is rarely equal. More accuracy can come at the cost of more coding time and, in particular, longer running time. For

example, in the exercise with the stochastic neoclassical growth model reported in Fig. 15, we could subdivide the finite elements as much as we want and use modern scientific libraries such as the *GNU multiple precision arithmetic library* to achieve any arbitrary level of accuracy, but at the cost of longer running times and more memory requirements. The researcher must look at her needs and resources and, once inferior solution methods are rejected, select those that best fit her goals.

But if the goal is indeed dependent on achieving additional accuracy, there are different possibilities available. If a perturbation is being used, we can increase the order of the perturbation. If a projection method is being used, we can increase the number of elements in the basis. The researcher can also explore changes of variables to make the problem more linear or switch the solution method.

Once the error of the model has been assessed, we are finally ready to move to Part II and see how the DSGE model can account for the observed data.

PART II. ESTIMATING DSGE MODELS

8. CONFRONTING DSGE MODELS WITH DATA

The preceding sections discussed how to compute an approximate solution for a DSGE model conditional on its parameterization. Part II focuses on determining the DSGE model parameters based on the empirical evidence and assessing the model's fit. More specifically, we ask four fundamental questions: (i) How can one estimate the DSGE model parameters from the observed macroeconomic time series? (ii) How well does the estimated DSGE model capture salient features of the data? (iii) What are the quantitative implications of the estimated DSGE models with respect to, for instance, sources of business cycle fluctuations, propagation of exogenous shocks, the effect of changes in macroeconomic policies, and the future path of macroeconomic time series? (iv) How should one construct measures of uncertainty for the parameters and the quantitative implications of the DSGE model? To answer these questions, we begin by analytically solving a stylized New Keynesian DSGE model in Section 8.1 and studying its properties in Section 8.2. DSGE model-implied population moments, autocovariances, spectra, and impulse response functions have sample analogs in the data, which are examined in Section 8.3. Macroeconomic time series exhibit trends that may or may not be captured by the DSGE model, which is discussed in Section 8.4.

Part II of this chapter assumes that the reader has some basic familiarity with econometrics, at the level of a first-year PhD sequence in a US graduate program. With the exception of Canova (2007) and DeJong and Dave (2007) there are no textbooks that focus on the estimation of DSGE models. The literature has progressed quickly since these two books were first written. The subsequent sections contain, in addition to a critical introduction to "standard methods," an overview of the most recent developments in the literature, which include identification conditions for DSGE models,

identification-robust frequentist inference, and sequential Monte Carlo techniques for Bayesian analysis. Unlike the recent monograph by Herbst and Schorfheide (2015) which focuses on Bayesian computations, Part II of this chapter also contains extensive discussions of the consequences of misspecification for econometric inference and covers frequentist methods.

8.1 A Stylized DSGE Model

Throughout Part II we consider a stylized New Keynesian DSGE model in its loglinearized form.[ae] This model shares many of the features of its more realistic siblings that have been estimated in the literature. It is a stripped-down version of the model developed in the work by Christiano et al. (2005) and Smets and Wouters (2003). The specific version presented below is taken from Del Negro and Schorfheide (2008) and obtained by imposing several parameter restrictions. It is not suitable to be confronted with actual data, but it can be solved analytically, which is useful for the subsequent exposition. For brevity, we refer to this model as the stylized DSGE model in the remainder of this chapter.

   The model economy consists of households, intermediate goods producers, final goods producers, a monetary policy authority, and a fiscal authority. Macroeconomic fluctuations are generated by four exogenous processes: a technology growth shock, z_t, a shock that generates shifts in the preference for leisure, ϕ_t, a price markup shock, λ_t, and a monetary policy shock $\epsilon_{R,t}$. We assume that the level of productivity Z_t in the economy is evolving exogenously according to a random walk with drift:

$$\log Z_t = \log \gamma + \log Z_{t-1} + z_t, \quad z_t = \rho_z z_{t-1} + \sigma_z \epsilon_{z,t}. \tag{54}$$

The productivity process Z_t induces a stochastic trend in output X_t and real wages W_t. To facilitate the model solution, it is useful to detrend output and real wages by the level of technology, defining $x_t = X_t/Z_t$ and $w_t = W_t/Z_t$, respectively. In terms of the detrended variables, the model has the following steady state:

$$\bar{x} = x^*, \quad \bar{w} = \overline{lsh} = \frac{1}{1+\lambda}, \quad \bar{\pi} = \pi^*, \quad \bar{R} = \pi^* \frac{\gamma}{\beta}. \tag{55}$$

Here x^* and π^* are free parameters. The latter can be interpreted as the central bank's target inflation rate, whereas the former can in principle be derived from the weight on leisure in the households' utility function. The steady-state real wage $\bar{w}$ is equal to the steady-state labor share $\overline{lsh}$. The parameter λ can be interpreted as the steady-state markup charged by the monopolistically competitive intermediate goods producers, β is the discount factor of the households, and γ is the growth rate of technology. Under the assumption that the production technology is linear in labor and labor is the only

---

[ae] See Sections 4.1 and 4.5 for how to think about loglinearizations as a first-order perturbations.

factor of production, the steady state labor share equals the steady state of detrended wages. We also assume that all output is consumed, which means that x can be interpreted as aggregate consumption.

8.1.1 Loglinearized Equilibrium Conditions

In terms of log-deviations from the steady state (denoted by ˆ), ie, $\hat{x} = \log(x_t/\bar{x})$, $\hat{w}_t = \log(w_t/\bar{w})$, $\hat{\pi}_t = \log(\pi_t/\bar{\pi})$, and $\hat{R}_t = \log(R_t/\bar{R})$, the equilibrium conditions of the model can be stated as follows. The consumption Euler equation of the households takes the form

$$\hat{x}_t = \mathbb{E}_{t+1}[\hat{x}_{t+1}] - \left(\hat{R}_t - \mathbb{E}[\hat{\pi}_{t+1}]\right) + \mathbb{E}_t[z_{t+1}]. \tag{56}$$

The expected technology growth rate arises because the Euler equation is written in terms of output in deviations from the stochastic trend induced by Z_t. Assuming the absence of nominal wage rigidities, the intratemporal Euler equation for the households leads to the following labor supply equation:

$$\hat{w}_t = (1 + \nu)\hat{x}_t + \phi_t, \tag{57}$$

where $\hat{w}_t$ is the real wage, $1/(1 + \nu)$ is the Frisch labor supply elasticity, $\hat{x}_t$ is proportional to hours worked, and ϕ_t is an exogenous labor supply shifter

$$\phi_t = \rho_\phi \phi_{t-1} + \sigma_\phi \epsilon_{\phi, t}. \tag{58}$$

We refer to ϕ_t as preference shock.

The intermediate goods producers hire labor from the households and produce differentiated products, indexed by j, using a linear technology of the form $X_t(j) = Z_t L_t(j)$. After detrending and loglinearization around steady-state aggregate output, the production function becomes

$$\hat{x}_t(j) = \hat{L}_t(j). \tag{59}$$

Nominal price rigidity is introduced via the Calvo mechanism. In each period, firm j is unable to reoptimize its nominal price with probability ζ_p. In this case, the firm simply adjusts its price from the previous period by the steady-state inflation rate. With probability $1 - \zeta_p$, the firm can choose its price to maximize the expected sum of future profits. The intermediate goods are purchased and converted into an aggregate good X_t by a collection of perfectly competitive final goods producers using a constant-elasticity-of-substitution aggregator.

The optimality conditions for the two types of firms can be combined into the so-called New Keynesian Phillips curve, which can be expressed as

$$\hat{\pi}_t = \beta \mathbb{E}_t[\hat{\pi}_{t+1}] + \kappa_p(\hat{w}_t + \lambda_t), \quad \kappa_p = \frac{(1 - \zeta_p \beta)(1 - \zeta_p)}{\zeta_p}, \tag{60}$$

where β is the households' discount factor and λ_t can be interpreted as a price mark-up shock, which exogenously evolves according to

$$\lambda_t = \rho_\lambda \lambda_{t-1} + \sigma_\lambda \epsilon_{\lambda,t}. \tag{61}$$

It is possible to derive an aggregate resource constraint that relates the total amount of labor L_t hired by the intermediate goods producers to the total aggregate output X_t produced in the economy. Based on this aggregate resource constraint, it is possible to compute the labor share of income, which, in terms of deviations from steady state is given by

$$\widehat{lsh}_t = \hat{w}_t. \tag{62}$$

Finally, the central bank sets the nominal interest rate according to the feedback rule

$$\hat{R}_t = \psi \hat{\pi}_t + \sigma_R \epsilon_{R,t} \quad \psi = 1/\beta. \tag{63}$$

We abstract from interest rate smoothing and the fact that central banks typically also react to some measure of real activity, eg, the gap between actual output and potential output. The shock $\epsilon_{R,t}$ is an unanticipated deviation from the systematic part of the interest rate feedback rule and is called a monetary policy shock. We assume that $\psi = 1/\beta$, which ensures the existence of a unique stable solution to the system of linear rational expectations difference equations and, as will become apparent below, simplifies the solution of the model considerably. The fiscal authority determines the level of debt and lump-sum taxes such that the government budget constraint is satisfied.

8.1.2 Model Solution

To solve the model, note that the economic state variables are ϕ_t, λ_t, z_t, and $\epsilon_{R,t}$. Due to the fairly simple loglinear structure of the model, the aggregate laws of motion $\hat{x}(\cdot)$, $\hat{lsh}(\cdot)$, $\hat{\pi}(\cdot)$, and $\hat{R}(\cdot)$ are linear in the states and can be determined sequentially. We first eliminate the nominal interest rate from the consumption Euler equation using (63):

$$\hat{x}_t = \mathbb{E}_{t+1}[\hat{x}_{t+1}] - \left(\frac{1}{\beta}\hat{\pi}_t + \sigma_R \epsilon_{R,t} - \mathbb{E}[\hat{\pi}_{t+1}]\right) + \mathbb{E}_t[z_{t+1}]. \tag{64}$$

Now notice that the New Keynesian Phillips curve can be rewritten as

$$\frac{1}{\beta}\hat{\pi}_t - \mathbb{E}_t[\hat{\pi}_{t+1}] = \frac{\kappa_p}{\beta}((1+\nu)\hat{x}_t + \phi_t + \lambda_t). \tag{65}$$

Here we replaced wages $\hat{w}_t$ with the right-hand side of (57). Substituting (65) into (64) and rearranging terms leads to the following expectational difference equation for output $\hat{x}_t$

$$\hat{x}_t = \psi_p \mathbb{E}_t[\hat{x}_{t+1}] - \frac{\kappa_p \psi_p}{\beta}(\phi_t + \lambda_t) + \psi_p \mathbb{E}_t[z_{t+1}] - \psi_p \sigma_R \epsilon_{R,t}, \tag{66}$$

where $0 \leq \psi_p \leq 1$ is given by

$$\psi_p = \left(1 + \frac{\kappa_p}{\beta}(1+\nu)\right)^{-1}.$$

We now need to find a law of motion for output (and, equivalently, consumption) of the form

$$\hat{x}_t = \hat{x}(\phi_t, \lambda_t, z_t, \epsilon_{R,t}) = x_\phi \phi_t + x_\lambda \lambda_t + x_z z_t + x_{\epsilon_R} \epsilon_{R,t} \tag{67}$$

that solves the functional equation

$$
\begin{aligned}
\mathbb{E}_t \mathcal{H}(\hat{x}(\cdot)) \\
= \mathbb{E}_t &\left[\hat{x}(\phi_t, \lambda_t, z_t, \epsilon_{R,t}) - \psi_p \hat{x}\left(\rho_\phi \phi_t + \sigma_\phi \epsilon_{\phi,t+1}, \rho_\lambda \lambda_t + \sigma_\lambda \epsilon_{\lambda,t+1}, \rho_z z_t + \sigma_z \epsilon_{z,t+1}, \epsilon_{R,t+1}\right) \right. \\
&\left. + \frac{\kappa_p \psi_p}{\beta}(\phi_t + \lambda_t) - \psi_p z_{t+1} + \psi_p \sigma_R \epsilon_{R,t} \right] = 0.
\end{aligned}
\tag{68}
$$

Here, we used the laws of motion of the exogenous shock processes in (54), (58), and (61). Assuming that the innovations ϵ_t are Martingale difference sequences, it can be verified that the coefficients of the linear decision rule are given by

$$x_\phi = -\frac{\kappa_p \psi_p / \beta}{1 - \psi_p \rho_\phi}, \quad x_\lambda = -\frac{\kappa_p \psi_p / \beta}{1 - \psi_p \rho_\lambda}, \quad x_z = \frac{\rho_z \psi_p}{1 - \psi_p \rho_z} z_t, \quad x_{\epsilon_R} = -\psi_p \sigma_R. \tag{69}$$

After having determined the law of motion for output, we now solve for the labor share, inflation, and nominal interest rates. Using (57) and (62) we immediately deduce that the labor share evolves according to

$$\widehat{lsh}_t = \left[1 + (1+\nu)x_\phi\right]\phi_t + (1+\nu)x_\lambda \lambda_t + (1+\nu)x_z z_t + (1+\nu)x_{\epsilon_R} \epsilon_{R,t}. \tag{70}$$

To obtain the law of motion of inflation, we have to solve the following functional equation derived from the New Keynesian Phillips curve (60):

$$
\begin{aligned}
\mathbb{E}_t \mathcal{H}(\hat{\pi}(\cdot)) \\
= \mathbb{E}_t &\left[\hat{\pi}(\phi_t, \lambda_t, z_t, \epsilon_{R,t}) - \beta \hat{\pi}\left(\rho_\phi \phi_t + \sigma_\phi \epsilon_{\phi,t+1}, \rho_\lambda \lambda_t + \sigma_\lambda \epsilon_{\lambda,t+1}, \rho_z z_t + \sigma_z \epsilon_{z,t+1}, \epsilon_{R,t+1}\right) \right. \\
&\left. - \kappa_p \widehat{lsh}(\phi_t, \lambda_t, z_t, \epsilon_{R,t}) - \kappa_p \lambda_t \right] = 0,
\end{aligned}
\tag{71}
$$

where $\widehat{lsh}(\cdot)$ is given by (70). The solution takes the form

$$
\begin{aligned}
\hat{\pi}_t = {} & \frac{\kappa_p}{1 - \beta \rho_\phi}\left[1 + (1+\nu)x_\phi\right]\phi_t + \frac{\kappa_p}{1 - \beta \rho_\lambda}\left[1 + (1+\nu)x_\lambda\right]\lambda_t \\
& + \frac{\kappa_p}{(1 - \beta \rho_z)}(1+\nu)x_z z_t + \kappa_p(1+\nu)x_{\epsilon_R} \epsilon_{R,t}.
\end{aligned}
\tag{72}
$$

Finally, combining (72) with the monetary policy rule (63) yields the solution for the nominal interest rate

$$\hat{R}_t = \frac{\kappa_p/\beta}{1 - \beta\rho_\phi} \left[1 + (1 + \nu)x_\phi\right]\phi_t + \frac{\kappa_p/\beta}{1 - \beta\rho_\lambda}\left[1 + (1 + \nu)x_\lambda\right]\lambda_t$$
$$+ \frac{\kappa_p/\beta}{1 - \beta\rho_z}(1 + \nu)x_z z_t + \left[\kappa_p(1 + \nu)x_{\epsilon_R}/\beta + \sigma_R\right]\epsilon_{R,t}. \tag{73}$$

8.1.3 State-Space Representation

To confront the model with data, one has to account for the presence of the model-implied stochastic trend in aggregate output and to add the steady states to all model variables. Measurement equations for output growth, the labor share, net inflation rates and net interest rates take the form

$$\log(X_t/X_{t-1}) = \hat{x}_t - \hat{x}_{t-1} + z_t + \log\gamma$$
$$\log(lsh_t) = \widehat{lsh}_t + \log(lsh)$$
$$\log\pi_t = \hat{\pi}_t + \log\pi^* \tag{74}$$
$$\log R_t = \hat{R}_t + \log(\pi^*\gamma/\beta).$$

The DSGE model solution has the form of a generic state-space model. Define the $n_s \times 1$ vector of econometric state variables s_t as

$$s_t = [\phi_t, \lambda_t, z_t, \epsilon_{R,t}, \hat{x}_{t-1}]'$$

and the vector of DSGE model parameters[af]

$$\theta = [\beta, \gamma, \lambda, \pi^*, \zeta_p, \nu, \rho_\phi, \rho_\lambda, \rho_z, \sigma_\phi, \sigma_\lambda, \sigma_z, \sigma_R]'. \tag{75}$$

We omitted the steady-state output x^* from the list of parameters because it does not affect the law of motion of output growth. Using this notation, we can express the state transition equation as

$$s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t, \tag{76}$$

where the $n_\epsilon \times 1$ vector ϵ_t is defined as $\epsilon_t = [\epsilon_{\phi,t}, \epsilon_{\lambda,t}, \epsilon_{z,t}, \epsilon_{R,t}]'$. The coefficient matrices $\Phi_1(\theta)$ and $\Phi_\epsilon(\theta)$ are determined by (54), (58), (61), the identity $\epsilon_{R,t} = \epsilon_{R,t}$, and a lagged version of (69) to determine $\hat{x}_{t-1}$. If we define the $n_y \times 1$ vector of observables as

$$y_t = M_y'[\log(X_t/X_{t-1}), \log lsh_t, \log\pi_t, \log R_t]', \tag{77}$$

where M_y' is a matrix that selects rows of the vector $[\log(X_t/X_{t-1}), \log lsh_t, \log\pi_t, \log R_t]'$ then the measurement equation can be written as

$$y_t = \Psi_0(\theta) + \Psi_1(\theta)s_t. \tag{78}$$

[af] From now on, we will use θ to denote the parameters of the DSGE model as opposed to the coefficients of a decision rule conditional on a particular set of DSGE model parameters. Also, to reduce clutter, we no longer distinguish vectors and matrices from scalars by using boldfaced symbols.

The coefficient matrices $\Psi_0(\theta)$ and $\Psi_1(\theta)$ can be obtained from (74), the equilibrium law of motion for the detrended model variables given by (69), (70), (72), and (73). They are summarized in Table 4.

The state-space representation of the DSGE model given by (76) and (78) provides the basis for the subsequent econometric analysis. It characterizes the joint distribution of the observables y_t and the state variables s_t conditional on the DSGE model parameters θ

$$p(Y_{1:T}, S_{1:T}|\theta) = \int \left(\prod_{t=1}^{T} p(y_t|s_t, \theta) p(s_t|s_{t-1}, \theta) \right) p(s_0|\theta) ds_0, \tag{79}$$

where $Y_{1:t} = \{y_1, \ldots, y_t\}$ and $S_{1:t} = \{s_1, \ldots, s_t\}$. Because the states are (at least partially) unobserved, we will often work with the marginal distribution of the observables defined as

$$p(Y_{1:T}|\theta) = \int p(Y_{1:T}, S_{1:T}|\theta) dS_{1:T}. \tag{80}$$

As a function of θ the density $p(Y_{1:T}|\theta)$ is called the likelihood function. It plays a central role in econometric inference and its evaluation will be discussed in detail in Section 10.

Remark 28 First, it is important to distinguish economic state variables, namely, ϕ_t, λ_t, z_t, and $\epsilon_{R,t}$, that are relevant for the agents' intertemporal optimization problems, from the econometric state variables s_t, which are used to cast the DSGE model solution into the state-space form given by (76) and (78). The economic state variables of our simple model are all exogenous. As we have seen in Section 4.3, the vector of state variables of a richer DSGE model also may include one or more endogenous variables, eg, the capital stock. Second, output growth in the measurement equation could be replaced by the level of output. This would require adding x^* to the parameter vector θ, eliminating $\hat{x}_{t-1}$ from s_t, adding $\log Z_t/\gamma^t$ to s_t, and accounting for the deterministic trend component $(\log\gamma)t$ in log output in the measurement equation. Third, the measurement Eq. (78) could be augmented by measurement errors. Fourth, if a DSGE model is solved with a higher-order perturbation or projection method, then, depending on how exactly the state vector s_t is defined, the state-transition Eq. (76), the measurement Eq. (78), or both are nonlinear.

8.2 Model Implications

Once we specify a distribution for the innovation vector ϵ_t the probability distribution of the DSGE model variables is fully determined. Recall that the innovation standard deviations were absorbed into the definition of the matrix $\Phi_\epsilon(\theta)$ in (76). For the sake of concreteness, we assume that

$$\epsilon_t \sim iidN(0, I), \tag{81}$$

where I denotes the identity matrix. Based on the probabilistic structure of the DSGE model, we can derive a number of implications from the DSGE model that will later

Table 4 System matrices for DSGE model

State-space representation:
$$y_t = \Psi_0(\theta) + \Psi_1(\theta)s_t$$
$$s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t$$

System matrices:

$$\Psi_0(\theta) = M_\gamma' \begin{bmatrix} \log\gamma \\ \log(lsh) \\ \log\pi^* \\ \log(\pi^*\gamma/\beta) \end{bmatrix}, \quad x_\phi = -\frac{\kappa_p\psi_p/\beta}{1-\psi_p\rho_\phi}, \quad x_\lambda = -\frac{\kappa_p\psi_p/\beta}{1-\psi_p\rho_\lambda}, \quad x_z = \frac{\rho_z\psi_p}{1-\psi_p\rho_z}, \quad x_{\epsilon_R} = -\psi_p\sigma_R$$

$$\Psi_1(\theta) = M_\gamma' \begin{bmatrix} x_\phi & x_\lambda & x_z+1 & x_{\epsilon_R} & -1 \\[2mm] 1+(1+\nu)x_\phi & (1+\nu)x_\lambda & (1+\nu)x_z & (1+\nu)x_{\epsilon_R} & 0 \\[2mm] \dfrac{\kappa_p}{1-\beta\rho_\phi}(1+(1+\nu)x_\phi) & \dfrac{\kappa_p}{1-\beta\rho_\lambda}(1+(1+\nu)x_\lambda) & \dfrac{\kappa_p}{1-\beta\rho_z}(1+\nu)x_z & +\kappa_p(1+\nu)x_{\epsilon_R} & 0 \\[2mm] \dfrac{\kappa_p/\beta}{1-\beta\rho_\phi}(1+(1+\nu)x_\phi) & \dfrac{\kappa_p/\beta}{1-\beta\rho_\lambda}(1+(1+\nu)x_\lambda) & \dfrac{\kappa_p/\beta}{1-\beta\rho_z}(1+\nu)x_z & (\kappa_p(1+\nu)x_{\epsilon_R}/\beta+\sigma_R) & 0 \end{bmatrix}$$

$$\Phi_1(\theta) = \begin{bmatrix} \rho_\phi & 0 & 0 & 0 & 0 \\ 0 & \rho_\lambda & 0 & 0 & 0 \\ 0 & 0 & \rho_z & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ x_\phi & x_\lambda & x_z & x_{\epsilon_R} & 0 \end{bmatrix}, \quad \Phi_\epsilon(\theta) = \begin{bmatrix} \sigma_\phi & 0 & 0 & 0 \\ 0 & \sigma_\lambda & 0 & 0 \\ 0 & 0 & \sigma_z & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 \end{bmatrix}$$

M_γ' is an $n_\gamma \times 4$ selection matrix that selects rows of Ψ_0 and Ψ_1.

Table 5 Parameter values for stylized DSGE model

Parameter	Value	Parameter	Value
β	1/1.01	γ	$\exp(0.005)$
λ	0.15	π^*	$\exp(0.005)$
ζ_p	0.65	ν	0
ρ_ϕ	0.94	ρ_λ	0.88
ρ_z	0.13		
σ_ϕ	0.01	σ_λ	0.01
σ_z	0.01	σ_R	0.01

be used to construct estimators of the parameter vector θ and evaluate the fit of the model. For now, we fix θ to the values listed in Table 5.

8.2.1 Autocovariances and Forecast Error Variances

DSGE models are widely used for business cycle analysis. In this regard, the model-implied variances, autocorrelations, and cross-correlations are important objects. For linear DSGE models it is straightforward to compute the autocovariance function from the state-space representation given by (76) and (78).[ag] Using the notation

$$\Gamma_{yy}(h) = \mathbb{E}[y_t y_{t-h}], \quad \Gamma_{ss}(h) = \mathbb{E}[s_t s_{t-h}], \quad \text{and} \quad \Gamma_{ys}(h) = \mathbb{E}[y_t s'_{t-h}]$$

and the assumption that $\mathbb{E}[\epsilon_t \epsilon_t'] = I$, we can express the autocovariance matrix of s_t as the solution to the following Lyapunov equation:[ah]

$$\Gamma_{ss}(0) = \Phi_1 \Gamma_{ss}(0) \Phi_1' + \Phi_\epsilon \Phi_\epsilon'. \tag{82}$$

Once the covariance matrix of s_t has been determined, it is straightforward to compute the autocovariance matrices for $h \neq 0$ according to

$$\Gamma_{ss}(h) = \Phi_1^h \Gamma_{ss}(0). \tag{83}$$

Finally, using the measurement Eq. (78), we deduce that

$$\Gamma_{yy}(h) = \Psi_1 \Gamma_{ss}(h) \Psi_1', \quad \Gamma_{ys}(h) = \Psi_1 \Gamma_{ss}(h). \tag{84}$$

[ag] For the parameters in Table 5, the largest (in absolute value) eigenvalue of the matrix $\Phi_1(\theta)$ in (76) is less than one, which implies that the VAR(1) law of motion for s_t is covariance stationary.

[ah] Efficient numerical routines to solve Lyapunov equations are readily available in many software packages, eg, the function *dylap* in MATLAB.
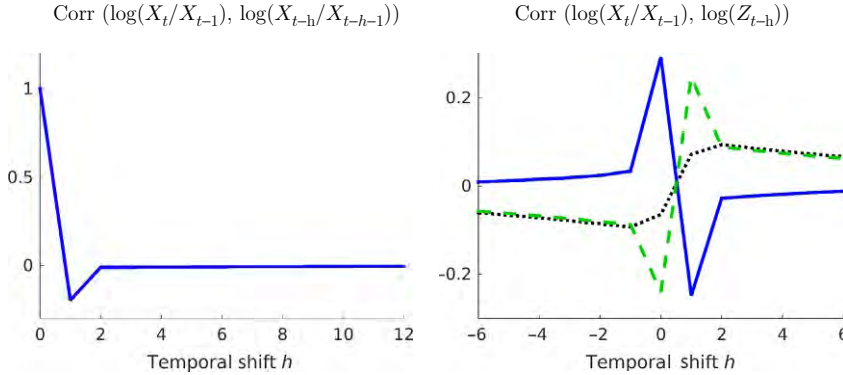
Fig. 16 Autocorrelations. *Notes: Right panel*: correlations of output growth with labor share (*solid*), inflation (*dotted*), and interest rates (*dashed*).

Correlations can be easily computed by normalizing the entries of the autocovariance matrices using the respective standard deviations. Fig. 16 shows the model-implied autocorrelation function of output growth and the cross-correlations of output growth with the labor share, inflation, and interest rates as a function of the temporal shift h.

The law of motion for the state vector s_t can also be expressed as the infinite-order vector moving average (MA) process

$$y_t = \Psi_0 + \Psi_1 \sum_{s=0}^{\infty} \Phi_1^s \Phi_\epsilon \epsilon_{t-s}. \tag{85}$$

Based on the moving average representation, it is straightforward to compute the h-step-ahead forecast error, which is given by

$$e_{t|t-h} = y_t - \mathbb{E}_{t-h}[y_t] = \Psi_1 \sum_{s=0}^{h-1} \Phi_1^s \Phi_\epsilon \epsilon_{t-s}. \tag{86}$$

The h-step-ahead forecast error covariance matrix is given by

$$\mathbb{E}[e_{t|t-h} e_{t|t-h}'] = \Psi_1 \left(\sum_{s=0}^{h-1} \Phi_1^s \Phi_\epsilon \Phi_\epsilon' \Phi_1^{s'} \right) \Psi_1' \quad \text{with} \quad \lim_{h\to\infty} \mathbb{E}[e_{t|t-h} e_{t|t-h}'] = \Gamma_{ss}(0). \tag{87}$$

Under the assumption that $\mathbb{E}[\epsilon_t \epsilon_t'] = I$, it is possible to decompose the forecast error covariance matrix as follows. Let $I^{(j)}$ be defined by setting all but the j-th diagonal element of the identity matrix I to zero. Then we can write

$$I = \sum_{j=1}^{n_\epsilon} I^{(j)}. \tag{88}$$

Moreover, we can express the contribution of shock j to the forecast error for y_t as

$$e_{t|t-h}^{(j)} = \Psi_1 \sum_{s=0}^{h-1} \Phi_1^s \Phi_\epsilon I^{(j)} \epsilon_{t-s}.$$

(89)

Thus, the contribution of shock j to the forecast error variance of observation $y_{i,t}$ is given by the ratio

$$\text{FEVD}(i,j,h) = \frac{\left[\Psi_1 \left(\sum_{s=0}^{h-1} \Phi_1^s \Phi_\epsilon I^{(j)} \Phi_\epsilon' \Phi_1^{s'}\right)\Psi_1'\right]_{ii}}{\left[\Psi_1 \left(\sum_{s=0}^{h-1} \Phi_1^s \Phi_\epsilon \Phi_\epsilon' \Phi_1^{s'}\right)\Psi_1'\right]_{ii}},$$

(90)

where $[A]_{ij}$ denotes element (i,j) of a matrix A. Fig. 17 shows the contribution of the four shocks to the forecast error variance of output growth, the labor share, inflation, and interest rates in the stylized DSGE model. Given the choice of parameters θ in
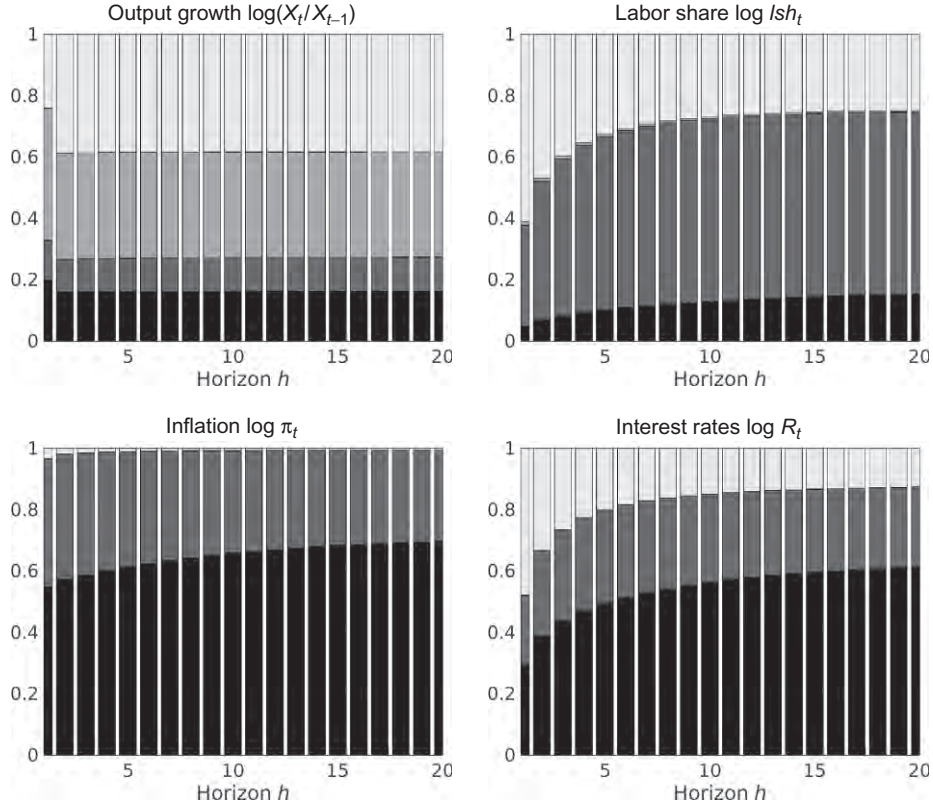

Fig. 17 Forecast error variance decomposition. *Notes*: The stacked bar plots represent the cumulative forecast error variance decomposition. The bars, from darkest to lightest, represent the contributions of ϕ_t, λ_t, z_t, and $\varepsilon_{R,t}$.

Table 5, most of the variation in output growth is due to the technology and the monetary policy shocks. The labor share fluctuations are dominated by the mark-up shock λ_t, in particular in the long run. Inflation and interest rate movements are strongly influenced by the preference shock ϕ_t and the mark-up shock λ_t.

8.2.2 Spectrum

Instead of studying DSGE model implications over different forecasting horizons, one can also consider different frequency bands. There is a long tradition of frequency domain analysis in the time series literature. A classic reference is Priestley (1981). We start with a brief discussion of the linear cyclical model, which will be useful for interpreting some of the formulas presented subsequently. Suppose that y_t is a scalar time series that follows the process

$$y_t = 2\sum_{j=1}^{m} a_j \left(\cos\theta_j \cos(\omega_j t) - \sin\theta_j \sin(\omega_j t) \right), \tag{91}$$

where $\theta_j \sim iidU[-\pi,\pi]$ and $0 \leq \omega_j \leq \omega_{j+1} \leq \pi$. The random variables θ_j cause a phase shift of the cycle and are assumed to be determined in the infinite past. In a nutshell, the model in (91) expresses the variable y_t as the sum of sine and cosine waves that differ in their frequency. The interpretation of the ω_j's depends on the length of the period t. Suppose the model is designed for quarterly data and $\omega_j = (2\pi)/32$. This means that it takes 32 periods to complete the cycle. Business cycles typically comprise cycles that have a duration of 8–32 quarters, which would correspond to $\omega_j \in [0.196, 0.785]$ for quarterly t.

Using Euler's formula, we rewrite the cyclical model in terms of an exponential function:

$$y_t = \sum_{j=-m}^{m} A(\omega_j)e^{i\omega_j t}, \tag{92}$$

where $\omega_{-j} = -\omega_j$, $i = \sqrt{-1}$, and

$$A(\omega_j) = \begin{cases} a_j\left(\cos\theta_{|j|} + i\sin\theta_{|j|} \right) & \text{if } j > 0 \\ a_j\left(\cos\theta_{|j|} - i\sin\theta_{|j|} \right) & \text{if } j < 0 \end{cases} \tag{93}$$

It can be verified that expressions (91) and (92) are identical. The function $A(\omega_j)$ captures the amplitude of cycles with frequency ω_j.

The spectral distribution function of y_t on the interval $\omega \in (-\pi,\pi]$ is defined as

$$F_{yy}(\omega) = \sum_{j=-m}^{m} \mathbb{E}[A(\omega_j)\overline{A(\omega_j)}]\mathbb{I}\{\omega_j \leq \omega\}, \tag{94}$$

where $\mathbb{I}\{\omega_j \leq \omega\}$ denotes the indicator function that is one if $\omega_j \leq \omega$ and $\bar{z} = x - iy$ is the complex conjugate of $z = x + iy$. If $F_{yy}(\omega)$ is differentiable with respect to ω, then we can define the spectral density function as

$$f_{\gamma\gamma}(\omega) = dF_{\gamma\gamma}(\omega)d\omega. \tag{95}$$

If a process has a spectral density function $f_{\gamma\gamma}(\omega)$, then the covariances can be expressed as

$$\Gamma_{\gamma\gamma}(h) = \int_{(-\pi,\pi]} e^{ih\omega} f_{\gamma\gamma}(\omega)d\omega. \tag{96}$$

For the linear cyclical model in (91) the autovariances are given by

$$\Gamma_{\gamma\gamma}(h) = \sum_{j=-m}^{m} \mathbb{E}[A(\omega_j)\overline{A(\omega_j)}]e^{i\omega_j h} = \sum_{j=-m}^{m} a_j^2 e^{i\omega_j h}. \tag{97}$$

The spectral density uniquely determines the entire sequence of autocovariances. Moreover, the converse is also true. The spectral density can be obtained from the auto-covariances of γ_t as follows:

$$f_{\gamma\gamma}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma_{\gamma\gamma}(h)e^{-i\omega h}. \tag{98}$$

The formulas (96) and (98) imply that the spectral density function and the sequence of autocovariances contain the same information. Their validity is not restricted to the linear cyclical model and they extend to vector-valued γ_t's. Recall that for the DSGE model defined by the state-space system (76) and (78) the autocovariance function for the state vector s_t was defined as $\Gamma_{ss}(h) = \Phi_1^h \Gamma_{ss}(0)$. Thus,

$$\begin{aligned} f_{ss}(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Phi_1^h \Gamma_{ss}(0)e^{-i\omega h} \\ &= \frac{1}{2\pi}\left(I - \Phi_1' e^{i\omega}\right)^{-1} \Phi_\epsilon \Phi_\epsilon' \left(I - \Phi_1 e^{-i\omega}\right)^{-1}. \end{aligned} \tag{99}$$

The contribution of shock j to the spectral density is given by

$$f_{ss}^{(j)}(\omega) = \frac{1}{2\pi}\left(I - \Phi_1' e^{i\omega}\right)^{-1} \Phi_\epsilon \mathcal{I}^{(j)} \Phi_\epsilon' \left(I - \Phi_1 e^{-i\omega}\right)^{-1}. \tag{100}$$

The spectral density for the observables γ_t (and the contribution of shock j to the spectral density) can be easily obtained as

$$f_{\gamma\gamma}(\omega) = \Psi_1 f_{ss}(\omega)\Psi_1' \quad \text{and} \quad f_{\gamma\gamma}^{(j)}(\omega) = \Psi_1 f_{ss}^{(j)}(\omega)\Psi_1'. \tag{101}$$

Fig. 18 depicts the spectral density functions for output growth, the labor share, inflation, and interest rates for the stylized DSGE model conditional on the parameters in Table 5. Note that $f_{\gamma\gamma}(\omega)$ is a matrix valued function. The four panels correspond to the diagonal elements of this function, providing a summary of the univariate autocovariance properties of the four series. Each panel stacks the contributions of the four shocks to the spectral densities. Because the shocks are independent and evolve according to AR(1) processes, the spectral density peaks at the origin and then decays as the frequency increases.
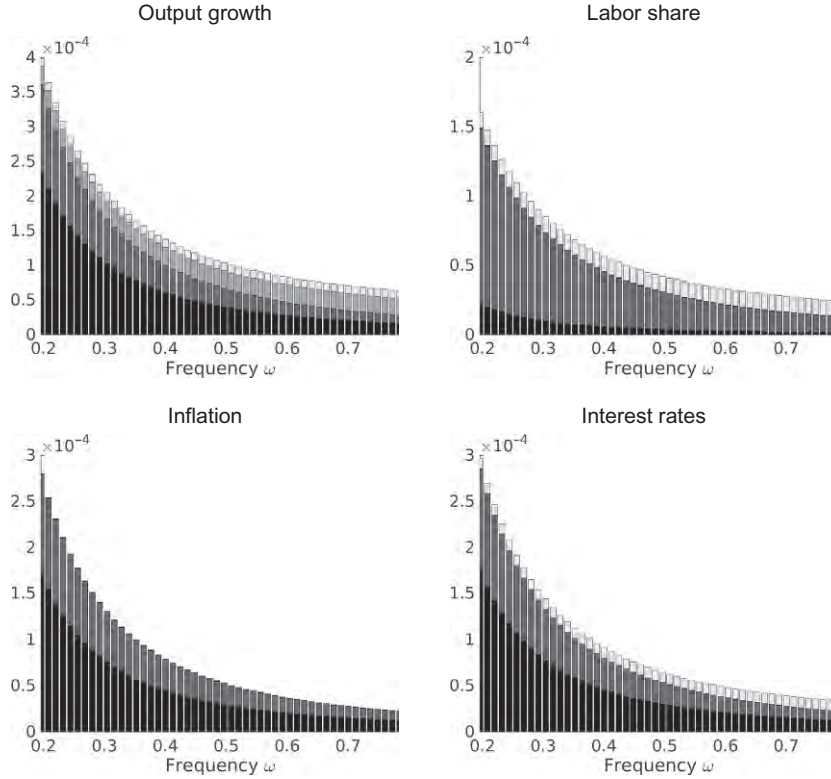
Fig. 18 Spectral decomposition. *Notes*: The stacked bar plots depict cumulative spectral densities. The bars, from darkest to lightest, represent the contributions of ϕ_t, λ_t, z_t, and $\varepsilon_{R,t}$.

8.2.3 Impulse Response Functions

An important tool for studying the dynamic effects of exogenous shocks are impulse response functions (IRFs). Formally, impulse responses in a DSGE model can be defined as the difference between two conditional expectations:

$$\text{IRF}(i, j, h|s_{t-1}) = \mathbb{E}\left[y_{i,t+h} \mid s_{t-1}, \epsilon_{j,t} = 1\right] - \mathbb{E}\left[y_{i,t+h} \mid s_{t-1}\right]. \tag{102}$$

Both expectations are conditional on the initial state s_{t-1} and integrate over current and future realizations of the shocks ϵ_t. However, the first term also conditions on $\epsilon_{j,t} = 1$, whereas the second term averages of $\epsilon_{j,t}$. In a linearized DSGE model with a state–space representation of the form (76) and (78), we can use the linearity and the property that $\mathbb{E}[\epsilon_{t+h}|s_{t-1}] = 0$ for $h = 0, 1, \ldots$ to deduce that

$$\text{IRF}(., j, h) = \Psi_1 \frac{\partial}{\partial \epsilon_{j,t}} s_{t+h} = \Psi_1 \Phi_1^h [\Phi_\epsilon]_{.j}, \tag{103}$$

where $[A]_{.j}$ is the j-th column of a matrix A. We dropped s_{t-1} from the conditioning set to simplify the notation.
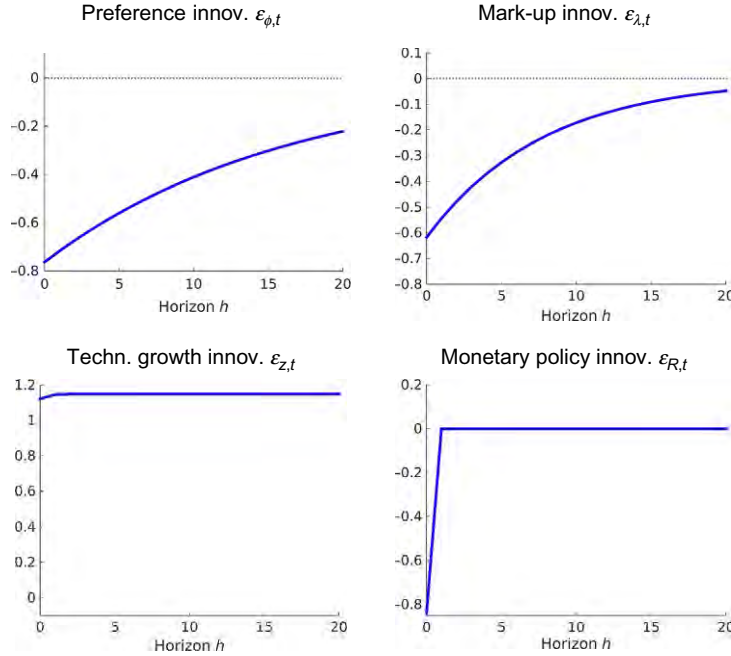
Fig. 19 Impulse responses of log output $100 \log (X_{t+h}/X_t)$.

Fig. 19 depicts the impulse response functions for the stylized DSGE model of log output to the four structural shocks, which can be easily obtained from (69) and the laws of motion of the exogenous shock processes. The preference and mark–up shocks lower output upon impact. Subsequently, output reverts back to its steady state. The speed of the reversion is determined by the autoregressive coefficient associated with the exogenous shock process. The technology growth shock raises the log level of output permanently, whereas a monetary policy shock has only a one-period effect on output.

8.2.4 Conditional Moment Restrictions

The intertemporal optimality conditions take the form of conditional moment restrictions. For instance, rearranging the terms in the New Keynesian Phillips (60) curve, we can write

$$\mathbb{E}_{t-1}\left[\hat{\pi}_{t-1} - \beta\hat{\pi}_t - \kappa_p\left(\widehat{lsh}_{t-1} + \lambda_{t-1}\right)\right] = 0. \tag{104}$$

The conditional moment condition can be converted into a vector of unconditional moment conditions as follows. Let $\mathcal{F}_t$ denote the sigma algebra generated by the infinite histories of $\{\gamma_\tau, s_\tau, \epsilon_\tau\}_{\tau=-\infty}^{t}$ and let $\tilde{Z}_t$ be a vector of random variables that is measurable with respect to $\mathcal{F}_t$, meaning that its value is determined based on information on current and past $(\gamma_t, s_t, \epsilon_t)$. Then for every such vector $\tilde{Z}_{t-1}$,

$$\mathbb{E}\left[\widetilde{Z}_{t-1}\left(\hat{\pi}_{t-1} - \beta\hat{\pi}_t - \kappa_p(\widehat{lsh}_{t-1} + \lambda_{t-1})\right)\right]$$
$$= \mathbb{E}\left[\widetilde{Z}_{t-1}\mathbb{E}_{t-1}\left[\hat{\pi}_{t-1} - \beta\hat{\pi}_t - \kappa_p(\widehat{lsh}_{t-1} + \lambda_{t-1})\right]\right] = 0, \tag{105}$$

where $\mathbb{E}_{t-1}[\cdot] = \mathbb{E}[\cdot|\mathcal{F}_{t-1}]$.

The moment conditions derived from the New Keynesian Phillips curve involve the latent price mark-up shock λ_t, which will cause difficulties if one tries to use (105) in an estimation objective function. Now consider the consumption Euler equation (56) instead. Recall that the measurement equations imply that

$$\hat{x}_t - \hat{x}_{t-1} + z_t = \log X_t - \log X_{t-1} - \log\gamma \quad \text{and} \quad \hat{R}_t = \log R_t - \log(\pi^*\gamma/\beta).$$

Thus, we can write

$$\mathbb{E}_{t-1}[-\log(X_t/X_{t-1}) + \log R_{t-1} - \log\pi_t - \log(1/\beta)] = 0. \tag{106}$$

The terms γ and $\log\pi^*$ that appear in the steady-state formulas for the nominal interest rate and inflation cancel and the conditional moment condition only depends on observables and the model parameters, but not on latent variables. Finally, as long as the monetary policy shock satisfies the martingale difference sequences property $\mathbb{E}_{t-1}[\epsilon_{R,t}] = 0$, we obtain from the monetary policy rule the condition that

$$\mathbb{E}_{t-1}[\log R_t - \log(\gamma/\beta) - \psi\log\pi_t - (1-\psi)\log\pi^*] = 0. \tag{107}$$

Both (106) and (107) can be converted into an unconditional moment condition using an $\mathcal{F}_{t-1}$ measurable random vector $\mathcal{Z}_{t-1}$ as in (105).

8.2.5 Analytical Calculation of Moments vs Simulation Approximations

As previously shown, formulas for autocovariance functions, spectra, and impulse response functions for a linearized DSGE model can be derived analytically from the state-space representation. These analytical expressions can then be numerically evaluated for different vectors of parameter values θ. For DSGE models solved with perturbation methods, there are also analytical formulas available that exploit a conditionally linear structure of some perturbation solutions; see Andreasen et al. (2013). For a general nonlinear DSGE model, the implied moments have to be computed using Monte Carlo simulation. For instance, let $Y^*_{1:T}$ denote a sequence of observations simulated from the state-space representation of the DSGE model by drawing an initial state vector s_0 and innovations ϵ_t from their model-implied distributions, then

$$\frac{1}{T}\sum_{t=1}^{T} y_t^* \xrightarrow{a.s.} \mathbb{E}[y_t], \tag{108}$$

provided that the DSGE model–implied γ_t is strictly stationary and ergodic.[ai] The downside of Monte Carlo approximations is that they are associated with a simulation error. We will come back to this problem in Section 11.2, when we use simulation approximations of moments to construct estimators of θ.

8.3 Empirical Analogs

We now examine sample analogs of the population moments derived from the state-space representation of the DSGE model using US data. The time series were downloaded from the FRED database maintained by the Federal Reserve Bank of St. Louis and we report the series labels in parentheses. For real aggregate output, we use quarterly, seasonally adjusted GDP at the annual rate that has been pegged to 2009 dollars (GDPC96). We turn GDP into growth rates by taking logs and then differencing. The labor share is defined as Compensation of Employees (COE) divided by nominal GDP (GDP). Both series are quarterly and seasonally adjusted at the annual rate. We use the log labor share as the observable. Inflation rates are computed from the implicit price deflator (GDPDEF) by taking log differences. Lastly, for the interest rate, we use the Effective Federal Funds Rate (FEDFUNDS), which is monthly, and not seasonally adjusted. Quarterly interest rates are obtained by taking averages of the monthly rates. Throughout this section we focus on the post-Great Moderation and pre-Great Recession period and restrict our sample from 1984:Q1 to 2007:Q4.

8.3.1 Autocovariances
The sample analog of the population autocovariance $\Gamma_{yy}(h)$ is defined as

$$\hat{\Gamma}_{yy}(h) = \frac{1}{T}\sum_{t=h}^{T}(\gamma_t - \hat{\mu}_y)(\gamma_{t-h} - \hat{\mu}_y)', \quad \text{where} \quad \hat{\mu}_y = \frac{1}{T}\sum_{t=1}^{T}\gamma_t. \tag{109}$$

Under suitable regularity conditions, eg, covariance stationarity of the vector process γ_t, a sufficiently fast decay of the serial correlation in γ_t, and some bounds on higher-order moments of γ_t, the sample autocovariance $\hat{\Gamma}_{yy}(h)$ converges to the population autocovariance $\Gamma_{yy}(h)$, satisfying a strong law of large numbers (SLLN) and a central limit theorem (CLT).

   If the object of interest is a sequence of autocovariance matrices, then it might be more efficient to first estimate an auxiliary model and then convert the parameter estimates of the auxiliary model into estimates of the autocovariance sequence. A natural class of auxiliary models is provided by linear vector autoregressions (VARs). For illustrative purposes consider the following VAR(1):

[ai] A sequence of random variables X_T converges to a limit random variable X almost surely (a.s.) if the set of trajectories for which $X_T \nrightarrow X$ has probability zero.

$$y_t = \Phi_1 y_{t-1} + \Phi_0 + u_t, \quad u_t \sim iid(0, \Sigma). \tag{110}$$

The OLS estimator of Φ_1 can be approximated by

$$\hat{\Phi}_1 = \hat{\Gamma}_{yy}(1)\hat{\Gamma}_{yy}^{-1}(0) + O_p(T^{-1}), \quad \hat{\Sigma} = \hat{\Gamma}_{yy}(0) - \hat{\Gamma}_{yy}(1)\hat{\Gamma}_{yy}^{-1}(0)\hat{\Gamma}'_{yy}(1) + O_p(T^{-1}) \tag{111}$$

The $O_p(T^{-1})$ terms arise because the range of the summations in the definition of the sample autocovariances in (109) and the definition of the OLS estimator are not exactly the same.[aj] Suppose that now we plug the OLS estimator into the autocovariance formulas associated with the VAR(1) (see (82) and (83)), then:

$$\hat{\Gamma}_{yy}^V(0) = \hat{\Gamma}_{yy}(0) + O_p(T^{-1}), \quad \hat{\Gamma}_{yy}^V(h) = \left(\hat{\Gamma}_{yy}(1)\hat{\Gamma}_{yy}^{-1}(0)\right)^h \hat{\Gamma}_{yy}(0) + O_p(T^{-1}). \tag{112}$$

Note that for $h = 0, 1$ we obtain $\hat{\Gamma}_{yy}^V(1) = \hat{\Gamma}_{yy}(1) + O_p(T^{-1})$. For $h > 1$ the VAR(1) plug-in estimate of the autocovariance matrix differs from the sample autocovariance matrix. If the actual time series are well approximated by a VAR(1), then the plug-in autocovariance estimate tends to be more efficient than the sample autocovariance estimate $\hat{\Gamma}_{yy}(h)$; see, for instance, Schorfheide (2005b).

In practice, a VAR(1) may be insufficient to capture the dynamics of a time series y_t. In this case the autocovariances can be obtained from a VAR(p)

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_0 + u_t, \quad u_t \sim iid(0, \Sigma). \tag{113}$$

The appropriate lag length p can be determined with a model selection criterion, eg, the Schwarz (1978) criterion, which is often called the Bayesian information criterion (BIC). The notationally easiest way (but not the computationally fastest way) is to rewrite the VAR(p) in companion form. This entails expressing the law of motion for the stacked vector $\tilde{y}_t = [y_t', y_{t-1}', \ldots, y_{t-p+1}']$ as VAR(1):

$$\tilde{y}_t = \tilde{\Phi}_1 \tilde{y}_{t-1} + \tilde{\Phi}_0 + \tilde{u}_t, \quad \tilde{u}_t \sim iid(0, \tilde{\Sigma}), \tag{114}$$

where

$$\tilde{\Phi}_1 = \begin{bmatrix} \Phi_1 & \ldots & \Phi_{p-1} & \Phi_p \\ I_{n\times n} & \ldots & 0_{n\times n} & 0_{n\times n} \\ \vdots & \ddots & \vdots & \vdots \\ 0_{n\times n} & \ldots & I_{n\times n} & 0_{n\times n} \end{bmatrix}, \quad \tilde{\Phi}_0 = \begin{bmatrix} \Phi_0 \\ 0_{n(p-1)\times 1} \end{bmatrix},$$

$$\tilde{\epsilon}_t = \begin{bmatrix} \epsilon_t \\ 0_{n(p-1)\times 1} \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} \Sigma & 0_{n\times n(p-1)} \\ 0_{n(p-1)\times n} & 0_{n(p-1)\times n(p-1)} \end{bmatrix}.$$

[aj] We say that a sequence of random variables is $O_p(T^{-1})$ if TX_T is stochastically bounded as $T \to \infty$.
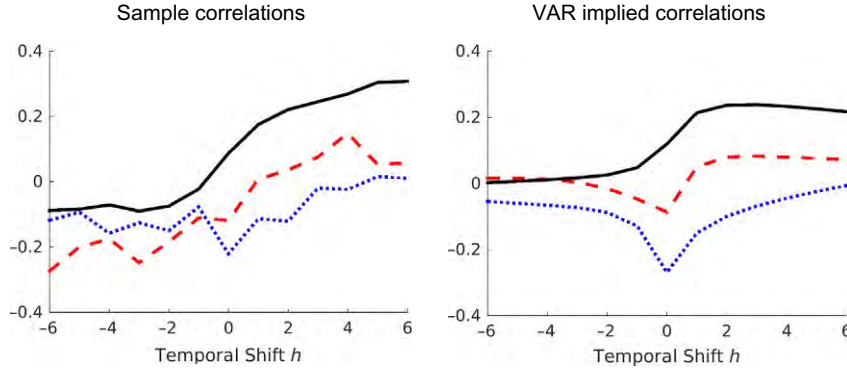
Fig. 20 Empirical cross-correlations $\mathrm{Corr}(\log(X_t/X_{t-1}), \log Z_{t-h})$. *Notes*: Each plot shows the correlation of output growth $\log(X_t/X_{t-1})$ with interest rates (*solid*), inflation (*dashed*), and the labor share (*dotted*), respectively. *Left panel*: correlation functions are computed from sample autocovariance matrices $\hat{\Gamma}_{yy}(h)$. *Right panel*: correlation functions are computed from estimated VAR(1).

The autocovariances for $\tilde{y}_t$ are then obtained by adjusting the VAR(1) formulas (112) to $\tilde{y}_t$ and reading off the desired submatrices that correspond to the autocovariance matrices for y_t using the selection matrix $M' = [I_n, 0_{n \times n(p-1)}]$ such that $y_t = M' \tilde{y}_t$.

We estimate a VAR for output growth, labor share, inflation, and interest rates. The lag length $p = 1$ is determined by the BIC. The left panel of Fig. 20 shows sample cross-correlations (obtained from $\hat{\Gamma}_{yy}(h)$ in (109)) between output growth and leads and lags of the labor share, inflation, and interest rates, respectively. The right panel depicts correlation functions derived from the estimated VAR(1). The two sets of correlation functions are qualitatively similar but quantitatively different. Because the VAR model is more parsimonious, the VAR–implied correlation functions are smoother.

8.3.2 Spectrum
An intuitively plausible estimate of the spectrum is the sample periodogram, defined as

$$\hat{f}_{yy}(\omega) = \frac{1}{2\pi} \sum_{h=-T+1}^{T-1} \hat{\Gamma}_{yy}(h) e^{-i\omega h} = \frac{1}{2\pi} \left(\hat{\Gamma}_{yy}(0) + \sum_{h=1}^{T-1} (\hat{\Gamma}_{yy}(h) + \hat{\Gamma}_{yy}(h)') \cos \omega h \right). \quad (115)$$

While the sample periodogram is an asymptotically unbiased estimator of the population spectral density, it is inconsistent because its variance does not vanish as the sample size $T \to \infty$. A consistent estimator can be obtained by smoothing the sample periodogram across adjacent frequencies. Define the fundamental frequencies

$$\omega_j = j \frac{2\pi}{T}, \quad j = 1, \ldots, (T-1)/2$$

and let $K(x)$ denote a kernel function with the property that $\int K(x) dx = 1$. A smoothed periodogram can be defined as

$$\bar{f}_{\gamma\gamma}(\omega) = \frac{\pi}{\lambda(T-1)/2} \sum_{j=1}^{(T-1)/2} K\left(\frac{\omega_j - \omega}{\lambda}\right) \hat{f}_{\gamma\gamma}(\omega_j). \tag{116}$$

An example of a simple kernel function is

$$K\left(\frac{\omega_j - \omega}{\lambda}\right) \hat{f}_{\gamma\gamma}(\omega_j) = \mathbb{I}\left\{-\frac{1}{2} < \frac{\omega_j - \omega}{\lambda} < \frac{1}{2}\right\} = \mathbb{I}\{\omega_j \in B(\omega|\lambda)\},$$

where $B(\omega|\lambda)$ is a frequency band. The smoothed periodogram estimator $\bar{f}_{\gamma\gamma}(\omega)$ is consistent, provided that the bandwidth shrinks to zero, that is, $\lambda \to 0$ as $T \to \infty$, and the number of ω_j's in the band, given by $\lambda T(2\pi)$, tends to infinity. In the empirical application below we use a Gaussian kernel, meaning that $K(x)$ equals the probability density function of a standard normal random variable.

An estimate of the spectral density can also be obtained indirectly through the estimation of the VAR(p) in (113). Define

$$\Phi = [\Phi_1, \dots, \Phi_p, \Phi_0]' \quad \text{and} \quad M(z) = [Iz, \dots, Iz^p],$$

and let $\hat{\Phi}$ be an estimator of Φ. Then a VAR(p) plug-in estimator of the spectral density is given by

$$\hat{f}_{\gamma\gamma}^V(\omega) = \frac{1}{2\pi}[I - \hat{\Phi}'M'(e^{-i\omega})]^{-1}\hat{\Sigma}[I - M(e^{-i\omega})\hat{\Phi}]^{-1}. \tag{117}$$

This formula generalizes the VAR(1) spectral density in (99) to a spectral density for a VAR(p).

Estimates of the spectral densities of output growth, log labor share, inflation, and interest rates are reported in Fig. 21. The shaded areas highlight the business cycle frequencies. Because the autocorrelation of output growth is close to zero, the spectral density is fairly flat. The other three series have more spectral mass at the low frequency, which is a reflection of the higher persistence. The labor share has a pronounced hump-shaped spectral density, whereas the other spectral densities of interest and inflation rates are monotonically decreasing in the frequency ω. The smoothness of the periodogram estimates $\bar{f}_{\gamma\gamma}(\omega)$ depends on the choice of the bandwidth. The figure is based on a Gaussian kernel with standard deviation 0.15, which, roughly speaking, averages the sample periodogram over a frequency band of 0.6. While the shapes of the smoothed periodograms and the VAR–based spectral estimates are qualitatively similar, the spectral density is lower according to the estimated VAR.

8.3.3 Impulse Response Functions
The VAR(p) in (113) is a so-called reduced-form VAR because the innovations u_t do not have a specific structural interpretation—they are simply one-step-ahead forecast errors. The impulse responses that we constructed for the DSGE model are responses to
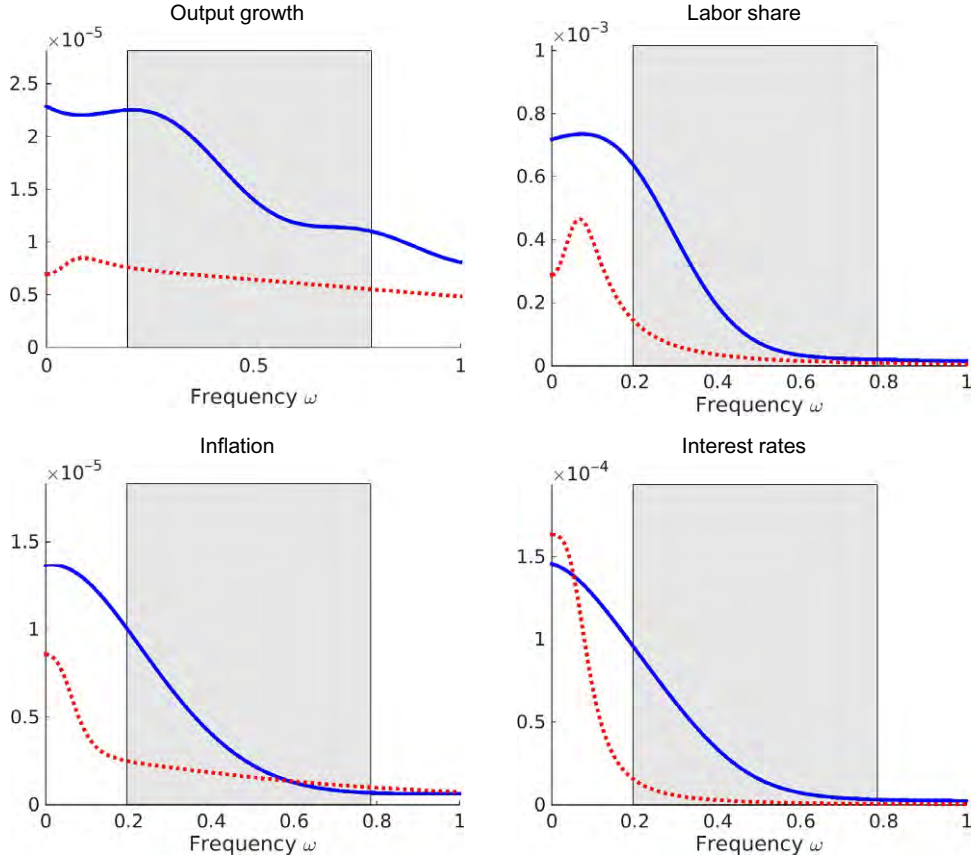
Fig. 21 Empirical spectrum. *Notes*: The *dotted lines* are spectra computed from an estimated VAR(1); the *solid lines* are smoothed periodograms based on a Gaussian kernel with standard deviation 0.15. The *shaded areas* indicate business cycle frequencies (0.196–0.785).

innovations in the structural shock innovations that contribute to the forecast error for several observables simultaneously. In order to connect VAR-based impulse responses to DSGE model-based responses, one has to link the one-step-ahead forecast errors to a vector of structural innovations ϵ_t. We assume that

$$u_t = \Phi_\epsilon \epsilon_t = \Sigma_{tr} \Omega \epsilon_t, \tag{118}$$

where Σ_{tr} is the unique lower-triangular Cholesky factor of Σ with nonnegative diagonal elements, and Ω is an $n \times n$ orthogonal matrix satisfying $\Omega \Omega' = I$. The second equality ensures that the covariance matrix of u_t is preserved in the sense that

$$\Phi_\epsilon \Phi'_\epsilon = \Sigma_{tr} \Omega \Omega' \Sigma'_{tr} = \Sigma. \tag{119}$$

By construction, the covariance matrix of the forecast error is invariant to the choice of Ω, which implies that it is not possible to identify Ω from the data. In turn, much of the literature on structural VARs reduces to arguments about an appropriate set of restrictions for the matrix Ω. Detailed surveys about the restrictions, or identification schemes, that have been used in the literature to identify innovations to technology, monetary policy, government spending, and other exogenous shocks can be found, for instance, in Cochrane (1994), Christiano et al. (1999), Stock and Watson (2001), and Ramey (2016). Conditional on an estimate of the reduced-form coefficient matrices Φ and Σ and an identification scheme for one or more columns of Ω, it is straightforward to express the impulse response as

$$\widehat{IRF}^V(.,j,h) = C_h(\hat{\Phi})\hat{\Sigma}_{tr}[\Omega]_{.j}, \tag{120}$$

where the moving average coefficient matrix $C_h(\hat{\Phi})$ can be obtained from the companion form representation of the VAR in (114): $C_h(\Phi) = M'\tilde{\Phi}_1^h M$ with $M' = [I_n, 0_{n \times n(p-1)}]$.

For illustrative purposes, rather than conditioning the computation of impulse response functions on a particular choice of Ω, we follow the recent literature on sign restrictions; see Faust (1998), Canova and De Nicoló (2002), and Uhlig (2005). The key idea of this literature is to restrict the matrices Ω to a set $\mathcal{O}(\Phi, \Sigma)$ such that the implied impulse response functions satisfy certain sign restrictions. This means that the magnitude of the impulse responses are only set-identified. Using our estimated VAR(1) in output growth, log labor share, inflation, and interest rates, we impose the condition that in response to a contractionary monetary policy shock interest rates increase and inflation is negative for four quarters. Without loss of generality, we can assume that the shocks are ordered such that the first column of Ω, denoted by q, captures the effect of the monetary policy shock. Conditional on the reduced-form VAR coefficient estimates $(\hat{\Phi}, \hat{\Sigma})$, we can determine the set of unit-length vectors q such that the implied impulse responses satisfy the sign restrictions. The bands depicted in Fig. 22 delimit the upper and lower bounds of the estimated identified sets for the pointwise impulse responses of output, labor share, inflation, and interest rates to a monetary policy shock. The sign restrictions that are imposed on the monetary policy shock are not sufficiently strong to determine the sign of the output and labor share responses to a monetary policy shock. Note that if a researcher selects a particular q (possibly as a function of the reduced-form parameters Φ and Σ), then the bands in the figure would reduce to a single line, which is exemplified by the solid line in Fig. 22.

8.3.4 Conditional Moment Restrictions
The unconditional moment restrictions derived from the equilibrium conditions of the DSGE model discussed in Section 8.2.4 have sample analogs in which the population expectations are replaced by sample averages. A complication arises if the moment
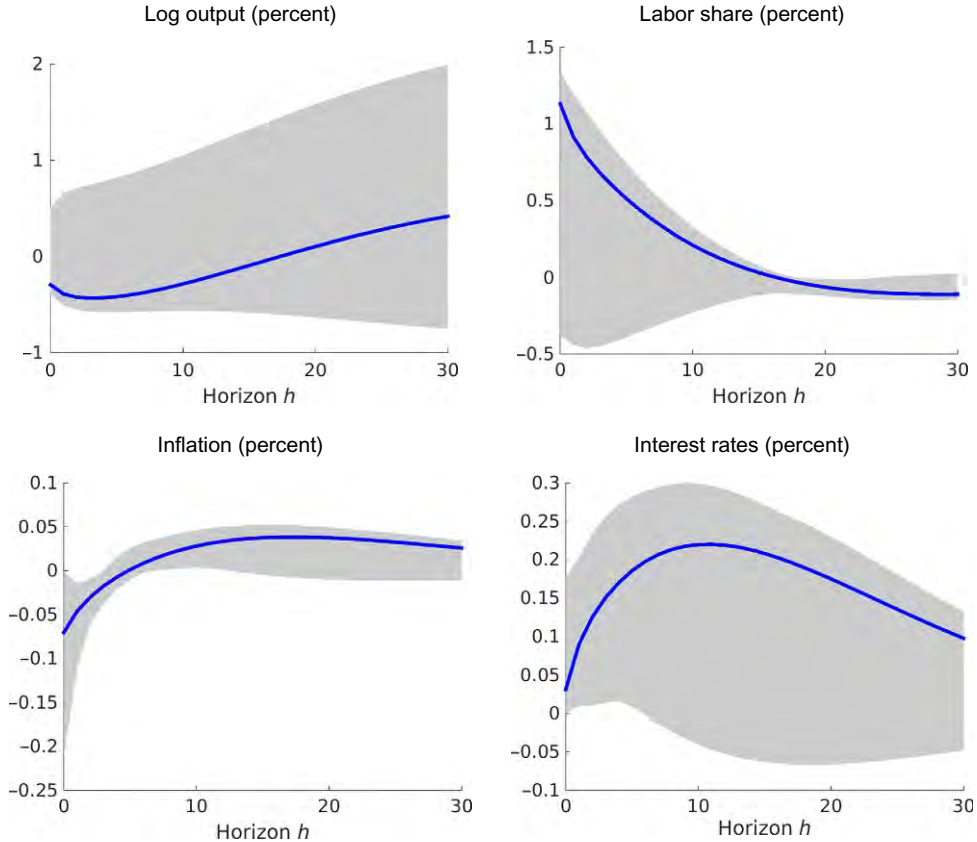
Fig. 22 Impulse responses to a monetary policy shock. *Notes*: Impulse responses to a one-standard-deviation monetary policy shock. Inflation and interest rate responses are not annualized. The bands indicate pointwise estimates of identified sets for the impulse responses based on the assumption that a contractionary monetary policy shock raises interest rates and lowers inflation for four quarters. The *solid line* represents a particular impulse response function contained in the identified set.

conditions contain latent variables, eg, the shock process λ_t in the moment condition (105) derived from the New Keynesian Phillips curve. Sample analogs of population moment conditions can be used to form generalized method of moments (GMM) estimators, which are discussed in Section 11.4.

8.4 Dealing with Trends

Trends are a salient feature of macroeconomic time series. The stylized DSGE model presented in Section 8.1 features a stochastic trend generated by the productivity process $\log Z_t$, which evolves according to a random walk with drift. While the trend in productivity induces a common trend in consumption, output, and real wages, the model

Fig. 23 Consumption-output ratio and Labor share (in logs).

specification implies that the log consumption–output ratio and the log labor share are stationary. Fig. 23 depicts time series of the US log consumption–output ratio and the log labor share for the United States from 1965 to 2014. Here the consumption–output ratio is defined as Personal Consumption Expenditure on Services (PCESV) plus Personal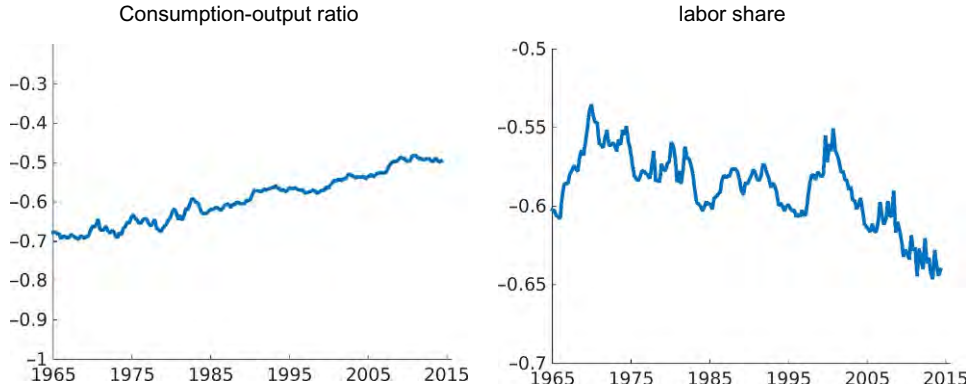 Consumption Expenditure on nondurable goods (PCND) divided by nominal GDP. The consumption–output ratio has a clear upward trend and the labor share has been falling since the late 1990s. Because these trends are not captured by the DSGE model, they lead to a first-order discrepancy between actual US and model–generated data.

Most DSGE models that are used in practice have counterfactual trend implications because they incorporate certain cotrending restrictions, eg, a balanced growth path along which output, consumption, investment, the capital stock, and real wages exhibit a common trend and hours worked and the return on capital are stationary, that are to some extent violated in the data as we have seen in the above example. Researchers have explored various remedies to address the mismatch between model and data, including: (i) detrending each time series separately and fitting the DSGE model to detrended data; (ii) applying an appropriate trend filter to both actual data and model–implied data when confronting the DSGE model with data; (iii) creating a hybrid model, eg, Canova (2014) that consists of a flexible, nonstructural trend component and uses the structural DSGE model to describe fluctuations around the reduced-form trend; and (iv) incorporating more realistic trends directly into the structure of the DSGE model. From a modeling perspective, option (i) is the least desirable and option (iv) is the most desirable choice.

9. STATISTICAL INFERENCE

DSGE models have a high degree of theoretical coherence. This means that the functional forms and parameters of equations that describe the behavior of macroeconomic aggregates are tightly restricted by optimality and equilibrium conditions. In turn, the

family of probability distributions $p(Y|\theta)$, $\theta \in \Theta$, generated by a DSGE model tends to be more restrictive than the family of distributions associated with an atheoretical model, such as a reduced-form VAR as in (113). This may place the empirical researcher in a situation in which the data favor the atheoretical model and the atheoretical model generates more accurate forecasts, but a theoretically coherent model is required for the analysis of a particular economic policy. The subsequent discussion of statistical inference will devote special attention to this misspecification problem.

The goal of statistical inference is to infer an unknown parameter vector θ from observations Y; to provide a measure of uncertainty about θ; and to document the fit of the statistical model. The implementation of these tasks becomes more complicated if the statistical model suffers from misspecification. Confronting DSGE models with data can essentially take two forms. If it is reasonable to assume that the probabilistic structure of the DSGE model is well specified, then one can ask how far the observed data $Y_{1:T}^o$ or sample statistics $\mathcal{S}(Y_{1:T}^o)$ computed from the observed data fall into the tails of the model–implied distribution derived from $p(Y_{1:T}|\theta)$. The parameter vector θ can be chosen to ensure that the density (likelihood) of $\mathcal{S}(Y_{1:T}^o)$ is high under the distribution $p(Y_{1:T}|\theta)$. If, on the other hand, there is a strong belief (possibly supported by empirical evidence) that the probabilistic structure of the DSGE model is not rich enough to capture the salient features of the observed data, it is more sensible to consider a reference model with a well-specified probabilistic structure, use it to estimate some of the population objects introduced in Section 8.2, and compare these estimates to their model counterparts.

In Section 9.1 we ask the question whether the DSGE model parameters can be determined based on observations Y and review the recent literature on identification. We then proceed by reviewing two modes of statistical inference: frequentist and Bayesian.[ak] We pay special attention to the consequences of model misspecification. Frequentist inference, introduced in Section 9.2, takes a preexperimental perspective and focuses on the behavior of estimators and test statistics, which are functions of the observations Y, in repeated sampling under the distribution $\mathbb{P}_\theta^Y$. Frequentist inference is conditioned on a "true" but unknown parameter θ, or on a data–generating process (DGP), which is a hypothetical probability distribution under which the data are assumed to be generated. Frequentist procedures have to be well behaved for all values of $\theta \in \Theta$. Bayesian inference, introduced in Section 9.3, takes a postexperimental perspective by treating the unknown parameter θ as a random variable and updating a prior distribution $p(\theta)$ in view of the data Y using Bayes Theorem to obtain the posterior distribution $p(\theta|Y)$.

Estimation and inference requires that the model be solved many times for different parameter values θ. The subsequent numerical illustrations are based on the stylized

---

[ak] A comparison between econometric inference approaches and the calibration approach advocated by Kydland and Prescott (1982) can be found in Ríos-Rull et al. (2012).

DSGE model introduced in Section 8.1, for which we have a closed-form solution. However, such closed-form solutions are the exception and typically not available for models used in serious empirical applications. Thus, estimation methods, both frequentist and Bayesian, have to be closely linked to model solution procedures. This ultimately leads to a trade-off: given a fixed amount of computational resources, the more time is spent on solving a model conditional on a particular θ, eg, through the use of a sophisticated projection technique, the less often an estimation objective function can be evaluated. For this reason, much of the empirical work relies on first-order perturbation approximations of DSGE models, which can be obtained very quickly. The estimation of models solved with numerically sophisticated projection methods is relatively rare, because it requires a lot of computational resources. Moreover, as discussed in Part I, perturbation solutions are more easily applicable to models with a high–dimensional state vector and such models, in turn, are less prone to misspecification and are therefore more easily amenable to estimation. However, the recent emergence of low-cost parallel programming environments and cloud computing will make it feasible for a broad group of researchers to solve and estimate elaborate nonlinear DSGE models in the near future.

9.1 Identification

The question of whether a parameter vector θ is identifiable based on a sample Y is of fundamental importance for statistical inference because one of the main objectives is to infer the unknown θ based on the sample Y. Suppose that the DSGE model generates a family of probability distributions $p(Y|\theta)$, $\theta \in \Theta$. Moreover, imagine a stylized setting in which data are in fact generated from the DSGE model conditional on some "true" parameter θ_0. The parameter vector θ_0 is globally identifiable if

$$p(Y|\theta) = p(Y|\theta_0) \quad \text{implies} \quad \theta = \theta_0. \tag{121}$$

The statement is somewhat delicate because it depends on the sample Y. From a preexperimental perspective, the sample is unobserved and it is required that (121) holds with probability one under the distribution $p(Y|\theta_0)$. From a postexperimental perspective, the parameter θ may be identifiable for some trajectories Y, but not for others. The following example highlights the subtle difference. Suppose that

$$y_{1,t}|(\theta, y_{2,t}) \sim iidN(\theta y_{2,t}, 1), \quad y_{2,t} = \begin{cases} 0 & \text{w.p. } 1/2 \\ \sim iidN(0,1) & \text{w.p. } 1/2 \end{cases}$$

Thus, with probability (w.p.) 1/2, one observes a trajectory along which θ is not identifiable because $y_{2,t} = 0$ for all t. If, on the other hand, $y_{2,t} \neq 0$, then θ is identifiable.

9.1.1 Local Identification
If condition (121) is only satisfied for values of θ in an open neighborhood of θ_0, then θ_0 is locally identified. Most of the literature has focused on devising procedures to check local

identification in linearized DSGE models with Gaussian innovations. In this case the distribution of $Y|\theta$ is a joint normal distribution and can be characterized by a $Tn_y \times 1$ vector of means $\mu(\theta)$ (where n is the dimension of the vector y_t) and a $Tn_y \times Tn_y$ covariance matrix $\Sigma(\theta)$. Defining $m(\theta) = [\mu(\theta)', vech(\Sigma(\theta))']'$, where $vech(\cdot)$ vectorizes the nonredundant elements of a symmetric matrix, we can restate the identification condition as

$$m(\theta) = m(\theta_0) \text{ implies } \theta = \theta_0. \tag{122}$$

Thus, verifying the local identification condition is akin to checking whether the Jacobian

$$\mathcal{J}(\theta) = \frac{\partial}{\partial \theta'} m(\theta) \tag{123}$$

is of full rank. This approach was proposed and applied by Iskrev (2010) to examine the identification of linearized DSGE models. If the joint distribution of Y is not Gaussian, say because the DSGE model innovations ϵ_t are non-Gaussian or because the DSGE model is nonlinear, then it is possible that θ_0 is not identifiable based on the first and second moments $m(\theta)$, but that there are other moments that make it possible to distinguish θ_0 from $\tilde{\theta} \neq \theta_0$.

Local identification conditions are often stated in terms of the so-called information matrix. Using Jensen's inequality, it is straightforward to verify that the Kullback–Leibler discrepancy between $p(Y|\theta_0)$ and $p(Y|\theta)$ is nonnegative:

$$\Delta_{KL}(\theta|\theta_0) = - \int \log\left(\frac{p(Y|\theta)}{p(Y|\theta_0)}\right) p(Y|\theta_0) dY \geq 0. \tag{124}$$

Under a nondegenerate probability distribution for Y, the relationship holds with equality only if $p(Y|\theta) = p(Y|\theta_0)$. Thus, we deduce that the Kullback–Leibler distance is minimized at $\theta = \theta_0$ and that θ_0 is identified if θ_0 is the unique minimizer of $\Delta_{KL}(\theta|\theta_0)$. Let $\ell(\theta|Y) = \log p(Y|\theta)$ denote the log-likelihood function and $\nabla_\theta^2 \ell(\theta|Y)$ denote the matrix of second derivatives of the log-likelihood function with respect to θ (Hessian), then (under suitable regularity conditions that allow the exchange of integration and differentiation)

$$\nabla_{\theta^2} \Delta_{KL}(\theta_0|\theta_0) = \int \nabla_{\theta^2} \ell(\theta_0|Y) p(Y|\theta_0) dY. \tag{125}$$

In turn, the model is locally identified at θ_0 if the expected Hessian matrix is nonsingular.

For linearized Gaussian DSGE models that can be written in the form $Y \sim N(\mu(\theta), \Sigma(\theta))$ we obtain

$$\int \nabla_\theta^2 \ell(\theta_0|Y) p(Y|\theta_0) dY = \mathcal{J}(\theta)' \Omega \mathcal{J}(\theta), \tag{126}$$

where Ω is the Hessian matrix associated with the unrestricted parameter vector $m = [\mu', vech(\Sigma)']'$ of a $N(\mu, \Sigma)$. Because Ω is a symmetric full-rank matrix of dimension $\dim(m)$, we deduce that the Hessian is of full rank whenever the Jacobian matrix in (123) is of full rank.

Qu and Tkachenko (2012) focus on the spectral density matrix of the process y_t. Using a frequency domain approximation of the likelihood function and utilizing the information matrix equality, they express the Hessian as the outer product of the Jacobian matrix of derivatives of the spectral density with respect to θ

$$G(\theta_0) = \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial \theta'} vec(f_{yy}(\omega)') \right)' \left(\frac{\partial}{\partial \theta'} vec(f_{yy}(\omega)) \right) d\omega \tag{127}$$

and propose to verify whether $G(\theta_0)$ is of full rank. The identification checks of Iskrev (2010) and Qu and Tkachenko (2012) have to be implemented numerically. For each conjectured θ_0 the user has to compute the rank of the matrices $\mathcal{J}(\theta_0)$ or $G(\theta_0)$, respectively. Because in a typical implementation the computation of the matrices relies on numerical differentiation (and integration), careful attention has be paid to the numerical tolerance level of the procedure that computes the matrix rank. Detailed discussions can be found in the two referenced papers.

Komunjer and Ng (2011) take a different route to assess the local identification of linearized DSGE models. They examine the relationship between the coefficients of the state-space representation of the DSGE model and the parameter vector θ. Recall that the state-space representation takes the form

$$y_t = \Psi_0(\theta) + \Psi_1(\theta), \quad s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t. \tag{128}$$

The notation highlights the dependence of the coefficient matrices on θ. Now stack the coefficients of the Ψ and Φ matrices in the vector ϕ:

$$\phi = \left[vec(\Psi_0)', vec(\Psi_1)', vec(\Phi_1)', vec(\Phi_\epsilon)' \right]'.$$

It is tempting to conjecture that θ is locally identifiable if the Jacobian matrix associated with the mapping from economic parameters θ to the reduced-form state-space parameters ϕ

$$\frac{\partial}{\partial \theta'} \phi(\theta) \tag{129}$$

has full column rank at θ_0. The problem with this conjecture is that the reduced-form parameters ϕ themselves are not identifiable. Let A be a nonsingular $n_s \times n_s$ matrix and Ω an $n_\epsilon \times n_\epsilon$ orthogonal matrix, then we can define

$$\tilde{s}_t = As_t, \quad \tilde{\epsilon}_t = \Omega\epsilon_t, \quad \tilde{\Psi}_1 = \Psi_1 A^{-1}, \quad \tilde{\Phi}_1 = \Phi_1 A^{-1}, \quad \tilde{\Phi}_\epsilon = A\Phi_\epsilon \Omega'$$

to obtain an observationally equivalent state-space system

$$y_t = \Psi_0 + \widetilde{\Psi}_1 \, \widetilde{s}_t, \quad s_t = \widetilde{\Phi}_1 \widetilde{s}_{t-1} + \widetilde{\Phi}_\epsilon \epsilon_t \tag{130}$$

with $\phi \neq \widetilde{\phi}$. Thus, the number of identifiable reduced-form parameters is smaller than the number of elements in the Ψ and Φ matrices. The main contribution in Komunjer and Ng (2011) is to account for the nonidentifiability of the reduced-form state-space parameters when formulating a rank condition along the lines of (129). In many DSGE models a subset of the state transitions are deterministic, which complicates the formal analysis.

Identification becomes generally more tenuous the fewer variables are included in the vector y_t. For instance, in the context of the stylized DSGE model, suppose y_t only includes the labor share. According to (70) the law of motion for the labor share is the sum of three AR(1) processes and an *iid* monetary policy shock. It can be rewritten as an ARMA(3,3) process and therefore has at most 8 identifiable reduced-form parameters. Thus, the upper bound on the number of reduced-form parameters is less than the number of DSGE model parameters, which is 13. In turn, it is not possible to identify the entire θ vector.

9.1.2 Global Identification

Global identification is more difficult to verify than local identification. Consider the following example from Schorfheide (2013):

$$y_t = [1 \;\; 1] s_t, \quad s_t = \begin{bmatrix} \theta_1^2 & 0 \\ 1 - \theta_1^2 - \theta_1 \theta_2 & (1 - \theta_1^2) \end{bmatrix} s_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \epsilon_t, \quad \epsilon_t \sim iidN(0,1). \tag{131}$$

Letting L denote the lag operator with the property that $Ly_t = y_{t-1}$, one can write the law of motion of y_t as an restricted ARMA(2,1) process:

$$\left(1 - \theta_1^2 L\right)\left(1 - (1 - \theta_1^2)L\right) y_t = (1 - \theta_1 \theta_2 L)\epsilon_t. \tag{132}$$

It can be verified that given θ_1 and θ_2, an observationally equivalent process can be obtained by choosing $\widetilde{\theta}_1$ and $\widetilde{\theta}_2$ such that

$$\widetilde{\theta}_1 = \sqrt{1 - \theta_1^2}, \quad \widetilde{\theta}_2 = \theta_1 \theta_2 / \widetilde{\theta}_1.$$

Here we switched the values of the two roots of the autoregressive lag polynomial. Qu and Tkachenko (2014) propose to check for global identification by searching for solutions to the equation

$$0 = \triangle_{KL}(\theta | \theta_0), \quad \theta \in \Theta. \tag{133}$$

If θ_0 is the unique solution, then the DSGE model is globally identified. The authors evaluate the Kullback–Leibler discrepancy using a frequency domain transformation. The computational challenge is to find all the roots associated with (133). Kociecki and Kolasa (2015) follow a slightly different approach that is attractive because it requires the user to solve the DSGE model only at θ_0, but not at all the other values of $\theta \in \Theta$.

9.2 Frequentist Inference

The fundamental problem of statistical inference is to infer the parameter vector θ, in our case the DSGE model parameters, based on a random sample Y. Frequentist inference adopts a preexperimental perspective and examines the sampling distribution of estimators and test statistics, which are transformations of the random sample Y, conditional on a hypothetical DGP. We will distinguish between two cases. First, we consider the stylized case in which the DSGE model is correctly specified. Formally, this means that Y is sampled from $p(Y|\theta_0)$, where the density $p(Y|\theta_0)$ is derived from the DSGE model and θ is the "true" but unknown parameter vector.[al] Second, we consider the case of misspecification, meaning the DSGE model is too stylized to capture some of the key features of the data Y. As a consequence, the sampling distribution of Y has to be characterized by a reference model, for instance, a VAR or a linear process. In terms of notation, we will distinguish between the DSGE model, denoted by M_1, and the reference model M_0. To avoid confusion about which model generates the sampling distribution of Y, we add the model indicator to the conditioning set and write, eg, $p(Y|\theta,M_1)$ or $p(Y|M_0)$. We also use the notation $\|a\|_W = a'Wa$.

9.2.1 "Correct" Specification of DSGE Model

Under the assumption of correct specification, the DSGE model itself is the DGP and $p(Y|\theta_0,M_1)$ describes the sampling distribution of Y under which the behavior of estimators and test statistics is being analyzed. In this case it is desirable to let the model–implied probability distribution $p(Y|\theta_0,M_1)$ determine the choice of the objective function for estimators and test statistics to obtain a statistical procedure that is efficient (meaning that the estimator is close to θ_0 with high probability in repeated sampling). In this regard, the maximum likelihood (ML) estimator

$$\hat{\theta}_{ml} = \text{argmax}_{\theta \in \Theta} \ \log p(Y|\theta,M_1) \tag{134}$$

plays a central role in frequentist inference, because it is efficient under fairly general regularity conditions. One of these conditions is that θ_0 is identifiable.

Alternative estimators can be obtained by constructing an objective function $Q_T(\theta|Y)$ that measures the discrepancy between sample statistics $\hat{m}_T(Y)$ (see Section 8.3) and model–implied population statistics $\mathbb{E}[\hat{m}_T(Y)|\theta,M_1]$ (see Section 8.2). Examples of the vector $\hat{m}_T(Y)$ are, for instance, vectorized sample autocovariances such as

---

[al] In reality, of course, the observed Y is never generated from a probabilistic mechanism. Instead it reflects measured macroeconomic activity. Thus, by "correct specification of a DSGE model" we mean that we believe that its probabilistic structure is rich enough to assign high probability to the salient features of macroeconomic time series.

$$\hat{m}_T(Y) = \left[vech(\hat{\Gamma}_{yy}(0))', vec(\hat{\Gamma}_{yy}(1))'\right] = \frac{1}{T}\sum_{t=1}^{T} m(y_{t-1:t})$$

or the OLS estimator of the coefficients of a VAR(1) (here without intercept)

$$\hat{m}_T(Y) = vec\left(\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}y'_{t-1}\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}y'_t\right).$$

We write the estimation objective function as

$$Q_T(\theta|Y) = \|\hat{m}_T(Y) - \mathbb{E}[\hat{m}_T(Y)|\theta, M_1]\|_{W_T}, \tag{135}$$

where W_T is a symmetric positive-definite weight matrix. Under the assumption of a correctly specified DSGE model, the optimal choice of the weight matrix W_T is the inverse of the DSGE model-implied covariance matrix of $\hat{m}_T(Y)$. Thus, more weight is assigned to sample moments that accurately approximate the underlying population moment. The minimum distance (MD) estimator of θ is defined as

$$\hat{\theta}_{md} = \text{argmax}_{\theta \in \Theta}\ Q_T(\theta|Y). \tag{136}$$

Econometric inference is based on the sampling distribution of the estimator $\hat{\theta}_{md}$ and confidence sets and test statistics derived from $\hat{\theta}_{md}$ and $Q_T(\theta|Y)$ under the distribution $p(Y|\theta_0, M_1)$.

9.2.2 Misspecification and Incompleteness of DSGE Models

Model misspecification can be interpreted as a violation of the cross-coefficient restrictions embodied in the mapping from the DSGE model parameters θ into the system matrices Ψ_0, Ψ_1, Φ_1, and Φ_ϵ of the state-space representation in (76) and (78). An example of an incomplete model is a version of the stylized DSGE model in which we do not fully specify the law of motion for the exogenous shock processes and restrict our attention to certain moment conditions, such as the consumption Euler equation. In some cases, incompleteness and misspecification are two sides of the same coin. Consider a version of the stylized DSGE model with only one structural shock, namely, the monetary policy shock. This version does not contain sufficiently many shocks to explain the observed variability in output growth, the labor share, inflation, and the interest rate. More specifically, the one-shock DSGE model implies, for instance, that the linear combination

$$\frac{1}{\kappa_p(1+\nu)x_{\epsilon_R}/\beta + \sigma_R}\hat{R}_t - \frac{1}{\kappa_p(1+\nu)x_{\epsilon_R}}\hat{\pi}_t = 0$$

is perfectly predictable; see (72) and (73). This prediction is clearly counterfactual. We could regard the model as misspecified, in the sense that its predictions are at odds with

the data; or as incomplete, in the sense that adding more structural shocks could reduce the gap between model and reality.

Regardless of whether the DSGE model is incomplete or misspecified, it does not produce a sampling distribution for the data Y that can be used to determine the frequentist behavior of estimators and test statistics. In order to conduct a frequentist analysis, we require a reference model M_0 that determines the distribution of the data $p(Y|M_0)$ and can be treated as a DGP. The reference model could be a fully specified parametric model such as a VAR, $p(Y|\phi, M_0)$, where ϕ is a finite-dimensional parameter vector. Alternatively, the reference model could be a general stochastic process for $\{y_t\}$ that satisfies a set of regularity conditions necessary to establish large sample approximations of the sampling distributions of estimators and test statistics.

If the DSGE model is incompletely specified, it is still possible to uphold the notion of a "true" parameter vector θ_0, in the sense that one could imagine the DGP to be the incompletely specified DSGE model augmented by a set of equations (potentially with additional parameters). If the DSGE model is misspecified, then the concept of a "true" parameter value has to be replaced by the notion of a pseudo-true (or pseudo-optimal) parameter value. The definition of a pseudo-true parameter value requires a notion of discrepancy between the DGP $p(Y|M_0)$ and the DSGE model $p(Y|\theta, M_1)$. Different discrepancies lead to different pseudo-optimal values. Likelihood-based inference is associated with the Kullback–Leibler discrepancy and would lead to

$$\theta_0(KL) = \mathrm{argmin}_{\theta \in \Theta} \ - \int \log\left(\frac{p(Y|\theta, M_1)}{p(Y|M_0)}\right) p(Y|M_0) dY. \tag{137}$$

Moment-based inference based on the sample objective function $Q_T(\theta|Y)$ is associated with a pseudo-optimal value

$$\theta_0(Q, W) = \mathrm{argmin}_{\theta \in \Theta} \ Q(\theta|M_0), \tag{138}$$

where

$$Q(\theta|M_0) = \left\| \mathbb{E}[\hat{m}_T(Y)|M_0] - \mathbb{E}[\hat{m}(Y)|\theta, M_1] \right\|_W.$$

Ultimately, the sampling properties of estimators and test statistics have to be derived from the reference model M_0.

9.3 Bayesian Inference

Under the Bayesian paradigm, the calculus of probability is used not only to deal with uncertainty about shocks ϵ_t, states s_t, and observations y_t, but also to deal with uncertainty about the parameter vector θ. The initial state of knowledge (or ignorance) is summarized by a prior distribution with density $p(\theta)$. This prior is combined with the conditional distribution of the data given θ, ie, the likelihood function, to characterize the joint distribution of parameters and data. Bayes Theorem is applied to obtain the conditional

distribution of the parameters given the observed data Y. This distribution is called the posterior distribution:

$$p(\theta|Y,M_1) = \frac{p(Y|\theta,M_1)p(\theta|M_1)}{p(Y|M_1)}, \quad p(Y|M_1) = \int p(Y|\theta,M_1)p(\theta|M_1)d\theta. \quad (139)$$

The posterior distribution contains all the information about θ conditional on sample information Y. In a Bayesian setting a model comprises the likelihood function $p(Y|\theta, M_1)$ *and* the prior $p(\theta|M_1)$.

The posterior distribution of transformations of the DSGE model parameters θ, say, $h(\theta)$, eg, autocovariances and impulse response functions, can be derived from $p(\theta|Y, M_1)$. For instance,

$$\mathbb{P}_Y\{h(\theta) \leq \bar{h}\} = \int_{\theta \mid h(\theta) \leq \bar{h}} p(\theta|Y,M_1)d\theta. \quad (140)$$

Solutions to inference problems can generally be obtained by specifying a suitable loss function, stating the inference problem as a decision problem, and minimizing posterior expected loss. For instance, to obtain a point estimator for $h(\theta)$, let $L(h(\theta),\delta)$ describe the loss associated with reporting δ if $h(\theta)$ is correct. The optimal decision δ_* is obtained by minimizing the posterior expected loss:

$$\delta_* = \mathrm{argmin}_{\delta \in \mathcal{D}} \int L(h(\theta),\delta)p(\theta|Y,M_1)d\theta. \quad (141)$$

If the loss function is quadratic, then the optimal point estimator is the posterior mean of $h(\theta)$.

The most difficult aspect of Bayesian inference is the characterization of the posterior moments of $h(\theta)$. Unfortunately, it is not possible to derive these moments analytically for DSGE models. Thus, researchers have to rely on numerical methods. The Bayesian literature has developed a sophisticated set of algorithms to generate draws θ^i from the posterior distribution, such that averages of these draws converge to posterior expectations:

$$\mathbb{E}[h(\theta)|Y,M_1] = \int h(\theta)p(\theta|Y,M_1)d\theta \approx \frac{1}{N}\sum_{i=1}^{N}h(\theta^i). \quad (142)$$

Several of these computational techniques are discussed in more detail in .

9.3.1 "Correct" Specification of DSGE Models
The use of Bayes Theorem to learn about the DSGE model parameters implicitly assumes that the researcher regards the probabilistic structure of the DSGE model as well specified in the sense that there are parameters θ in the support of the prior distribution conditional on which the salient features of the data Y are assigned a high probability. Of course, in practice there is always concern that an alternative DSGE model may deliver a better

description of the data. The Bayesian framework is well suited to account for model uncertainty.

Suppose the researcher contemplates two model specifications M_1 and M_2, assuming that one of them is correct. It is natural to place prior probabilities on the two models, which we denote by $\pi_{j,0}$. Ratios of model probabilities are called model odds. The posterior odds of M_1 vs M_2 are given by

$$\frac{\pi_{1,T}}{\pi_{2,T}} = \frac{\pi_{1,0}}{\pi_{2,0}}\frac{p(Y|M_1)}{p(Y|M_2)}, \tag{143}$$

where the first factor on the right-hand side captures the prior odds and the second factor, called Bayes factor, is the ratio of marginal data densities. Note that $p(Y|M_i)$ appears in the denominator of Bayes Theorem (139). Posterior model odds and probabilities have been widely used in the DSGE model literature to compare model specification or to take averages across DSGE models. Prominent applications include Rabanal and Rubio-Ramírez (2005) and Smets and Wouters (2007).

9.3.2 Misspecification of DSGE Models

As in the frequentist case, model misspecification complicates inference. Several approaches have been developed in the literature to adapt Bayesian analysis to the potential misspecification of DSGE models. In general, the model space needs to be augmented by a more densely parameterized reference model, M_0, that provides a more realistic probabilistic representation of the data.

Schorfheide (2000) considers a setting in which a researcher is interested in the relative ability of two (or more) DSGE models, say, M_1 and M_2, to explain certain population characteristics φ, eg, autocovariances or impulse responses.[am] However, the DSGE models may be potentially misspecified and the researcher considers a reference model M_0. As long as it is possible to form a posterior distribution for φ based on the reference model, the overall posterior can be described by

$$p(\varphi|Y) = \sum_{j=0,1,2} \pi_{j,T}p(\varphi|Y,M_j). \tag{144}$$

If one of the DSGE models is well specified, this model receives high posterior probability and dominates the mixture. If both DSGE models are at odds with the data, the posterior probability of the reference model will be close to one. Given a loss function over predictions of φ, one can compute DSGE model-specific predictions:

$$\hat{\varphi}_{(j)} = \operatorname{argmin}_{\tilde{\varphi}} \int L(\tilde{\varphi},\varphi)p(\varphi|Y,M_j)d\varphi, \quad j=1,2. \tag{145}$$

[am] Frequentist versions of this approach have been developed in Hnatkosvaka et al. (2012) and Marmer and Otsu (2012).

Finally, the two DSGE models can be ranked based on the posterior risk

$$\int L(\hat{\varphi}_{(j)}, \varphi) p(\varphi|Y) d\varphi. \tag{146}$$

Geweke (2010) assumes that the researcher regards the DSGE models not as models of the data Y, but as models of some population moments φ. A reference model M_0, eg, a VAR, provides the model for Y, but also permits the computation of implied population moments. He shows that under these assumptions, one can define the posterior odds of DSGE models as

$$\frac{\pi_{1,T}}{\pi_{2,T}} = \frac{\pi_{1,0}}{\pi_{2,0}} \frac{\int p(\varphi|M_1) p(\varphi|Y, M_0) d\varphi}{\int p(\varphi|M_2) p(\varphi|Y, M_0) d\varphi}. \tag{147}$$

Roughly, if we were able to observe φ, then $p(\varphi|M_j)$ is the marginal likelihood. However, φ is unobservable and therefore replaced by a posterior predictive distribution obtained from a reference model M_0. The odds in favor of model M_1 are high if there is a lot of overlap between the preditive distribution for the population moments φ under the DSGE model, and the posterior distribution of φ obtained when estimating the reference model M_0.

Building on work by Ingram and Whiteman (1994); Del Negro and Schorfheide (2004) do not treat the DSGE model as a model of the data Y, but instead use it to construct a prior distribution for a VAR. Consider the companion form VAR in (114). Use the DSGE model to generate a prior distribution for $(\widetilde{\Phi}_1, \widetilde{\Phi}_0, \widetilde{\Sigma})$ and combine this prior with the VAR likelihood function

$$p(Y, \widetilde{\Phi}_0, \widetilde{\Phi}_1, \widetilde{\Sigma}, \theta|\lambda) = p(Y|\widetilde{\Phi}_0, \widetilde{\Phi}_1, \widetilde{\Sigma}) p(\widetilde{\Phi}_0, \widetilde{\Phi}_1, \widetilde{\Sigma}|\theta, \lambda) p(\theta). \tag{148}$$

The resulting hierarchical model is called a DSGE-VAR. The prior $p(\widetilde{\Phi}_0, \widetilde{\Phi}_1, \widetilde{\Sigma}|\theta, \lambda)$ is centered on restriction functions

$$\widetilde{\Phi}_0^*(\theta), \quad \widetilde{\Phi}_1^*(\theta), \quad \widetilde{\Sigma}^*(\theta),$$

but allows for deviations from these restriction functions to account for model misspecification. The parameter λ is a hyperparameter that controls the magnitude of the deviations (prior variance) from the restriction function. This framework can be used for forecasting, to assess the fit of DSGE models, eg, Del Negro et al. (2007), and to conduct policy analysis, eg, Del Negro and Schorfheide (2009).

In a setting in which the reference model M_0 plays a dominating role, Fernández-Villaverde and Rubio-Ramírez (2004) show that choosing the DSGE model that attains the highest posterior probability (among, say, competing DSGE models M_1 and M_2) leads asymptotically to the specification that is closest to M_0 in a Kullback–Leibler sense. Rather than using posterior probabilities to select among or average across two DSGE models, one can form a prediction pool, which is essentially a linear combination of two predictive densities:

$$\lambda p(\gamma_t | Y_{1:t-1}, M_1) + (1 - \lambda) p(\gamma_t | Y_{1:t-1}, M_2).$$

The weight $\lambda \in [0,1]$ can be determined based on

$$\prod_{t=1}^{T} [\lambda p(\gamma_t | Y_{1:t-1}, M_1) + (1 - \lambda) p(\gamma_t | Y_{1:t-1}, M_2)].$$

This objective function could either be maximized with respect to λ or it can be treated as a likelihood function for λ and embedded in a Bayesian inference procedure. This idea is developed in Geweke and Amisano (2011) and Geweke and Amisano (2012). Dynamic versions with λ depending on time t are provided by Waggoner and Zha (2012) and Del Negro et al. (2014).

10. THE LIKELIHOOD FUNCTION

The likelihood function plays a central role in both frequentist and Bayesian inference. The likelihood function treats the joint density of the observables conditional on the parameters, $p(Y_{1:T} | \theta)$, as a function of θ. The state-space representation of the DSGE model leads to a joint distribution $p(Y_{1:T}, S_{1:T} | \theta)$; see (79). In order to obtain the likelihood function, one needs to integrate out the (hidden) states $S_{1:T}$. This can be done recursively, using an algorithm that is a called a *filter*.

This section focuses on the numerical evaluation of the likelihood function conditional on a particular parameterization θ through the use of linear and nonlinear filters. We assume that the DSGE model has the following, possibly nonlinear, state-space representation:

$$\begin{aligned} \gamma_t &= \Psi(s_t, t; \theta) + u_t, \quad u_t \sim F_u(\cdot; \theta) \\ s_t &= \Phi(s_{t-1}, \epsilon_t; \theta), \quad \epsilon_t \sim F_\epsilon(\cdot; \theta). \end{aligned} \tag{149}$$

The state-space system is restricted in two dimensions. First, the errors in the measurement equation enter in an additively separable manner. This implies that the conditional density $p(\gamma_t | s_t, \theta)$ is given by $p_u(\gamma_t - \Psi(s_t, t; \theta) | \theta)$, where $p_u(\cdot | \theta)$ is the pdf associated with the measurement error distribution $F_u(\cdot; \theta)$. In the absence of measurement errors, the distribution $\gamma_t | (s_t, \theta)$ is a pointmass at $\Psi(s_t, t; \theta)$. Second, the state-transition equation has a first-order Markov structure.[an] Owing to the first-order Markov structure of the state-transition equation, neither the states $s_{t-2}, s_{t-3}, \ldots$ nor the observations $\gamma_{t-1}, \gamma_{t-2}, \ldots$ provide any additional information about s_t conditional on s_{t-1}. Thus,

---

[an] Additional lags of the state vector could be easily incorporated using a companion form representation of the state vector as in (114).

$$p(s_t|s_{t-1},\theta) = p(s_t|s_{t-1}, S_{1:t-2},\theta) = p(s_t|s_{t-1}, S_{1:t-2}, Y_{1:t-1},\theta). \tag{150}$$

For the linearized DSGE model of Section 8.1 with normally distributed measurement errors $u_t \sim N(0,\Sigma_u)$ the conditional distributions are given by $s_t|(s_{t-1},\theta) \sim N(\Phi_1 s_{t-1}, \Phi_\epsilon \Phi_\epsilon')$ and $y_t|(s_t,\theta) \sim N(\Psi_0 + \Psi_1 s_t, \Sigma_u)$.

10.1 A Generic Filter

We now describe a generic filter that can be used to recursively compute the conditional distributions $p(s_t|Y_{1:t},\theta)$ and $p(y_t|Y_{1:t-1},\theta)$, starting from an initialization $p(s_0|\theta)$. The distributions $p(s_t|Y_{1:t},\theta)$ are a by-product of the algorithm and summarize the information about the state s_t conditional on the current and past observations $Y_{1:t}$, which may be of independent interest. The sequence of predictive distributions $p(y_t|Y_{1:t-1},\theta)$, $t = 1,\ldots,T$, can be used to obtain the likelihood function, which can be factorized as follows

$$p(Y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|Y_{1:t-1},\theta). \tag{151}$$

The filter is summarized in Algorithm 5. In the description of the filter we drop the parameter θ from the conditioning set to simplify the notation.

Algorithm 5 (Generic Filter). Let $p(s_0) = p(s_0|Y_{1:0})$ be the initial distribution of the state. For $t = 1$ to T:
1. Forecasting t given $t - 1$:
   (a) Transition equation:

   $$p(s_t|Y_{1:t-1}) = \int p(s_t|s_{t-1}, Y_{1:t-1})p(s_{t-1}|Y_{1:t-1})ds_{t-1}$$

   (b) Measurement equation:

   $$p(y_t|Y_{1:t-1}) = \int p(y_t|s_t, Y_{1:t-1})p(s_t|Y_{1:t-1})ds_t$$

2. Updating with Bayes Theorem. Once y_t becomes available:

   $$p(s_t|Y_{1:t}) = p(s_t|y_t, Y_{1:t-1}) = \frac{p(y_t|s_t, Y_{1:t-1})p(s_t|Y_{1:t-1})}{p(y_t|Y_{1:t-1})}.$$

10.2 Likelihood Function for a Linearized DSGE Model

For illustrative purposes, consider the prototypical DSGE model. Owing to the simple structure of the model, we can use (69), (70), (72), and (73) to solve for the latent shocks ϕ_t, λ_t, z_t, and $\epsilon_{R,t}$ as a function of $\hat{x}_t$, $l\hat{s}h_t$, $\hat{\pi}_t$, and $\hat{R}_t$. Thus, we can deduce from (78) and the definition of s_t that conditional on $\hat{x}_0$, the states s_t can be uniquely inferred from the observables y_t in a recursive manner, meaning that the conditional distributions

$p(s_t|Y_{1:t}, \hat{x}_0)$ are degenerate. Thus, the only uncertainty about the state stems from the initial condition.

Suppose that we drop the labor share and the interest rates from the definition of γ_t. In this case it is no longer possible to uniquely determine s_t as a function of γ_t and $\hat{x}_0$, because we only have two equations, (69) and (72), and four unknowns. The filter in Algorithm 5 now essentially solves an underdetermined system of equations, taking into account the probability distribution of the four hidden processes. For our linearized DSGE model with Gaussian innovations, all the distributions that appear in Algorithm 5 are Gaussian. In this case the Kalman filter can be used to compute the means and covariance matrices of these distributions recursively. To complete the model specification, we make the following distributional assumptions about the initial state s_0:

$$s_0 \sim N\left(\bar{s}_{0|0}, P_{0|0}\right).$$

In stationary models it is common to set $\bar{s}_{0|0}$ and $P_{0|0}$ equal to the unconditional first and second moments of the invariant distribution associated with the law of motion of s_t in (76). The four conditional distributions in the description of Algorithm 5 for a linear Gaussian state-space model are summarized in Table 6. Detailed derivations can be found in textbook treatments of the Kalman filter and smoother, eg, Hamilton (1994) or Durbin and Koopman (2001).

To illustrate the Kalman filter algorithm, we simulate $T = 50$ observations from the stylized DSGE model conditional on the parameters in Table 5. The two left panels of Fig. 24 depict the filtered shock processes ϕ_t and z_t based on observations of only output growth, which are defined as $\mathbb{E}[s_t|Y_{1:t}]$. The bands delimit 90% credible intervals which are centered around the filtered estimates and based on the standard deviations $\sqrt{\mathbb{V}[s_t|Y_{1:t}]}$. The information in the output growth series is not sufficient to generate a precise estimate of the preference shock process ϕ_t, which, according to the forecast

Table 6 Conditional distributions for the Kalman filter

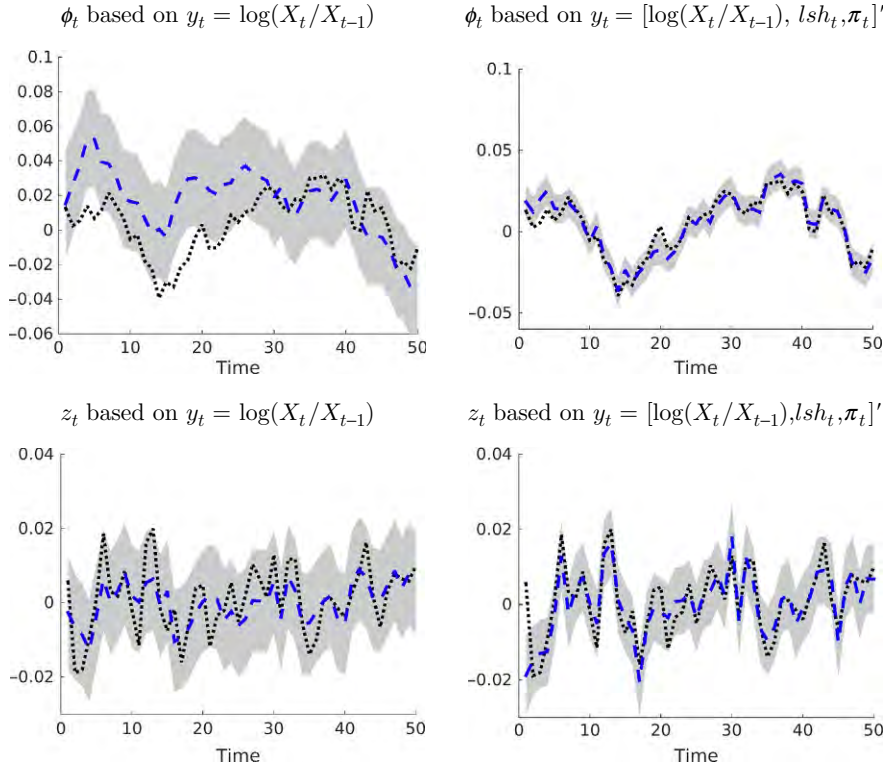|	Distribution	Mean and variance
$s_{t-1}|Y_{1:t-1}$	$N\left(\bar{s}_{t-1|t-1}, P_{t-1|t-1}\right)$	Given from Iteration $t-1$
$s_t|Y_{1:t-1}$	$N\left(\bar{s}_{t|t-1}, P_{t|t-1}\right)$	$\bar{s}_{t|t-1} = \mathbf{\Phi}_1 \bar{s}_{t-1|t-1}$
		$P_{t|t-1} = \mathbf{\Phi}_1 P_{t-1|t-1} \mathbf{\Phi}_1' + \mathbf{\Phi}_\epsilon \Sigma_\epsilon \mathbf{\Phi}_\epsilon'$
$\gamma_t|Y_{1:t-1}$	$N\left(\bar{y}_{t|t-1}, F_{t|t-1}\right)$	$\bar{y}_{t|t-1} = \mathbf{\Psi}_0 + \mathbf{\Psi}_1 \bar{s}_{t|t-1}$
		$F_{t|t-1} = \mathbf{\Psi}_1 P_{t|t-1} \mathbf{\Psi}_1' + \Sigma_u$
$s_t|Y_{1:t}$	$N\left(\bar{s}_{t|t}, P_{t|t}\right)$	$\bar{s}_{t|t} = \bar{s}_{t|t-1} + P_{t|t-1} \mathbf{\Psi}_1' F_{t|t-1}^{-1}\left(\gamma_t - \bar{y}_{t|t-1}\right)$
		$P_{t|t} = P_{t|t-1} - P_{t|t-1} \mathbf{\Psi}_1' F_{t|t-1}^{-1} \mathbf{\Psi}_1 P_{t|t-1}$
$s_t|(S_{t+1:T}, Y_{1:T})$	$N\left(\bar{s}_{t|t+1}, P_{t|t+1}\right)$	$\bar{s}_{t|t+1} = \bar{s}_{t|t} + P_{t|t} \mathbf{\Phi}_1' P_{t+1|t}^{-1}\left(s_{t+1} - \mathbf{\Phi}_1 \bar{s}_{t|t}\right)$
		$P_{t|t+1} = P_{t|t} - P_{t|t} \mathbf{\Phi}_1' P_{t+1|t}^{-1} \mathbf{\Phi}_1 P_{t|t}$

Fig. 24 Filtered states. *Notes*: The filtered states are based on a simulated sample of $T = 50$ observations. Each panel shows the true state s_t (*dotted*), the filtered state $\mathbb{E}[s_t|Y_{1:t}]$ (*dashed*), and 90% credible bands based on $p(s_t|Y_{1:t})$ (*grey area*).

error variance decomposition in Fig. 17, only explains a small fraction of the variation in output growth. The two right panels of Fig. 24 show what happens to the inference about the hidden states if inflation and labor share are added to the set of observables. Conditional on the three series, it is possible to obtain fairly sharp estimates of both the preference shock ϕ_t and the technology growth shock z_t.

Instead of using the Kalman filter, in a linearized DSGE model with Gaussian innovations it is possible to characterize the joint distribution of the observables directly. Let Y be a $T \times n_y$ matrix composed of rows y_t'. Then the joint distribution of Y is given by

$$
vec(Y)|\theta \sim N\left(I \otimes \Phi_0(\theta), \begin{bmatrix} \Gamma_{yy}(0|\theta) & \Gamma_{yy}(1|\theta) & \dots & \Gamma_{yy}(T-1|\theta) \\ \Gamma_{yy}'(1|\theta) & \Gamma_{yy}(0|\theta) & \dots & \Gamma_{yy}(T-2|\theta) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{yy}'(T-1|\theta) & \Gamma_{yy}'(T-2|\theta) & \dots & \Gamma_{yy}(0|\theta) \end{bmatrix} \right).
$$

$$(152)$$

The evaluation of the likelihood function requires the calculation of the autocovariance sequence and the inversion of an $n_y T \times n_y T$ matrix. For large T the joint density can be approximated by the so-called Whittle likelihood function

$$p_W(Y|\theta) \propto \left(\prod_{j=0}^{T-1} \left| 2\pi f_{\gamma\gamma}^{-1}(\omega_j|\theta) \right| \right)^{1/2} \exp\left\{ -\frac{1}{2} \sum_{j=0}^{T-1} tr\left[f_{\gamma\gamma}^{-1}(\omega_j|\theta) \hat{f}_{\gamma\gamma}(\omega_j) \right] \right\} \qquad (153)$$

where $f_{\gamma\gamma}(\omega|\theta)$ is the DSGE model-implied spectral density, $\hat{f}_{\gamma\gamma}(\omega)$ is the sample periodogram, and the ω_j's are the fundamental frequencies. The attractive feature of this likelihood function is that the researcher can introduce weights for the different frequencies, and, for instance, only consider business cycle frequencies in the construction of the likelihood function. For the estimation of DSGE models, the Whittle likelihood has been used, for instance, by Christiano and Vigfusson (2003), Qu and Tkachenko (2012), and Sala (2015).

10.3 Likelihood Function for Nonlinear DSGE Models

If the DSGE model is solved using a nonlinear approximation technique, then either the state-transition equation, or the measurement equation, or both become nonlinear. As a consequence, analytical representations of the densities $p(s_{t-1}|Y_{1:t-1})$, $p(s_t|Y_{1:t-1})$, and $p(y_t|Y_{1:t-1})$ that appear in Algorithm 5 are no longer available. While there exists a large literature on nonlinear filtering (see for instance Crisan and Rozovsky, 2011) we focus on the class of particle filters. Particle filters belong to the class of sequential Monte Carlo algorithms. The basic idea is to approximate the distribution $s_t|Y_{1:t}$ through a swarm of particles $\{s_t^j, W_t^j\}_{j=1}^M$ such that

$$\bar{h}_{t,M} = \frac{1}{M} \sum_{j=1}^M h(s_t^j) W_t^j \xrightarrow{a.s.} \mathbb{E}[h(s_t)|Y_{1:t}],$$

$$\sqrt{M}(\bar{h}_{t,M} - \mathbb{E}[h(s_t)|Y_{1:t}]) \implies N(0, \Omega_t[h]), \qquad (154)$$

where $\implies$ denotes convergence in distribution.[ao] Here the s_t^j's are particle values and the W_t^j's are the particle weights. The conditional expectation of $h(s_t)$ is approximated by a weighted average of the (transformed) particles $h(s_t^j)$. Under suitable regularity conditions, the Monte Carlo approximation satisfies an SLLN and a CLT. The covariance matrix $\Omega_t[h]$ characterizes the accuracy of the Monte Carlo approximation. Setting $h(s_t) = p(y_{t+1}|s_t)$ yields the particle filter approximation of the likelihood increment $p(y_{t+1}|Y_{1:t}) = \mathbb{E}[p(y_{t+1}|s_t)|Y_{1:t}]$. Each iteration of the filter manipulates the particle values and weights to recursively track the sequence of conditional distributions $s_t|Y_{1:t}$. The paper by Fernández-Villaverde and Rubio-Ramírez (2007) was the first to

---

[ao] A sequence of random variables X_T converges in distribution to a random variable X if for every measurable and bounded function $f(\cdot)$ that is continuous almost everywhere $\mathbb{E}[f(X_T)] \to \mathbb{E}[f(X)]$.

approximate the likelihood function of a nonlinear DSGE model using a particle filter and many authors have followed this approach.

Particle filters are widely used in engineering and statistics. Surveys and tutorials are provided, for instance, in Arulampalam et al. (2002), Cappé et al. (2007), Doucet and Johansen (2011), and Creal (2012). The basic bootstrap particle filter algorithm is remarkably straightforward, but may perform quite poorly in practice. Thus, much of the literature focuses on refinements of the bootstrap filter that increases the efficiency of the algorithm; see, for instance, Doucet et al. (2001). Textbook treatments of the statistical theory underlying particle filters can be found in Liu (2001), Cappé et al. (2005), and Del Moral (2013).

10.3.1 Generic Particle Filter

The subsequent exposition draws from Herbst and Schorfheide (2015), who provide a detailed presentation of particle filtering techniques in the context of DSGE model applications as well as a more extensive literature survey. In the basic version of the particle filter, the time t particles are generated based on the time $t-1$ particles by simulating the state-transition equation forward. The particle weights are then updated based on the likelihood of the observation y_t under the s_t^j particle, $p(y_t|s_t^j)$. The more accurate the prediction of y_t based on s_t^j, the larger the density $p(y_t|s_t^j)$, and the larger the relative weight that will be placed on particle j. However, the naive forward simulation ignores information contained in the current observation y_t and may lead to a very uneven distribution of particle weights, in particular, if the measurement error variance is small or if the model has difficulties explaining the period t observation in the sense that for most particles s_t^j the actual observation y_t lies far in the tails of the model-implied distribution of $y_t|s_t^j$. The particle filter can be generalized by allowing s_t^j in the forecasting step to be drawn from a generic importance sampling density $g_t(\cdot|s_{t-1}^j)$, which leads to the following algorithm:[ap]

Algorithm 6 (Generic Particle Filter).
1. **Initialization.** Draw the initial particles from the distribution $s_0^j \overset{iid}{\sim} p(s_0)$ and set $W_0^j = 1$, $j = 1,\dots,M$.
2. **Recursion.** For $t = 1,\dots,T$:
   (a) **Forecasting** s_t. Draw $\tilde{s}_t^j$ from density $g_t(\tilde{s}_t^j|s_{t-1}^j)$ and define the importance weights

$$\omega_t^j = \frac{p(\tilde{s}_t^j|s_{t-1}^j)}{g_t(\tilde{s}_t^j|s_{t-1}^j)}. \tag{155}$$

An approximation of $\mathrm{E}[h(s_t)|Y_{1:t-1}]$ is given by

---

[ap] To simplify the notation, we omit θ from the conditioning set.

$$\hat{h}_{t,M} = \frac{1}{M}\sum_{j=1}^{M} h(\tilde{s}_t^j)\omega_t^j W_{t-1}^j.$$
(156)

(b) Forecasting γ_t. Define the incremental weights

$$\tilde{w}_t^j = p(\gamma_t|\tilde{s}_t^j)\omega_t^j.$$
(157)

The predictive density $p(\gamma_t|Y_{1:t-1})$ can be approximated by

$$\hat{p}(\gamma_t|Y_{1:t-1}) = \frac{1}{M}\sum_{j=1}^{M} \tilde{w}_t^j W_{t-1}^j.$$
(158)

(c) Updating. Define the normalized weights

$$\tilde{W}_t^j = \frac{\tilde{w}_t^j W_{t-1}^j}{\frac{1}{M}\sum_{j=1}^{M}\tilde{w}_t^j W_{t-1}^j}.$$
(159)

An approximation of $\mathbb{E}[h(s_t)|Y_{1:t},\theta]$ is given by

$$\tilde{h}_{t,M} = \frac{1}{M}\sum_{j=1}^{M} h(\tilde{s}_t^j)\,\tilde{W}_t^j.$$
(160)

(d) Selection. Resample the particles via multinomial resampling. Let $\{s_t^j\}_{j=1}^{M}$ denote M *iid* draws from a multinomial distribution characterized by support points and weights $\{\tilde{s}_t^j, \tilde{W}_t^j\}$ and set $W_t^j = 1$ for $j =,1\ldots, M$. An approximation of $\mathbb{E}[h(s_t)|Y_{1:t},\theta]$ is given by

$$\bar{h}_{t,M} = \frac{1}{M}\sum_{j=1}^{M} h(s_t^j) W_t^j.$$
(161)

3. **Likelihood Approximation.** The approximation of the log likelihood function is given by

$$\log\hat{p}(Y_{1:T}|\theta) = \sum_{t=1}^{T}\log\left(\frac{1}{M}\sum_{j=1}^{M}\tilde{w}_t^j W_{t-1}^j\right).$$
(162)

Conditional on the stage $t-1$ weights W_{t-1}^j the accuracy of the approximation of the likelihood increment $p(\gamma_t|Y_{1:t-1})$ depends on the variability of the incremental weights $\tilde{\omega}_t^j$ in (157). The larger the variance of the incremental weights, the less accurate the particle filter approximation of the likelihood function. In this regard, the most important choice for the implementation of the particle filter is the choice of the proposal distribution $g_t(\tilde{s}_t^j|s_{t-1}^j)$, which is discussed in more detail below.

The selection step is included in the filter to avoid a degeneracy of particle weights. While it adds additional noise to the Monte Carlo approximation, it simultaneously equalizes the particle weights, which increases the accuracy of subsequent approximations. In the absence of the selection step, the distribution of particle weights would become more uneven from iteration to iteration. The selection step does not have to be executed in every iteration. For instance, in practice, users often apply a threshold rule according to which the selection step is executed whenever the following measure falls below a threshold, eg, 25% or 50% of the nominal number of particles:

$$\widehat{ESS}_t = M / \left(\frac{1}{M} \sum_{j=1}^{M} (\tilde{W}_t^j)^2 \right). \tag{163}$$

The effective sample size $\widehat{ESS}_t$ (in terms of number of particles) captures the variance of the particle weights. It is equal to M if $\tilde{W}_t^j = 1$ for all j and equal to 1 if one of the particles has weight M and all others have weight 0. The resampling can be executed with a variety of algorithms. We mention multinomial resampling in the description of Algorithm 6. Multinomial resampling is easy to implement and satisfies a CLT. However, there are more efficient algorithms (meaning they are associated with a smaller Monte Carlo variance), such as stratified or systematic resampling. A detailed textbook treatment can be found in Liu (2001) and Cappé et al. (2005).

10.3.2 Bootstrap Particle Filter

The bootstrap particle filter draws $\tilde{s}_t^j$ from the state-transition equation and sets

$$g_t(\tilde{s}_t^j | s_{t-1}^j) = p(\tilde{s}_t^j | s_{t-1}^j). \tag{164}$$

This implies that $\omega_t^j = 1$ and the incremental weight is given by the likelihood $p(y_t | \tilde{s}_t^j)$, which unfortunately may be highly variable. Fig. 25 provides an illustration of the bootstrap particle filter with $M = 100$ particles using the same experimental design as for the Kalman filter in Section 10.2. The observables are output growth, labor share, and inflation and the observation equation is augmented with measurement errors. The measurement error variance amounts to 10% of the total variance of the simulated data. Because the stylized DSGE is loglinearized, the Kalman filter provides exact inference and any discrepancy between the Kalman and particle filter output reflects the approximation error of the particle filter. In this application the particle filter approximations are quite accurate even with a small number of particles. The particle filtered states z_t and $\epsilon_{R,t}$ appear to be more volatile than the exactly filtered states from the Kalman filter.

Fig. 26 illustrates the accuracy of the likelihood approximation. The left panel compares log–likelihood increments $\log p(y_t | Y_{1:t-1}, \theta)$ obtained from the Kalman filter and a single run of the particle filter. The left panel shows the distribution of the approximation errors of the log-likelihood function: $\log \hat{p}(Y_{1:T} | \theta) - \log p(Y_{1:T} | \theta)$. It has been shown,
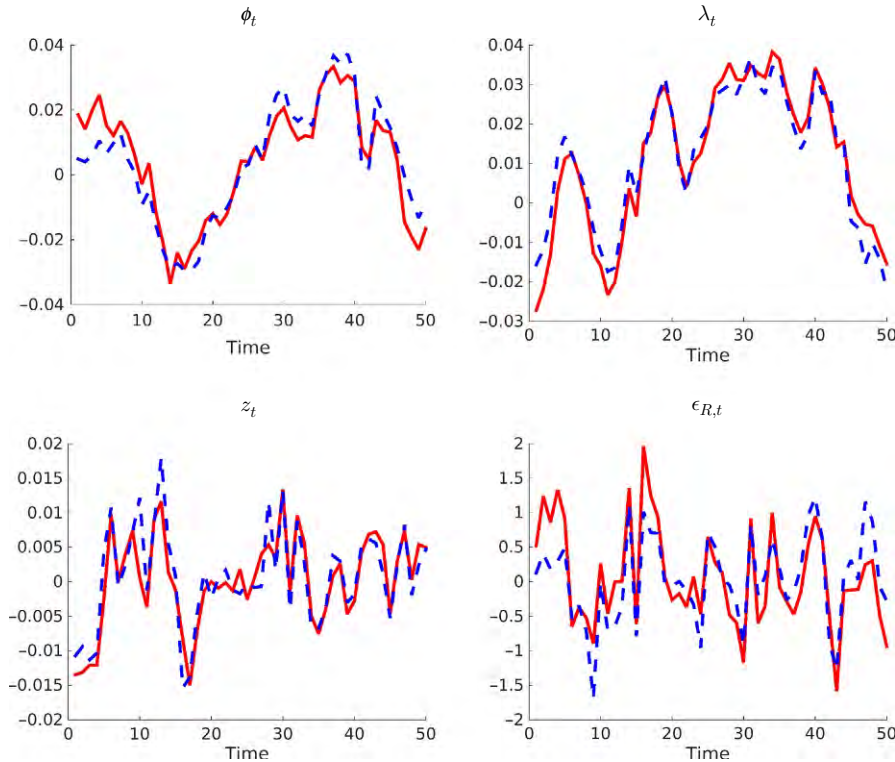
Fig. 25 Particle-filtered states. *Notes*: We simulate a sample of $T = 50$ observations y_t and states s_t from the stylized DSGE model. The four panels compare filtered states from the Kalman filter (*solid*) and a single run of the particle filter (*dashed*) with $M = 100$ particles. The observables used for filtering are output growth, labor share, and inflation. The measurement error variances are 10% of the total variance of the data.

eg, by Del Moral (2004) and Pitt et al. (2012), that the particle filter approximation of the likelihood function is unbiased, which implies that the approximation of the *log*-likelihood function has a downward bias, which is evident in the figure. Under suitable regularity conditions the particle filter approximations satisfy a CLT. The figure clearly indicates that the distribution of the approximation errors becomes more concentrated as the number of particles is increased from $M = 100$ to $M = 500$.

The accuracy of the bootstrap particle filter crucially depends on the quality of the fit of the DSGE model and the magnitude of the variance of the measurement errors u_t. Recall that for the bootstrap particle filter, the incremental weights $\widetilde{w}_t^j = p(y_t|\widetilde{s}_t^j)$. If the model fits poorly, then the one-step-ahead predictions conditional on the particles $\widetilde{s}_t^j$ are inaccurate and the density of the actual observation y_t falls far in the tails of the predictive distribution. Because the density tends to decay quickly in the tails, the
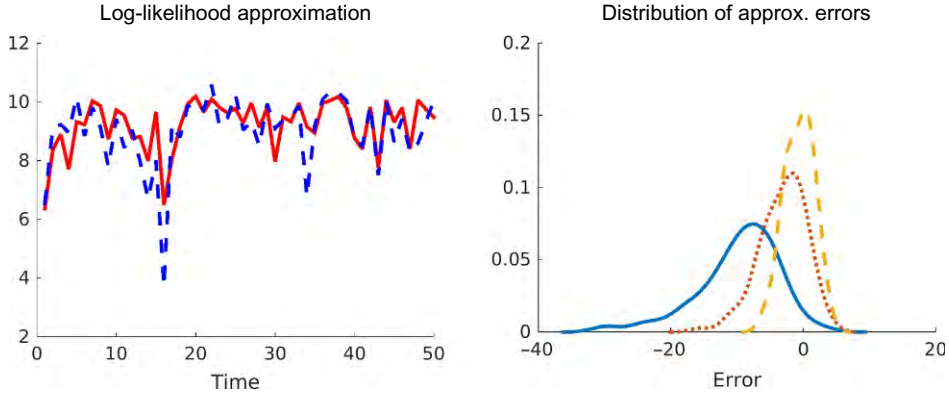
Fig. 26 Particle-filtered log-likelihood. *Notes*: We simulate a sample of $T = 50$ observations y_t and states s_t from the stylized DSGE model. The *left panel* compares log-likelihood increments from the Kalman filter (*solid*) and a single run of the particle filter (*dashed*) with $M = 100$ particles. The *right panel* shows a density plot for approximation errors of $\log \hat{p}(Y_{1:T}|\theta) - \log p(Y_{1:T}|\theta)$ based on $N_{run} = 100$ repetitions of the particle filter for $M = 100$ (*solid*), $M = 200$ (*dotted*), and $M = 500$ (*dashed*) particles. The measurement error variances are 10% of the total variance of the data.

incremental weights will have a high variability, which means that Monte Carlo approximations based on these incremental weights will be inaccurate.

The measurement error defines a metric between the observation y_t and the conditional mean prediction $\Psi(s_t,t;\theta)$. Consider the extreme case in which the measurement error is set to zero. This means that any particle that does not predict y_t exactly would get weight zero. In a model in which the error distribution is continuous, the probability of drawing a $\tilde{s}_t^j$ that receives a nonzero weight is zero, which means that the algorithm would fail in the first iteration. By continuity, the smaller the measurement error variance, the smaller the number of particles that would receive a nontrivial weight, and the larger the variance of the approximation error of particle filter approximations. In practice, it is often useful to start the filtering with a rather large measurement error variance, eg, 10% or 20% of the variance of the observables, and then observing the accuracy of the filter as the measurement error variance is reduced.

10.3.3 (Approximately) Conditionally Optimal Particle Filter
The conditionally optimal particle filter sets

$$g_t(\tilde{s}_t|s_{t-1}^j) = p(\tilde{s}_t|y_t, s_{t-1}^j), \tag{165}$$

that is, $\tilde{s}_t$ is sampled from the posterior distribution of the period t state given (y_t, s_{t-1}^j). In this case

$$\tilde{w}_t^j = \int p(y_t|s_t)p(s_t|s_{t-1}^j)ds_t = p(y_t|s_{t-1}^j). \tag{166}$$

Unfortunately, in a typical nonlinear DSGE model applications it is not possible to sample directly from $p(\tilde{s}_t | y_t, s_{t-1}^j)$. In this case the researcher could try to approximate the conditionally optimal proposal density, which leads to an *approximately conditionally optimal particle filter.* For instance, if the DSGE model's nonlinearity arises from a higher-order perturbation solution and the nonlinearities are not too strong, then an approximately conditionally optimal importance distribution could be obtained by applying the one-step Kalman filter updating described in Table 6 to the first-order approximation of the DSGE model. More generally, as suggested in Guo et al. (2005), one could use the updating steps of a conventional nonlinear filter, such as an extended Kalman filter, unscented Kalman filter, or a Gaussian quadrature filter, to construct an efficient proposal distribution. Approximate filters for nonlinear DSGE models have been developed by Andreasen (2013) and Kollmann (2015).

Whenever one uses a proposal distribution that differs from $p(\tilde{s}_t^j | s_{t-1}^j)$ it becomes necessary to evaluate the density $p(\tilde{s}_t^j | s_{t-1}^j)$. In DSGE model applications, one typically does not have a closed-form representation for this density. It is implicitly determined by the distribution of ϵ_t and the state transition $\Phi(s_{t-1}, \epsilon_t)$. The problem of having to evaluate the DSGE model–implied density of $\tilde{s}_t^j$ can be avoided by sampling an innovation from a proposal density $g^\epsilon(\tilde{\epsilon}_t | s_{t-1}^j)$ and defining $\tilde{s}_t^j = \Phi(s_{t-1}^j, \tilde{\epsilon}_t)$. In this case the particle weights can be updated by the density ratio

$$\omega_t^j = \frac{p^\epsilon(\tilde{\epsilon}_t^j)}{g_t(\tilde{\epsilon}_t^j | s_{t-1}^j)}, \tag{167}$$

where $p^\epsilon(\cdot)$ is the model-implied pdf of the innovation ϵ_t.

Sometimes, DSGE models have a specific structure that may simplify the particle-filter-based likelihood approximation. In models that are linear conditional on a subset of state variables, eg, volatility states or Markov-switching regimes, it is possible to use the Kalman filter to represent the uncertainty about a subset of states. In models in which the number of shocks ϵ_t equals the number of observables y_t, it might be possible (in the absence of measurement errors) conditional on an initial state vector s_0 to directly solve for ϵ_t based on y_t and s_{t-1}, which means that it may be possible to evaluate the likelihood function $p(Y_{1:T} | \theta, s_0)$ recursively. A more detailed discussion of these and other issues related to particle filtering for DSGE models is provided in Herbst and Schorfheide (2015).

11. FREQUENTIST ESTIMATION TECHNIQUES

We will now consider four frequentist inference techniques in more detail: likelihood-based estimation (Section 11.1), simulated method of moments estimation (Section 11.2), impulse response function matching (Section 11.3), and GMM estimation (Section 11.4). All of these econometric techniques, with the exception of the impulse

response function matching approach, are widely used in other areas of economics and are associated with extensive literatures that we will not do justice to in this section. We will sketch the main idea behind each of the econometric procedures and then focus on adjustments that have been proposed to tailor the techniques to DSGE model applications. Each estimation method is associated with a model evaluation procedure that essentially assesses the extent to which the estimation objective has been achieved.

11.1 Likelihood-Based Estimation

Under the assumption that the econometric model is well specified, likelihood-based inference techniques enjoy many optimality properties. Because DSGE models deliver a joint distribution for the observables, maximum likelihood estimation of θ is very appealing. The maximum likelihood estimator $\hat{\theta}_{ml}$ was defined in (134). Altug (1989) and McGrattan (1994) are early examples of papers that estimated variants of a neoclassical stochastic growth model by maximum likelihood, whereas Leeper and Sims (1995) estimated a DSGE model meant to be usable for monetary policy analysis.

Even in a loglinearized DSGE model, the DSGE model parameters θ enter the coefficients of the state-space representation in a nonlinear manner, which can be seen in Table 4. Thus, a numerical technique is required to maximize the likelihood function. A textbook treatment of numerical optimization routines can be found, for instance, in Judd (1998) and Nocedal and Wright (2006). Some algorithms, eg, Quasi-Newton methods, rely on the evaluation of the gradient of the objective function (which requires differentiability), and other methods, such as simulated annealing, do not. This distinction is important if the likelihood function is evaluated with a particle filter. Without further adjustments, particle filter approximations of the likelihood function are nondifferentiable in θ even if the exact likelihood function is. This issue and possible solutions are discussed, for instance, in Malik and Pitt (2011) and Kantas et al. (2014).

11.1.1 Textbook Analysis of the ML Estimator

Under the assumption that θ is well identified and the log-likelihood function is sufficiently smooth with respect to θ, confidence intervals and test statistics for the DSGE model parameters can be based on a large sample approximation of the sampling distribution of the ML estimator. A formal analysis in the context of state-space models is provided, for instance, in the textbook by Cappé et al. (2005). We sketch the main steps of the approximation, assuming that the DSGE model is correctly specified and the data are generated by $p(Y|\theta_0, M_1)$. Of course, this analysis could be generalized to a setting in which the DSGE model is misspecified and the data are generated by a reference model $p(Y|M_0)$. In this case the resulting estimator is called quasi-maximum-likelihood estimator and the formula for the asymptotic covariance matrix presented below would have to be adjusted. A detailed treatment of quasi-likelihood inference is provided in White (1994).

Recall from Section 10 that the log-likelihood function can be decomposed as follows:

$$\ell_T(\theta|Y) = \sum_{t=1}^{T} \log p(y_t|Y_{1:t-1},\theta) = \sum_{t=1}^{T} \log \int p(y_t|s_t,\theta)p(s_t|Y_{1:t-1})ds_t. \qquad (168)$$

Owing to the time-dependent conditioning information $Y_{1:t-1}$ the summands are not stationary. However, under the assumption that the sequence $\{s_t, y_t\}$ is stationary if initialized in the infinite past, one can approximate the log-likelihood function by

$$\ell_T^s(\theta|Y) = \sum_{t=1}^{T} \log \int p(y_t|s_t,\theta)p(s_t|Y_{-\infty:t-1})ds_t, \qquad (169)$$

and show that the discrepancy $\left|\ell_T(\theta|Y) - \ell_T^s(\theta|Y)\right|$ becomes negligible as $T \to \infty$. The ML estimator is consistent if $T^{-1}\ell_T^s(\theta|Y) \xrightarrow{a.s.} \ell^s(\theta)$ uniformly almost surely (a.s.), where $\ell^s(\theta)$ is deterministic and maximized at the "true" θ_0. The consistency can be stated as

$$\hat{\theta}_{ml} \xrightarrow{a.s.} \theta_0. \qquad (170)$$

Frequentist asymptotics rely on a second-order approximation of the log-likelihood function. Define the score (vector of first derivatives) $\nabla_\theta \ell_T^s(\theta|Y)$ and the matrix of second derivatives (Hessian, multiplied by minus one) $-\nabla_\theta^2 \ell_T^s(\theta|Y)$ and let

$$\ell_T^s(\theta|Y) = \ell_T^s(\theta_0|Y) + T^{-1/2}\nabla_\theta \ell_T^s(\theta_0|Y)\sqrt{T}(\theta - \theta_0)$$
$$+ \frac{1}{2}\sqrt{T}(\theta - \theta_0)'\left[\nabla_\theta^2 \ell_T^s(\theta_0|Y)\right]\sqrt{T}(\theta - \theta_0) + \text{small}.$$

If the maximum is attained in the interior of the parameter space Θ, the first-order conditions can be approximated by

$$\sqrt{T}(\hat{\theta}_{ml} - \theta_0) = \left[-\nabla_\theta^2 \ell_T^s(\theta_0|Y)\right]^{-1} T^{-1/2}\nabla_\theta \ell_T^s(\theta_0|Y) + \text{small}. \qquad (171)$$

Under suitable regularity conditions, the score process satisfies a CLT:

$$T^{-1/2}\nabla_\theta \ell_T(\theta|Y) \Longrightarrow N(0, \mathcal{I}(\theta_0)), \qquad (172)$$

where $\mathcal{I}(\theta_0)$ is the Fisher information matrix.[aq] As long as the likelihood function is correctly specified, the term $\|-\nabla_\theta^2 \ell_T(\theta|Y) - \mathcal{I}(\theta_0)\|$ converges to zero uniformly in a neighborhood around θ_0, which is a manifestation of the so-called information matrix equality. This leads to the following result

$$\sqrt{T}(\hat{\theta}_{ml} - \theta_0) \Longrightarrow N(0, \mathcal{I}^{-1}(\theta_0)). \qquad (173)$$

[aq] The formal definition of the information matrix for this model is delicate and therefore omitted.

Thus, standard error estimates for t-tests and confidence intervals for elements of the parameter vector θ can be obtained from the diagonal elements of the inverse Hessian $\left[-\nabla_\theta^2 \ell_T(\theta|Y)\right]^{-1}$ of the log-likelihood function evaluated at the ML estimator.[ar] Moreover, the maximized likelihood function can be used to construct textbook Wald, Lagrange-multiplier, and likelihood ratio statistics. Model selection could be based on a penalized likelihood function such as the Schwarz (1978) information criterion.

11.1.2 Illustration

To illustrate the behavior of the ML estimator we repeatedly generate data from the stylized DSGE model, treating the values listed in Table 5 as "true" parameters. We fix all parameters except for the Calvo parameter ζ_p at their "true" values and use the ML approach to estimate ζ_p. The likelihood function is based on output growth, labor share, inflation, and interest rate data. The left panel of Fig. 27 depicts the likelihood function for a single simulated data set Y. The right panel shows the sampling distribution of $\hat{\zeta}_{p,ml}$, which is approximated by repeatedly generating data and evaluating the ML estimator. The sampling distribution peaks near the "true" parameter value and becomes more concentrated as the sample size is increased from $T = 80$ to $T = 200$.

In practice, the ML estimator is rarely as well behaved as in this illustration, because the maximization is carried out over a high-dimensional parameter space and the log-likelihood function may be highly nonelliptical. In the remainder of this subsection, we focus on two obstacles that arise in the context of the ML estimation of DSGE models.
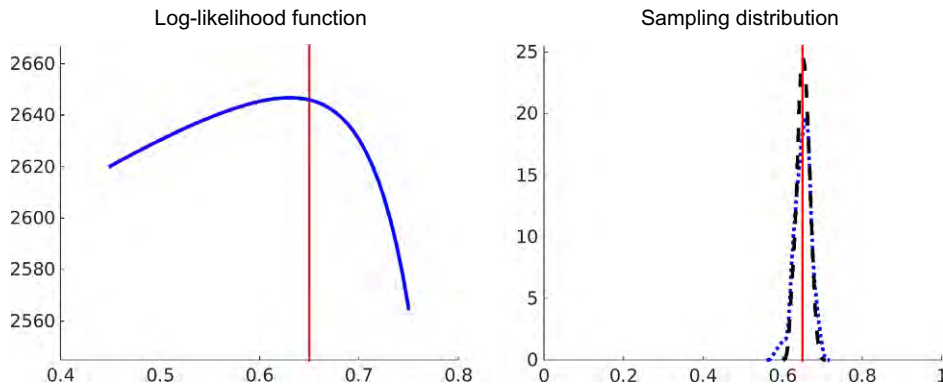

Fig. 27 Log-likelihood function and sampling distribution of $\hat{\zeta}_{p,ml}$. *Notes*: *Left panel*: log-likelihood function $\ell_T(\zeta_p|Y)$ for a single data set of size $T = 200$. *Right panel*: We simulate samples of size $T = 80$ (*dotted*) and $T = 200$ (*dashed*) and compute the ML estimator for the Calvo parameter ζ_p. All other parameters are fixed at their "true" value. The plot depicts densities of the sampling distribution of $\hat{\zeta}_p$. The vertical lines in the two panels indicate the "true" value of ζ_p.

[ar] Owing to the Information Matrix Equality, the standard error estimates can also be obtained from the outer product of the score: $\sum_{t=1}^{T} (\nabla_\theta \log p(y_t|Y_{1:t-1},\theta))(\nabla_\theta \log p(y_t|Y_{1:t-1},\theta))'$.

The first obstacle is the potential stochastic singularity of the DSGE model–implied conditional distribution of y_t given its past. The second obstacle is caused by a potential lack of identification of the DSGE model parameters.

11.1.3 Stochastic Singularity

Imagine removing all shocks except for the technology shock from the stylized DSGE model, while maintaining that y_t comprises output growth, the labor share, inflation, and the interest rate. In this case, we have one exogenous shock and four observables, which implies, among other things, that the DSGE model places probability one on the event that

$$\beta \log R_t - \log \pi_t = \beta \log (\pi^* \gamma / \beta) - \log \pi^*.$$

Because in the actual data $\beta \log R_t - \log \pi_t$ is time varying, the likelihood function is equal to zero and not usable for inference. The literature has adopted two types of approaches to address the singularity, which we refer to as the "measurement error" approach and the "more structural shocks" approach.

Under the measurement error approach (78) is augmented by a measurement error process u_t, which in general may be serially correlated. The term "measurement error" is a bit of a misnomer. It tries to blame the discrepancy between the model and the data on the accuracy of the latter rather than the quality of the former. In a typical DSGE model application, the blame should probably be shared by both. A key feature of the "measurement error" approach is that the agents in the model do not account for the presence of u_t when making their decisions. The "measurement error" approach has been particularly popular in the real business cycle literature—it was used, for instance, in Altug (1989). The real business cycle literature tried to explain business cycle fluctuations based on a small number of structural shocks, in particular, technology shocks.

The "more structural shocks" approach augments the DSGE model with additional structural shocks until the number of shocks is equal to or exceeds the desired number of observables stacked in the vector y_t. For instance, if we add the three remaining shock processes ϕ_t, λ_t, $\epsilon_{R,t}$ back into the prototypical DSGE model, then a stochastic singularity is no longer an obstacle for the evaluation of the likelihood function. Of course, at a deeper level, the stochastic singularity problem never vanishes, as we could also increase the dimension of the vector y_t. Because the policy functions in the solution of the DSGE model express the control variables as functions of the state variables, the set of potential observables y_t in any DSGE model exceeds the number of shocks (which are exogenous state variables from the perspective of the underlying agents' optimization problems). Most of the literature that estimates loglinearized DSGE models uses empirical specifications in which the number of exogenous shocks is at least as large as the number of observables. Examples are Schorfheide (2000), Rabanal and Rubio-Ramírez (2005), and Smets and Wouters (2007).

The converse of the "more structural shocks" approach would be a "fewer observables" approach, ie, one restricts the number of observables used in the construction of the likelihood function to the number of exogenous shocks included in the model. This raises the question of which observables to include in the likelihood function, which is discussed in Guerrón-Quintana (2010) and Canova et al. (2014). Qu (2015) proposes to use a composite likelihood to estimate singular DSGE models. A composite likelihood function is obtained by partitioning the vector of observables y_t into subsets, eg, $y_t' = [y_{1,t}', y_{2,t}', y_{3,t}']$ for which the likelihood function is nonsingular, eg, "composite likelihood" and then use the product of marginals $p(Y_{1,1:T}|\theta)p(Y_{2,1:T}|\theta)$ $p(Y_{3,1:T}|\theta)$ as the estimation objective function.

11.1.4 Dealing with Lack of Identification

In many applications it is quite difficult to maximize the likelihood function. This difficulty is in part caused by the presence of local extrema and/or weak curvature in some directions of the parameter space and may be a manifestation of identification problems. One potential remedy that has been widely used in practice is to fix a subset of the parameters at plausible values, where "plausible" means consistent with some empirical observations that are not part of the estimation sample Y. Conditional on the fixed parameters, the likelihood function for the remaining parameters may have a more elliptical shape and therefore may be easier to maximize. Of course, such an approach ignores the uncertainty with respect to those parameters that are being fixed. Moreover, if they are fixed at the "wrong" parameter values, inference about the remaining parameters will be distorted.

Building on the broader literature on identification-robust econometric inference, the recent literature has developed inference methods that remain valid even if some parameters of the DSGE model are only weakly or not at all identified. Guerrón-Quintana et al. (2013) propose a method that relies on likelihood-based estimates of the system matrices of the state-space representation $\hat{\Psi}_0$, $\hat{\Psi}_1$, $\hat{\Phi}_1$ and $\hat{\Phi}_\epsilon$. In view of the identification problems associated with the Ψ and Φ matrices discussed in Section 9.1, their approach requires a reparameterization of the state-space matrices in terms of an identifiable reduced-form parameter vector $\phi = f(\theta)$ that, according to the DSGE model, is a function of θ. In the context of our stylized DSGE model, such a reparameterization could be obtained based on the information in Table 4.

Let M_1^ϕ denote the state-space representation of the DSGE model in terms of ϕ and let $\hat{\phi}$ be the ML estimator of ϕ. The hypothesis $H_0 : \theta = \theta_0$ can be translated into the hypothesis $\phi = f(\theta_0)$ and the corresponding likelihood ratio statistic takes the form

$$LR(Y|\theta_0) = 2\left[\log p(Y|\hat{\phi}, M_1^\phi) - \log p(Y|f(\theta_0), M_1^\phi)\right] \Longrightarrow \chi^2_{\dim(\phi)}. \qquad (174)$$

The degrees of freedom of the χ^2 limit distribution depend on the dimension of ϕ (instead of θ), which means that it is important to reduce the dimension of ϕ as much as possible by using a minimal state-variable representation of the DSGE model solution and

to remove elements from the Ψ and Φ matrices that are zero for all values of θ. The likelihood ratio statistic can be inverted to generate a $1 - \alpha$ joint confidence set for the vector θ:

$$CS^\theta(Y) = \{\theta \mid LR(Y|\theta) \leq \chi^2_{crit}\}, \tag{175}$$

where χ^2_{crit} is the $1 - \alpha$ quantile of the $\chi^2_{\dim(\phi)}$ distribution. Subvector inference can be implemented by projecting the joint confidence set on the desired subspace. The inversion of test statistics is computationally tedious because the test statistic has to be evaluated for a wide range of θ values. However, it does not require the maximization of the likelihood function. Guerrón-Quintana et al. (2013) show how the computation of the confidence interval can be implemented based on the output from a Bayesian estimation of the DSGE model.

Andrews and Mikusheva (2015) propose an identification–robust Lagrange multiplier test. The test statistic is based on the score process and its quadratic variation

$$s_{T,t}(\theta) = \nabla_\theta \ell(\theta|Y_{1:t}) - \nabla_\theta \ell(\theta|Y_{1:t-1}), \quad J_T(\theta) = \sum_{t=1}^{T} s_{T,t}(\theta)s'_{T,t}(\theta)$$

and is defined as

$$LM(\theta|Y) = \nabla'_\theta \ell_T(\theta_0|Y)[J_T(\theta_0)]^{-1}\nabla_\theta \ell_T(\theta_0|Y) \Longrightarrow \chi^2_{\dim(\theta_0)}. \tag{176}$$

Note that the degrees of freedom of the χ^2 limit distribution now depend on the dimension of the parameter vector θ instead of the vector of identifiable reduced-form coefficients. A confidence set for θ can be obtained by replacing the LR statistic in (175) with the LM statistic. Andrews and Mikusheva (2015) also consider subvector inference based on a profile likelihood function that concentrates out a subvector of well–identified DSGE model parameters. A frequency domain version of the LM test based on the Whittle likelihood function is provided by Qu (2014). Both Andrews and Mikusheva (2015) and Qu (2014) provide detailed Monte Carlo studies to assess the performance of the proposed identification-robust tests.

11.2 (Simulated) Minimum Distance Estimation

Minimum distance (MD) estimation is based on the idea of minimizing the discrepancy between sample moments of the data, which we denoted by $\hat{m}_T(Y)$, and model–implied moments, which we denoted by $\mathbb{E}[\hat{m}_T(Y)|\theta, M_1]$. The MD estimator $\hat{\theta}_{md}$ was defined in (135) and (136). Examples of the sample statistics $\hat{m}_T(Y)$ are the sample autocovariances $\hat{\Gamma}_{yy}(h)$, a smoothed periodogram $\bar{f}_{yy}(\omega)$ as in Diebold et al. (1998), or estimates of the parameters of an approximating model, eg, the VAR(p) in (113) as in Smith (1993). If $\hat{m}_T(Y)$ consists of parameter estimates of a reference model, then the moment-based estimation is also called indirect inference; see Gourieroux et al. (1993). In some cases it is

possible to calculate the model-implied moments analytically. For instance, suppose that $\hat{m}_T(Y) = \frac{1}{T}\sum \gamma_t \gamma'_{t-1}$, then we can derive

$$\mathbb{E}[\hat{m}_T(Y)|\theta, M_1] = \frac{1}{T}\sum \mathbb{E}[\gamma_t \gamma'_{t-1}|\theta, M_1] = \mathbb{E}[\gamma_2 \gamma'_1|\theta, M_1] \tag{177}$$

from the state-space representation of a linearized DSGE model. Explict formulae for moments of pruned models solved with perturbation methods are provided by Andreasen et al. (2013) (recall Section 4.4). Alternatively, suppose that $\hat{m}_T(Y)$ corresponds to the OLS estimates of a VAR(1). In this case, even for a linear DSGE model, it is not feasible to compute

$$\mathbb{E}[\hat{m}_T(Y)] = \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T} \gamma_{t-1}\gamma'_{t-1}\right)^{-1} \frac{1}{T}\sum_{t=1}^{T} \gamma_{t-1}\gamma'_t \bigg| \theta, M_1\right]. \tag{178}$$

The model-implied expectation of the OLS estimator has to be approximated, for instance, by a population regression:

$$\hat{\mathbb{E}}[\hat{m}_T(Y)] = \left(\mathbb{E}[\gamma_{t-1}\gamma'_{t-1}|\theta, M_1]\right)^{-1} \mathbb{E}[\gamma_{t-1}\gamma'_t|\theta, M_1], \tag{179}$$

or the model-implied moment function has to be replaced by a simulation approximation, which will be discussed in more detail below.

11.2.1 Textbook Analysis

We proceed by sketching the asymptotic approximation of the frequentist sampling distribution of the MD estimator. Define the discrepancy

$$G_T(\theta|Y) = \hat{m}_T(Y) - \hat{\mathbb{E}}[\hat{m}_T(Y)|\theta, M_1], \tag{180}$$

such that the criterion function of the MD estimator in (135) can be written as

$$Q_T(\theta|Y) = \|G_T(\theta|Y)\|_{W_T}. \tag{181}$$

Suppose that there is a unique θ_0 with the property that[as]

$$\hat{m}_T(Y) - \mathbb{E}[\hat{m}_T(Y)|\theta_0, M_1] \xrightarrow{a.s.} 0 \tag{182}$$

and that the sample criterion function $Q_T(\theta|Y)$ converges uniformly almost surely to a limit criterion function $Q(\theta)$, then the MD estimator is consistent in the sense that $\hat{\theta}_{md} \xrightarrow{a.s.} \theta_0$.

The analysis of the MD estimator closely mirrors the analysis of the ML estimator, because both types of estimators are defined as the extremum of an objective function.

[as] In some DSGE models a subset of the series included in γ_t is nonstationary. Thus, moments are only well defined after a stationarity-inducing transformation has been applied. This problem is analyzed in Gorodnichenko and Ng (2010).

The sampling distribution of $\hat{\theta}_{md}$ can be derived from a second-order approximation of the criterion function $Q_T(\theta|Y)$ around θ_0:

$$TQ_T(\theta|Y) = \sqrt{T}\nabla_\theta Q_T(\theta_0|Y)\sqrt{T}(\theta-\theta_0)'$$
$$+ \frac{1}{2}\sqrt{T}(\theta-\theta_0)'\left[\frac{1}{T}\nabla_\theta^2 Q_T(\theta_0|Y)\right]\sqrt{T}(\theta-\theta_0) + \text{small.} \tag{183}$$

If the minimum of $Q_T(\theta|Y)$ is obtained in the interior, then

$$\sqrt{T}(\hat{\theta}_{md}-\theta_0) = \left[-\frac{1}{T}\nabla_\theta^2 Q_T(\theta_0|Y)\right]^{-1}\sqrt{T}\nabla_\theta Q_T(\theta_0|Y) + \text{small.} \tag{184}$$

Using (180), the "score" process can be expressed as

$$\sqrt{T}\nabla_\theta Q_T(\theta_0|Y) = (\nabla_\theta G_T(\theta_0|Y))W_T\sqrt{T}G_T(\theta_0|Y) \tag{185}$$

and its distribution depends on the distribution of

$$\sqrt{T}G_T(\theta_0|Y) = \sqrt{T}(\hat{m}_T(Y) - \mathbb{E}[\hat{m}_T(Y)|\theta_0,M_1])$$
$$+ \sqrt{T}\left(\hat{\mathbb{E}}[\hat{m}_T(Y)|\theta_0,M_1] - \mathbb{E}[\hat{m}_T(Y)|\theta_0,M_1]\right) \tag{186}$$
$$= I + II,$$

say. Term I captures the variability of the deviations of the sample moment $\hat{m}_T(Y)$ from its expected value $\mathbb{E}[\hat{m}_T(Y)|\theta_0,M_1]$ and term II captures the error due to approximating $\mathbb{E}[\hat{m}_T(Y)|\theta_0,M_1]$ by $\hat{\mathbb{E}}[\hat{m}_T(Y)|\theta_0,M_1]$. Under suitable regularity conditions

$$\sqrt{T}G_T(\theta_0|Y) \Longrightarrow N(0,\mathbf{\Omega}), \tag{187}$$

and

$$\sqrt{T}(\hat{\theta}_{md}-\theta_0) \Longrightarrow N\left(0,(DWD')^{-1}DW\mathbf{\Omega}WD'(DWD')^{-1}\right), \tag{188}$$

where W is the limit of the sequence of weight matrices W_T and the matrix D is defined as the probability limit of $\nabla_\theta G_T(\theta_0|Y)$. To construct tests and confidence sets based on the limit distribution, the matrices D and $\mathbf{\Omega}$ have to be replaced by consistent estimates. We will discuss the structure of $\mathbf{\Omega}$ in more detail below.

If the number of moment conditions exceeds the number of parameters, then the model specification can be tested based on the overidentifying moment conditions. If $W_T = [\hat{\mathbf{\Omega}}_T]^{-1}$, where $\hat{\mathbf{\Omega}}_T$ is a consistent estimator of $\mathbf{\Omega}$, then

$$TQ_T(\hat{\theta}_{md}|Y) \Longrightarrow \chi_{df}^2, \tag{189}$$

where the degrees of freedom df equal the number of overidentifying moment conditions. The sample objective function can also be used to construct hypothesis tests for θ. Suppose that the null hypothesis is $\theta = \theta_0$. A quasi-likelihood ratio test is based on $T(Q_T(\theta_0|Y) - Q_T(\hat{\theta}_{md}|Y))$; a quasi-Lagrange-multiplier test is based on a properly

standardized quadratic form of $\sqrt{T}\nabla_\theta Q_T(\theta_0|Y)$; and a Wald test is based on a properly standardized quadratic form of $\sqrt{T}(\hat\theta_{md} - \theta_0)$. Any of these test statistics can be inverted to construct a confidence set. Moreover, if the parameters suffer from identification problems, then the approach of Andrews and Mikusheva (2015) can be used to conduct identification-robust inference based on the quasi-Lagrange-multiplier test.

11.2.2 Approximating Model-Implied Moments

In many instances the model-implied moments $\mathbb{E}[m_T(Y)|\theta, M_1]$ are approximated by an estimate $\hat{\mathbb{E}}[m_T(Y)|\theta, M_1]$. This approximation affects the distribution of $\hat\theta_{md}$ through term II in (186). Consider the earlier example in (178) and (179) in which $\hat{m}_T(Y)$ corresponds to the OLS estimates of a VAR(1). Because the OLS estimator has a bias that vanishes at rate $1/T$, we can deduce that term II converges to zero and does not affect the asymptotic covariance matrix Ω.

The more interesting case is the one in which $\hat{\mathbb{E}}[m_T(Y)|\theta, M_1]$ is based on the simulation of the DSGE model. The asymptotic theory for simulation-based extremum estimators has been developed in Pakes and Pollard (1989). Lee and Ingram (1991) and Smith Jr. (1993) are the first papers that use simulated method of moments to estimate DSGE models. For concreteness, suppose that $m_T(Y)$ corresponds to the first-order (uncentered) sample autocovariances. We previously showed that, provided the y_t's are stationary, $\mathbb{E}[m_T(Y)|\theta, M_1]$ is given by the DSGE model population autocovariance matrix $\mathbb{E}[y_2 y_1'|\theta, M_1]$, which can be approximated by simulating a sample of length λT of artificial observations Y^* from the DSGE model M_1 conditional on θ. Based on these simulated observations one can compute the sample autocovariances $\hat{m}_{\lambda T}(Y^*(\theta, M_1))$. In this case term II is given by

$$II = \frac{1}{\sqrt{\lambda}}\sqrt{\lambda T}\left(\frac{1}{\lambda T}\sum_{t=1}^{\lambda T} y_t^* y_{t-1}^* - \mathbb{E}[y_2 y_1'|\theta_0, M_1]\right) \tag{190}$$

and satisfies a CLT. Because the simulated data are independent of the actual data, terms I and II in (186) are independent and we can write

$$\Omega = \mathbb{V}_\infty[I] + \mathbb{V}_\infty[II], \tag{191}$$

where

$$\mathbb{V}_\infty[II] = \frac{1}{\lambda}\left(\lim_{T\to\infty} T\mathbb{V}[\hat{m}_T(Y^*(\theta_0, M_1))]\right) \tag{192}$$

and can be derived from the DSGE model. The larger λ, the more accurate the simulation approximation and the contribution of $\mathbb{V}_\infty[II]$ to the overall covariance matrix Ω.

We generated the simulation approximation by simulating one long sample of observations from the DSGE model. Alternatively, we could have simulated λ samples Y^i, $i = 1, \lambda$ of size T. It turns out that for the approximation, say, of $\mathbb{E}[y_2 y_1'|\theta, M_1]$, it does not matter because $\hat{m}_T(Y^*(\theta, M_1))$ is an unbiased estimator of $\mathbb{E}[y_2 y_1'|\theta, M_1]$. However, if

$\hat{m}_T(Y)$ is defined as the OLS estimator of a VAR(1), then the small–sample bias of the OLS estimator generates an $O(T^{-1})$ wedge between

$$\left(\sum_{t=1}^{\lambda T} y_{t-1}^* y_{t-1}^{*\prime}\right)^{-1} \sum_{t=1}^{\lambda T} y_{t-1}^* y_{t-1}^{*\prime} \quad \text{and} \quad \mathbb{E}\left[\left(\sum_{t=1}^{T} y_{t-1} y_{t-1}'\right)^{-1} \sum_{t=1}^{T} y_{t-1} y_{t-1}' \middle| \theta, M_1\right].$$

For large values of λ, this wedge can be reduced by using

$$\hat{E}[m_T(Y)|\theta, M_1] = \frac{1}{\lambda}\sum_{i=1}^{\lambda}\left(\sum_{t=1}^{T} y_{t-1}^i y_{t-1}^{i\prime}\right)^{-1} \sum_{t=1}^{T} y_{t-1}^i y_{t-1}^{i\prime}$$

instead. Averaging OLS estimators from model–generated data reproduces the $O(T^{-1})$ bias of the OLS estimator captured by $\mathbb{E}[\hat{m}_T(Y)|\theta_0, M_1]$ and can lead to a final sample bias reduction in term *II*, which improves the small sample performance of $\hat{\theta}_{md}$.[at]

When implementing the simulation approximation of the moments, it is important to fix the random seed when generating the sample Y^* such that for each parameter value of θ the same sequence of random variables is used in computing $Y^*(\theta, M_1)$. This ensures that the sample objective function $Q_T(\theta|Y)$ remains sufficiently smooth with respect to θ to render the second-order approximation of the objective function valid.

11.2.3 Misspecification

Under the assumption that the DSGE model is correctly specified, the MD estimator has a well–defined almost–sure limit θ_0 and the asymptotic variance $\mathbb{V}_\infty[I]$ of term *I* in (186) is given by the model–implied variance

$$\mathbb{V}_\infty[I] = \left(\lim_{T\to\infty} T\mathbb{V}[\hat{m}_T(Y^*(\theta_0, M_1))]\right), \tag{193}$$

which up to the factor of $1/\lambda$ is identical to the contribution $\mathbb{V}_\infty[II]$ of the simulation approximation of the moments to the overall asymptotic variance Ω; see (192). Under the assumption of correct specification, it is optimal to choose the weight matrix W based on the accuracy with which the elements of the moment vector $\hat{m}_T(Y)$ measure the population analog $\mathbb{E}[\hat{m}_T(Y)|\theta_0, M_1]$. If the number of moment conditions exceeds the number of parameters, it is optimal (in the sense of minimizing the sampling variance of $\hat{\theta}_{md}$) to place more weight on matching moments that are accurately measured in the data, by setting $W = \Omega^{-1}$. In finite sample, one can construct W_T from a consistent estimator of Ω^{-1}.

[at] See Gourieroux et al. (2010) for a formal analysis in the context of a dynamic panel data model.

If the DSGE model is regarded as misspecified, then the sampling distribution of the MD estimator has to be derived under the distribution of a reference model $p(Y|M_0)$. In this case we can define

$$\theta_0(Q) = \lim_{T\to\infty} \text{argmin}_\theta \, \|\mathbb{E}[\hat{m}_T(Y)|M_0] - \mathbb{E}[\hat{m}|\theta, M_1]\|_W \tag{194}$$

and, under suitable regularity, the estimator $\hat{\theta}_{md}$ will converge to the pseudo-optimal value θ_0. Note that θ_0 is a function of the moments $\hat{m}_T(Y)$ that are being matched and the weight matrix W (indicated by the Q argument). Both $\hat{m}$ and W are chosen by the researcher based on the particular application. The vector $\hat{m}$ should correspond to a set of moments that are deemed to be informative about the desired parameterization of the DSGE model and reflect the ultimate purpose of the estimated DSGE model. The weight matrix W should reflect beliefs about the informativeness of certain sample moments with respect to the desired parameterization of the DSGE model.

To provide an example, consider the case of a DSGE model with stochastic singularity that attributes all business cycle fluctuations to technology shocks. To the extent that the observed data are not consistent with this singularity, the model is misspecified. A moment-based estimation of the model will ultimately lead to inflated estimates of the standard deviation of the technology shock innovation, because this shock alone has to generate the observed variability in, say, output growth, the labor share, and other variables. The extent to which the estimated shock variance is upwardly biased depends on exactly which moments the estimator is trying to match. If one of the priorities of the estimation exercise is to match the unconditional variance of output growth, then the weight matrix W should assign a large weight to this moment, even if it is imprecisely measured by its sample analog in the data.

The asymptotic variance $\mathbb{V}_\infty[I]$ of term I in (186) is now determined by the variance of the sample moments implied by the reference model M_0:

$$\mathbb{V}_\infty[I] = \left(\lim_{T\to\infty} T\mathbb{V}[\hat{m}_T(Y)|M_0]\right). \tag{195}$$

Suppose that $\hat{m}_T(Y) = \frac{1}{T}\sum_{t=1}^{T} y_t y_{t-1}'$, which under suitable regularity conditions converges to the population autocovariance matrix $\mathbb{E}[y_1 y_0'|M_0]$ under the reference model M_0. If the reference model is a linear process, then the asymptotic theory developed in Phillips and Solo (1992) can be used to determine the limit covariance matrix $\mathbb{V}_\infty[I]$. An estimate of $\mathbb{V}_\infty[I]$ can be obtained with a heteroskedasticity and autocorrelation consistent (HAC) covariance matrix estimator that accounts for the serial correlation in the matrix-valued sequence $\{y_t y_{t-1}'\}_{t=1}^{T}$. An extension of indirect inference in which $\hat{m}_T(Y)$ comprises estimates of an approximating model to the case of misspecified DSGE models is provided in Dridi et al. (2007).

11.2.4 Illustration

Detailed studies of the small–sample properties of MD estimators for DSGE models can be found in Ruge-Murcia (2007) and Ruge-Murcia (2012). To illustrate the behavior of the MD estimator we repeatedly generate data from the stylized DSGE model, treating the values listed in Table 5 as "true" parameters. We fix all parameters except for the Calvo parameter ζ_p at their "true" values and use two versions of the MD procedure to estimate ζ_p. The vector of moment conditions $\hat{m}_T(Y)$ is defined as follows. Let $y_t = [\log(X_t/X_{t-1}), \pi_t]'$ and consider a VAR(2) in output growth and inflation:

$$y_t = \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \Phi_0 + u_t. \tag{196}$$

Let $\hat{m}_T(Y) = \hat{\Phi}$ be the OLS estimate of $[\Phi_1, \Phi_2, \Phi_0]'$.

The results in the left panel of Fig. 28 are obtained by a simulation approximation of the model–implied expected value of $\hat{m}_T(Y)$. We simulate $N = 100$ trajectories of length $T + T_0$ and discarding the first T_0 observations. Let $Y_{1:T}^{(i)}(\theta)$ be the i-th simulated trajectory and define

$$\mathbb{E}[\hat{m}_T(Y)|\theta, M_1] \approx \frac{1}{N} \sum_{i=1}^{N} \hat{m}_T(Y^{(i)}(\theta)), \tag{197}$$

which can be used to evaluate the objective function (181). For the illustration we use the optimal weight matrix $W_T = \hat{\Sigma}^{-1} \otimes X'X$, where X is the matrix of regressors for the VAR(2) and $\hat{\Sigma}$ an estimate of the covariance matrix of the VAR innovations. Because we are estimating a single parameter, we compute the estimator $\hat{\theta}_{md}$ by grid search. It
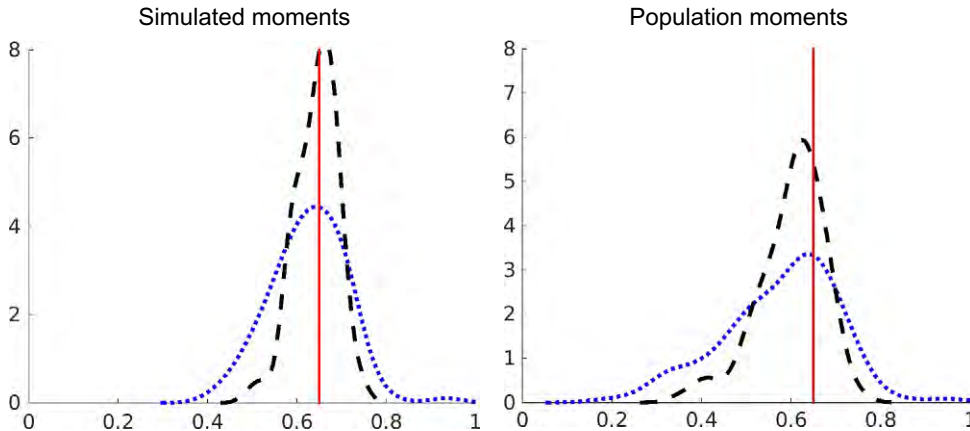

Fig. 28 Sampling distribution of $\hat{\zeta}_{p,md}$. *Notes*: We simulate samples of size $T = 80$ (*dotted*) and $T = 200$ (*dashed*) and compute two versions of an MD estimator for the Calvo parameter ζ_p. All other parameters are fixed at their "true" value. The plots depict densities of the sampling distribution of $\hat{\zeta}_{p,md}$. The vertical line indicates the "true" value of ζ_p.

is important to use the same sequence of random numbers for each value of $\theta \in \mathcal{T}$ to compute the simulation approximation $\mathbb{E}[\hat{m}_T(Y)|\theta, M_1]$. The results in the right panel of Fig. 28 are based on the VAR(2) approximation of the DSGE model based on a population regression. Let $x_t' = [y_{t-1}', y_{t-2}', 1]$ and let

$$\mathbb{E}[\hat{m}_T(Y)|\theta, M_1] \approx \left(\mathbb{E}[x_t x_t'|\theta, M_1]\right)^{-1} \mathbb{E}[x_t y_t'|\theta, M_1]. \tag{198}$$

Fig. 28 depicts density estimates of the sampling distribution of $\hat{\zeta}_{p,md}$. The vertical line indicates the "true" parameter value of ζ_p. As the sample size increases from $T = 80$ to $T = 200$, the sampling distribution concentrates around the "true" value and starts to look more like a normal distribution, as the asymptotic theory presented in this section suggests. The distribution of the estimator based on the simulated objective function is more symmetric around the "true" value and also less variable. However, even based on a sample size of 200 observations, there is considerable uncertainty about the Calvo parameter and hence the slope of the New Keynesian Phillips curve. A comparison with Fig. 27 indicates that the MD estimator considered in this illustration is less efficient than the ML estimator.

11.2.5 Laplace Type Estimators

In DSGE model applications the estimation objective function $Q_T(\theta|Y)$ is often difficult to optimize. Chernozhukov and Hong (2003) proposed computing a mean of a quasi-posterior density instead of computing an extremum estimator. The resulting estimator is called a Laplace-type (LT) estimator and defined as follows (provided the integral in the denominator is well defined):

$$\hat{\theta}_{LT} = \frac{\exp\left\{-\frac{1}{2}Q_T(\theta|Y)\right\}}{\int \exp\left\{-\frac{1}{2}Q_T(\theta|Y)\right\}d\theta}. \tag{199}$$

This estimator can be evaluated using the Metropolis–Hastings algorithm discussed in Section 12.2 or the sequential Monte Carlo algorithm presented in Section 12.3 below. The posterior computations may be more accurate than the computation of an extremum. Moreover, suppose that the objective function is multimodal. In repeated sampling, the extremum of the objective function may shift from one mode to the other, making the estimator appear to be unstable. On the other hand, owing to the averaging, the LT estimator may be more stable. Chernozhukov and Hong (2003) establish the consistency and asymptotic normality of LT estimators, which is not surprising because the sample objective function concentrates around its extremum as $T \to \infty$ and the discrepancy between the extremum and the quasi-posterior mean vanishes. DSGE model applications of LT estimators are provided in Kormilitsina and Nekipelov (2012, 2016).

LT estimators can be constructed not only from MD estimators but also from IRF matching estimators and GMM estimators discussed below.

11.3 Impulse Response Function Matching

As discussed previously, sometimes DSGE models are misspecified because researchers have deliberately omitted structural shocks that contribute to business cycle fluctuations. An example of such a model is the one developed by Christiano et al. (2005). The authors focus their analysis on the propagation of a single shock, namely, a monetary policy shock. If it is clear that if the DSGE model does not contain enough structural shocks to explain the variability in the observed data, then it is sensible to try to purge the effects of the unspecified shocks from the data, before matching the DSGE model to the observations. This can be done by "filtering" the data through the lens of a VAR that identifies the impulse responses to those shocks that are included in the DSGE model. The model parameters can then be estimated by minimizing the discrepancy between model–implied and empirical impulse response functions. A mismatch between the two sets of impulse responses provides valuable information about the misspecification of the propagation mechanism and can be used to develop better-fitting DSGE models. Influential papers that estimate DSGE models by matching impulse response functions include Rotemberg and Woodford (1997), Christiano et al. (2005), and Altig et al. (2011). The casual description suggests that impulse response function matching estimators are a special case of the previously discussed MD estimators (the DSGE model M_1 is misspecified and a structural VAR serves as reference model M_0 under which the sampling distribution of the estimator is derived). Unfortunately, several complications arise, which we will discuss in the remainder of this section. Throughout, we assume that the DSGE model has been linearized. An extension to the case of nonlinear DSGE models is discussed in Ruge-Murcia (2014).

11.3.1 Invertibility and Finite-Order VAR Approximations

The empirical impulse responses are based on a finite–order VAR, such as the one in (113). However, even linearized DSGE models typically cannot be written as a finite–order VAR. Instead, they take the form of a state-space model, which typically has a VARMA representation. In general we can distinguish the following three cases: (i) the solution of the DSGE model can be expressed as a VAR(p). For the stylized DSGE model, this is the case if γ_t is composed of four observables: output growth, the labor share, inflation, and interest rates. (ii) The moving average polynomial of the VARMA representation of the DSGE model is invertible. In this case the DSGE model can be expressed as an infinite–order VAR driven by the structural shock innovations ϵ_t. (iii) The moving average polynomial of the VARMA representation of the DSGE model is not invertible. In this case the innovation of the VAR(∞) approximation do not correspond to the structural innovations ϵ_t. Only in case (i) can one expect a direct match between the empirical IRFs and the DSGE

model IRFs. Cases (ii) and (iii) complicate econometric inference. The extent to which impulse–response–function–based estimation and model evaluation may be misleading has been fiercely debated in Christiano et al. (2007) and Chari et al. (2008).

Fernández-Villaverde et al. (2007) provide formal criteria to determine whether a DSGE model falls under case (i), (ii), or (iii). Rather than presenting a general analysis of this problem, we focus on a simple example. Consider the following two MA processes that represent the DSGE models in this example:

$$
\begin{aligned}
M_1 &: \ y_t = \epsilon_t + \theta\epsilon_{t-1} = (1 + \theta L)\epsilon_t \\
M_2 &: \ y_t = \theta\epsilon_t + \epsilon_{t-1} = (\theta + L)\epsilon_t,
\end{aligned}
\tag{200}
$$

where $0 < \theta < 1$, L denotes the lag operator, and $\epsilon_t \sim iidN(0,1)$. Models M_1 and M_2 are observationally equivalent, because they are associated with the same autocovariance sequence. The root of the MA polynomial of model M_1 is outside of the unit circle, which implies that the MA polynomial is invertible and one can express y_t as an AR(∞) process:

$$
AR(\infty) \text{ for } M_1 : \ y_t = -\sum_{j=1}^{\infty}(-\theta)^j y_{t-j} + \epsilon_t.
\tag{201}
$$

It is straightforward to verify that the AR(∞) approximation reproduces the impulse response function of M_1:

$$
\frac{\partial y_t}{\partial \epsilon_t} = 1, \quad \frac{\partial y_{t+1}}{\partial \epsilon_t} = \theta, \quad \frac{\partial y_{t+h}}{\partial \epsilon_t} = 0 \text{ for } h > 1.
$$

Thus, the estimation of an autoregressive model with many lags can reproduce the monotone impulse response function of model M_1.

The root of the MA polynomial of M_2 lies inside the unit circle. While M_2 could also be expressed as an AR(∞), it would be a representation in terms of a serially uncorrelated one-step-ahead forecast error u_t that is a function of the infinite history of the ϵ_t's: $u_t = (1+\theta L)^{-1}(\theta + L)$. As a consequence, the AR(∞) is unable to reproduce the hump-shaped IRF of model M_2. More generally, if the DSGE model is associated with a noninvertible moving average polynomial, its impulse responses cannot be approximated by a VAR(∞) and a direct comparison of VAR and DSGE IRFs may be misleading.

11.3.2 Practical Considerations

The objective function for the IRF matching estimator takes the same form as the criterion function of the method of moments estimator in (180) and (181), where $\hat{m}_T(Y)$ is the VAR IRF. For $\hat{\mathbb{E}}[\hat{m}_T(Y)|\theta, M_1]$ researchers typically just use the DSGE model impulse response, say, $IRF(\cdot|\theta, M_1)$. In view of the problems caused by noninvertible moving-average polynomials and finite-order VAR approximations of infinite-order VAR representations, a more prudent approach would be to replace

$IRF(\cdot|\theta,M_1)$ by average impulse response functions that are obtained by repeatedly simulating data from the DSGE model (given θ) and estimating a structural VAR, as in the indirect inference approach described in Section 11.2. Such a modification would address the concerns about IRF matching estimators raised by Chari et al. (2008).

The sampling distribution of the IRF matching estimator depends on the sampling distribution of the empirical VAR impulse responses $\hat{m}_T(Y)$ under the VAR M_0. An approximation of the distribution of $\hat{m}_T(Y)$ could be obtained by first-order asymptotics and the delta method as in Lütkepohl (1990) and Mittnik and Zadrozny (1993) for stationary VARs; or as in Phillips (1998), Rossi and Pesavento (2006), and Pesavento and Rossi (2007) for VARs with persistent components. Alternatively, one could use the bootstrap approximation proposed by Kilian (1998, 1999). If the number of impulse responses stacked in the vector $\hat{m}_T(Y)$ exceeds the number of reduced-form VAR coefficient estimates, then the sampling distribution of the IRFs becomes asymptotically singular. Guerrón-Quintana et al. (2014) use nonstandard asymptotics to derive the distribution of IRFs for the case in which there are more responses than reduced-form parameters.

Because for high-dimensional vectors $\hat{m}_T(Y)$ the joint covariance matrix may be close to singular, researchers typically choose a diagonal weight matrix W_T, where the diagonal elements correspond to the inverse of the sampling variance for the estimated response of variable i to shock j at horizon h. As discussed in Section 11.2, to the extent that the DSGE model is misspecified, the choice of weight matrix affects the probability limit of the IRF matching estimator and should reflect the researcher's loss function.

In fact, impulse response function matching is appealing only if the researcher is concerned about model misspecification. This misspecification might take two forms: First, the propagation mechanism of the DSGE model is potentially misspecified and the goal is to find pseudo-optimal parameter values that minimize the discrepancy between empirical and model-implied impulse responses. Second, the propagation mechanisms for the shocks of interest are believed to be correctly specified, but the model lacks sufficiently many stochastic shocks to capture the observed variation in the data. In the second case, it is in principle possible to recover the subset of "true" DSGE model parameters θ_0 that affect the propagation of the structural shock for which the IRF is computed. The consistent estimation would require that the DSGE model allow for a VAR(∞) representation in terms of the structural shock innovations ϵ_t; that the number of lags included in the empirical VAR increase with sample size T; and that the VAR identification scheme correctly identify the shock of interest if the data are generated from a version of the DSGE model that is augmented by additional structural shocks.

11.3.3 Illustration

To illustrate the properties of the IRF matching estimator, we simulate data from the stylized DSGE model using the parameter values given in Table 5. We assume that

the econometrician considers an incomplete version of the DSGE model that only includes the monetary policy shock and omits the remaining shocks. Moreover, we assume that the econometrician only has to estimate the degree of price stickiness captured by the Calvo parameter ζ_p. All other parameters are fixed at their "true" values during the estimation.

The empirical impulse response functions stacked in the vector $\hat{m}_T(Y)$ are obtained by estimating a VAR(p) for interest rates, output growth, and inflation:

$$y_t = [R_t - \pi_t/\beta, \log(X_t/X_{t-1}), \pi_t]'. \tag{202}$$

The first equation of this VAR represents the monetary policy rule of the DSGE model. The interest rate is expressed in deviations from the central bank's systematic reaction to inflation. Thus, conditional on β, the monetary policy shock is identified as the orthogonalized one-step–ahead forecast error in the first equation of the VAR. Upon impact, the response of y_t to the monetary policy shock is given by the first column of the lower-triangular Cholesky factor of the covariance matrix Σ of the reduced-form innovations u_t.

Because y_t excludes the labor share, the state-space representation of the DSGE model cannot be expressed as a finite-order VAR. However, we can construct a VAR approximation of the DSGE model as follows. Let $x_t = [y'_{t-1}, \ldots, y'_{t-p}, 1']'$ and define the functions[au]

$$\Phi_*(\theta) = \left(\mathbb{E}[x_t x'_t|\theta, M_1]\right)^{-1}\left(\mathbb{E}[x_t y'_t|\theta, M_1]\right),$$

$$\Sigma^*(\theta) = \mathbb{E}[y_t y'_t|\theta, M_1] - \mathbb{E}[y_t x'_t|\theta, M_1]\left(\mathbb{E}[x_t x'_t|\theta, M_1]\right)^{-1}\mathbb{E}[x_t y'_t|\theta, M_1]. \tag{203}$$

Note that $\Phi_*(\theta)$ and $\Sigma_*(\theta)$ are functions of the population autocovariances of the DSGE model. For a linearized DSGE model, these autocovariances can be expressed analytically as a function of the coefficient matrices of the model's state-space representation.

The above definition of $\Phi_*(\theta)$ and $\Sigma^*(\theta)$ requires that $\mathbb{E}[x_t x'_t|\theta, M_1]$ is nonsingular. This condition is satisfied as long as $n_y \leq n_\epsilon$. However, the appeal of IRF matching estimators is that they can be used in settings in which only a few important shocks are incorporated into the model and $n_y > n_\epsilon$. In this case, $\Phi_*(\theta)$ and $\Sigma^*(\theta)$ have to be modified, for instance, by computing the moment matrices based on $\tilde{y}_t = y_t + u_t$, where u_t is a "measurement error," or by replacing $\left(\mathbb{E}[x_t x'_t|\theta, M_1]\right)^{-1}$ with $\left(\mathbb{E}[x_t x'_t|\theta, M_1] + \lambda I\right)^{-1}$, where λ is a scalar and I is the identity matrix. In the subsequent illustration, we keep all the structural shocks in the DSGE model active, ie, $n_y \leq n_\epsilon$, such that the restriction functions can indeed be computed based on (203).

[au] For the evaluation of the moment matrices $\mathbb{E}[\cdot|\theta, M_1]$ see Section 8.2.1.
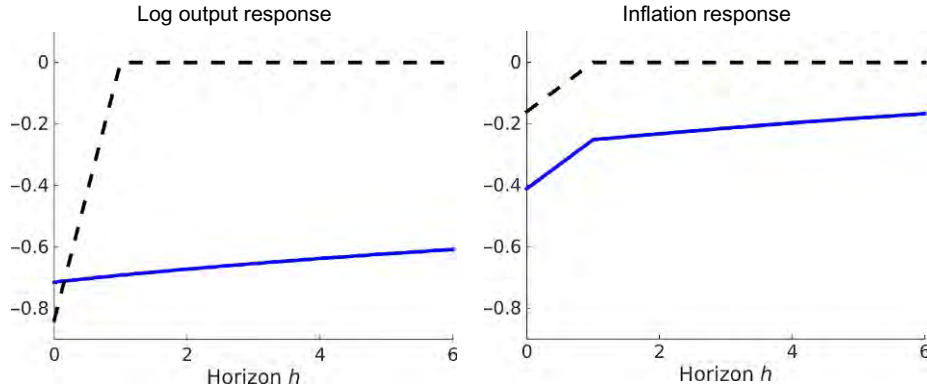
Fig. 29 DSGE model and VAR impulse responses to a monetary policy shock. *Notes*: The figure depicts impulse responses to a monetary policy shock computed from the state-space representation of the DSGE model (*dashed*) and the VAR(1) approximation of the DSGE model (*solid*).

Fig. 29 compares the impulse responses from the state-space representation and the VAR approximation of the DSGE model. It turns out that there is a substantial discrepancy. Because the monetary policy shock is *iid* and the stylized DSGE model does not have an endogenous propagation mechanism, both output and inflation revert back to the steady state after one period. The VAR response, on the other hand, is more persistent and the relative movement of output and inflation is distorted. Augmenting a VAR(1) with additional lags has no noticeable effect on the impulse response.

The IRF matching estimator minimizes the discrepancy between the empirical and the DSGE model-implied impulse responses by varying ζ_p. Fig. 30 illustrates the effect of ζ_p on the response of output and inflation. The larger ζ_p, the stronger the nominal rigidity, and the larger the effect of a monetary policy shock on output. Fig. 31 shows the sampling distribution of the IRF matching estimator for the sample sizes $T = 80$ and $T = 200$. We match IRFs over 10 horizons and use an identity weight matrix. If $\hat{\mathbb{E}}[\hat{m}_T(Y)|\theta, M_1]$ is defined as the IRF implied by the state-space representation, then the resulting estimator of ζ_p has a fairly strong downward bias. This is not surprising in view of the mismatch depicted in Figs. 29 and 30. If the state-space IRF is replaced by the IRF obtained from the VAR approximation of the DSGE model, then the sampling distribution is roughly centered at the "true" parameter value, though it is considerably more dispersed, also compared to the MD estimator in Fig. 28. This is consistent with the fact that the IRF matching estimator does not utilize variation in output and inflation generated by the other shocks.
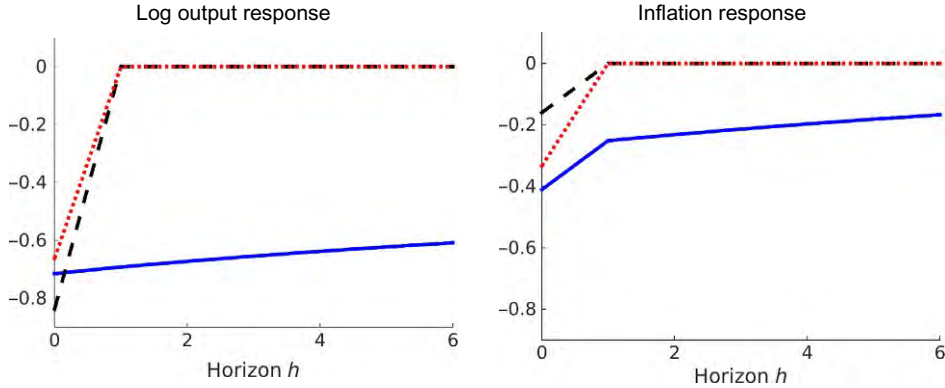
Fig. 30 Sensitivity of IRF to ζ_p. *Notes*: The *solid lines* indicate IRFs computed from the VAR approximation of the DSGE model. The other two lines depict DSGE model-implied IRFs based on $\zeta_p = 0.65$ (*dashed*) and $\zeta_p = 0.5$ (*dotted*).
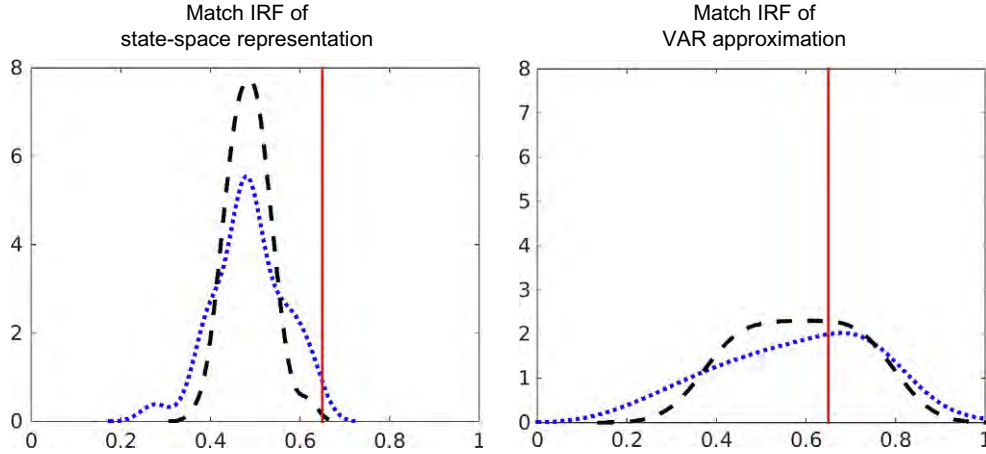

Fig. 31 Sampling distribution of $\hat{\zeta}_{p,irf}$. *Notes*: We simulate samples of size $T = 80$ and $T = 200$ and compute IRF matching estimators for the Calvo parameter ζ_p based on two choices of $\hat{\mathbb{E}}[\hat{m}_T(Y)|\theta, M_1]$. For the *left panel* we use the IRFs from the state-space representation of the DSGE model; for the *right panel* we use the IRF from the VAR approximation of the DSGE model. All other parameters are fixed at their "true" value. The plot depicts densities of the sampling distribution of $\hat{\zeta}_p$ for $T = 80$ (*dotted*) and $T = 200$ (*dashed*). The vertical line indicates the "true" value of ζ_p.

11.4 GMM Estimation

We showed in Section 8.2.4 that one can derive moment conditions of the form

$$\mathbb{E}[g(y_{t-p:t}|\theta, M_1)] = 0 \qquad (204)$$

for $\theta = \theta_0$ from the DSGE model equilibrium. For instance, based on (106) and (107) we could define

$$g(y_{t-p:t}|\theta,M_1) = \left[\begin{array}{c} (-\log(X_t/X_{t-1}) + \log R_{t-1} - \log\pi_t - \log(1/\beta))Z_{t-1} \\ (\log R_t - \log(\gamma/\beta) - \psi\log\pi_t - (1-\psi)\log\pi^*)Z_{t-1} \end{array} \right]. \tag{205}$$

The identifiability of θ requires that the moments be different from zero whenever $\theta \neq \theta_0$. A GMM estimator is obtained by replacing population expectations by sample averages. Let

$$G_T(\theta|Y) = \frac{1}{T}\sum_{t=1}^{T} g(y_{t-p:t}|\theta,M_1). \tag{206}$$

The GMM objective function is given by

$$Q_T(\theta|Y) = G_T(\theta|Y)' W_T G_T(\theta|Y) \tag{207}$$

and looks identical to the objective function studied in Section 11.2. In turn, the analysis of the sampling distribution of $\hat{\theta}_{md}$ carries over to the GMM estimator.

The theoretical foundations of GMM estimation were developed by Hansen (1982), who derived the first-order asymptotics for the estimator assuming that the data are stationary and ergodic. Christiano and Eichenbaum (1992) and Burnside et al. (1993) use GMM to estimate the parameters of real business cycle DSGE models. These papers use sufficiently many moment conditions to be able to estimate all the parameters of their respective DSGE models. GMM estimation can also be applied to a subset of the equilibrium conditions, eg, the consumption Euler equation or the New Keynesian Phillips curve to estimate the parameters related to these equilibrium conditions.

Unlike all the other estimators considered in this paper, the GMM estimators do not require the researchers to solve the DSGE model. To the extent that solving the model is computationally costly, this can considerably speed up the estimation process. Moreover, one can select moment conditions that do not require assumptions about the law of motion of exogenous driving processes, which robustifies the GMM estimator against misspecification of the exogenous propagation mechanism. However, it is difficult to exploit moment conditions in which some of the latent variables appear explicitly. For instance, consider the Phillips curve relationship of the stylized DSGE model, which suggests setting

$$g(y_{t-p:t}|\theta,M_1) = \left(\hat{\pi}_{t-1} - \beta\hat{\pi}_t - \kappa_p(\widehat{lsh}_{t-1}) \right) Z_{t-1}. \tag{208}$$

Note that λ_{t-1} is omitted from the definition of $g(y_{t-p:t}|\theta,M_1)$ because it is unobserved. However, as soon as Z_t is correlated with the latent variable λ_t the expected value of $g(y_{t-p:t}|\theta,M_1)$ is nonzero even for $\theta = \theta_0$:

$$\mathbb{E}[g(y_{t-p:t}|\theta_0,M_1)] = -\kappa_0\mathbb{E}[\lambda_{t-1}Z_{t-1}] \neq 0. \tag{209}$$

To the extent that λ_t is serially correlated, using higher–order lags of γ_t as instruments does not solve the problem.[av] Recent work by Gallant et al. (2013) and Shin (2014) considers extensions of GMM estimation to moment conditions with latent variables.

The recent literature on GMM estimation of DSGE models has focused on identification–robust inference in view of the weak identification of Phillips curve and monetary policy rule parameters. Generic identification problems in the context of monetary policy rule estimation are highlighted in Cochrane (2011) and methods to conduct identification–robust inference are developed in Mavroeidis (2010). Identification–robust inference for Phillips curve parameters is discussed in Mavroeidis (2005), Kleibergen and Mavroeidis (2009), and Mavroeidis et al. (2014). Dufour et al. (2013) consider identification–robust moment-based estimation of all of the equilibrium relationships of a DSGE model.

12. BAYESIAN ESTIMATION TECHNIQUES

Bayesian inference is widely used in empirical work with DSGE models. The first papers to estimate small–scale DSGE models using Bayesian methods were DeJong et al. (2000), Schorfheide (2000), Otrok (2001), Fernández-Villaverde and Rubio-Ramírez (2004), and Rabanal and Rubio-Ramírez (2005). Subsequent papers estimated open–economy DSGE models, eg, Lubik and Schorfheide (2006), and larger DSGE models tailored to the analysis of monetary policy, eg, Smets and Wouters (2003) and Smets and Wouters (2007). Because Bayesian analysis treats shock, parameter, and model uncertainty symmetrically by specifying a joint distribution that is updated in view of the observations Y, it provides a conceptually appealing framework for decision making under uncertainty. Levin et al. (2006) consider monetary policy analysis under uncertainty based on an estimated DSGE model and the handbook chapter by Del Negro and Schorfheide (2013) focuses on forecasting with DSGE models.

Conceptually, Bayesian inference is straightforward. A prior distribution is updated in view of the sample information contained in the likelihood function. This leads to a posterior distribution that summarizes the state of knowledge about the unknown parameter vector θ. The main practical difficulty is the calculation of posterior moments and quantiles of transformations $h(\cdot)$ of the parameter vector θ. The remainder of this section is organized as follows. We provide a brief discussion of the elicitation of prior distributions in Section 12.1. Sections 12.2 and 12.3 discuss two important algorithms to generate parameter draws from posterior distributions: Markov chain Monte Carlo (MCMC) and sequential Monte Carlo (SMC). Bayesian model diagnostics are reviewed in

---

[av] Under the assumption that λ_t follows an AR(1) process, one could quasi–difference the Phillips curve, which would replace the term $\lambda_{t-1}Z_{t-1}$ with $\epsilon_{\lambda,t-1}Z_{t-1}$. If Z_{t-1} is composed of lagged observables dated $t-2$ and earlier, then the validity of the moment condition is restored.

Section 12.4. Finally, we discuss the recently emerging literature on limited–information Bayesian inference in Section 12.5. Sections 12.1 and 12.3 are based on Herbst and Schorfheide (2015), who provide a much more detailed exposition. Section 12.4 draws from Del Negro and Schorfheide (2011).

12.1 Prior Distributions

There is some disagreement in the Bayesian literature about the role of prior information in econometric inference. Some authors advocate "flat" prior distributions that do not distort the shape of the likelihood function, which raises two issues: first, most prior distributions are not invariant under parameter transformations. Suppose a scalar parameter $\theta \sim U[-M,M]$. If the model is reparameterized in terms of $1/\theta$, the implied prior is no longer flat. Second, if the prior density is taken to be constant on the real line, say, $p(\theta) = c$, then the prior is no longer proper, meaning the total prior probability mass is infinite. In turn, it is no longer guaranteed that the posterior distribution is proper.

In many applications prior distributions are used to conduct inference in situations in which the number of unknown parameters is large relative to the number of sample observations. An example is a high–dimensional VAR. If the number of variables in the VAR is n and the number of lags is p, then each equation has at least np unknown parameters. For instance, a 4-variable VAR with $p = 4$ lags has 16 parameters. If this model is estimated based on quarterly post-Great Moderation and pre-Great Recession data, the data-to-parameter ratio is approximately 6, which leads to very noisy parameter estimates. A prior distribution essentially augments the estimation sample Y by artificial observations Y^* such that the model is estimated based on the combined sample (Y,Y^*).

Prior distributions can also be used to "regularize" the likelihood function by giving the posterior density a more elliptical shape. Finally, a prior distribution can be used to add substantive information about model parameters not contained in the estimation sample θ to the inference problem. Bayesian estimation of DSGE models uses prior distributions mostly to add information contained in data sets other than Y and to smooth out the likelihood function, down-weighing regions of the parameter space in which implications of the structural model contradict nonsample information and the model becomes implausible. An example would be a DSGE model with a likelihood that has a local maximum at which the discount factor is, say, $\beta = 0.5$. Such a value of β would strongly contradict observations of real interest rates. A prior distribution that implies that real interest rates are between 0% and 10% with high probability would squash the undesirable local maximum of the likelihood function.

To the extent that the prior distribution is "informative" and affects the shape of the posterior distribution, it is important that the specification of the prior distribution be carefully documented. Del Negro and Schorfheide (2008) developed a procedure to construct prior distributions based on information contained in presamples or in time series

that are not directly used for the estimation of the DSGE model. To facilitate the elicitation of a prior distribution it is useful to distinguish three groups of parameters: steady-state-related parameters, exogenous shock parameters, and endogenous propagation parameters.

In the context of the stylized DSGE model, the steady-state-related parameters are given by β (real interest rate), π^* (inflation), γ (output growth rate), and λ (labor share). A prior for these parameters could be informed by presample averages of these series. The endogenous propagation parameters are ζ_p (Calvo probability of not being able to reoptimize price) and ν (determines the labor supply elasticity). Micro-level information about the frequency of price changes and labor supply elasticities can be used to specify a prior distribution for these two parameters. Finally, the exogenous shock parameters are the autocorrelation parameters ρ and the shock standard deviations σ.

Because the exogenous shocks are latent, it is difficult to specify a prior distribution for these parameters directly. However, it is possible to map beliefs about the persistence and volatility of observables such as output growth, inflation, and interest rates into beliefs about the exogenous shock parameters. This can be done using the formal procedure described in Del Negro and Schorfheide (2008) or, informally, by generating draws of θ from the prior distribution, simulating artificial observations from the DSGE model, and computing the implied sample moments of the observables. If the prior predictive distribution of these sample moments appears implausible, say, in view of sample statistics computed from a presample of actual observations, then one can adjust the prior distribution of the exogenous shock parameters and repeat the simulation until a plausible prior is obtained. Table 7 contains an example of a prior distribution for our stylized DSGE model. The joint distribution for θ is typically generated as a product of marginal distributions for the elements (or some transformations thereof) of the vector θ.[aw] In most applications this product of marginals is truncated to ensure that the model has a unique equilibrium.

12.2 Metropolis–Hastings Algorithm

Direct sampling from the posterior distribution of θ is unfortunately not possible. One widely used algorithm to generate draws from $p(\theta|Y)$ is the Metropolis–Hastings (MH) algorithm, which belongs to the class of MCMC algorithms. MCMC algorithms produce a sequence of serially correlated parameter draws θ^i, $i = 1,\dots,N$ with the property that the random variables θ^i converge in distribution to the target posterior distribution, which we abbreviate as

[aw] In high-dimensional parameter spaces it might be desirable to replace some of the θ elements by transformations, eg, steady states, that are more plausibly assumed to be independent. This transformation essentially generates nonzero correlations for the original DSGE model parameters. Alternatively, the method discussed in Del Negro and Schorfheide (2008) also generates correlations between parameters.

Table 7 Prior distribution

Name	Domain	Prior		
		Density	Para (1)	Para (2)
Steady-state-related parameters $\theta_{(ss)}$				
$100(1/\beta - 1)$	$\mathbb{R}^+$	Gamma	0.50	0.50
$100\log\pi^*$	$\mathbb{R}^+$	Gamma	1.00	0.50
$100\log\gamma$	$\mathbb{R}$	Normal	0.75	0.50
λ	$\mathbb{R}^+$	Gamma	0.20	0.20
Endogenous propagation parameters $\theta_{(endo)}$				
ζ_p	$[0,1]$	Beta	0.70	0.15
$1/(1 + \nu)$	$\mathbb{R}^+$	Gamma	1.50	0.75
Exogenous shock parameters $\theta_{(exo)}$				
ρ_ϕ	$[0,1)$	Uniform	0.00	1.00
ρ_λ	$[0,1)$	Uniform	0.00	1.00
ρ_z	$[0,1)$	Uniform	0.00	1.00
$100\sigma_\phi$	$\mathbb{R}^+$	InvGamma	2.00	4.00
$100\sigma_\lambda$	$\mathbb{R}^+$	InvGamma	0.50	4.00
$100\sigma_z$	$\mathbb{R}^+$	InvGamma	2.00	4.00
$100\sigma_r$	$\mathbb{R}^+$	InvGamma	0.50	4.00

Notes: Marginal prior distributions for each DSGE model parameter. Para (1) and Para (2) list the means and the standard deviations for Beta, Gamma, and Normal distributions; the upper and lower bound of the support for the Uniform distribution; s and ν for the Inverse Gamma distribution, where $p_{\mathcal{IG}}(\sigma|\nu,s) \propto \sigma^{-\nu-1}e^{-\nu s^2/2\sigma^2}$. The joint prior distribution of θ is truncated at the boundary of the determinacy region.

$$\pi(\theta) = p(\theta|Y) = \frac{p(Y|\theta)p(\theta)}{p(Y)}, \tag{210}$$

as $N \to \infty$. More important, under suitable regularity conditions sample averages of draws converge to posterior expectations:

$$\frac{1}{N - N_0} \sum_{i=N_0+1}^{N} h(\theta^i) \xrightarrow{a.s.} \mathbb{E}_\pi[h(\theta)]. \tag{211}$$

Underlying this convergence result is the fact that the algorithm generates a Markov transition kernel $K(\theta^i|\theta^{i-1})$, characterizing the distribution of θ^i conditional on θ^{i-1}, with the invariance property

$$\int K(\theta^i|\theta^{i-1})\pi(\theta^{i-1})d\theta^{i-1} = \pi(\theta^i). \tag{212}$$

Thus, if θ^{i-1} is a draw from the posterior distribution, then so is θ^i. Of course, this invariance property is not sufficient to guarantee the convergence of the θ^i draws. Chib and

Greenberg (1995) provide an excellent introduction to MH algorithms and detailed text-book treatments can be found, for instance, in Robert and Casella (2004) and Geweke (2005).

12.2.1 The Basic MH Algorithm

The key ingredient of the MH algorithm is a proposal distribution $q(\vartheta|\theta^{i-1})$, which potentially depends on the draw θ^{i-1} in iteration $i-1$ of the algorithm. With probability $\alpha(\vartheta|\theta^{i-1})$ the proposed draw is accepted and $\theta^i = \vartheta$. If the proposed draw is not accepted, then the chain does not move and $\theta^i = \theta^{i-1}$. The acceptance probability is chosen to ensure that the distribution of the draws converges to the target posterior distribution. The algorithm takes the following form:

Algorithm 7 (Generic MH Algorithm). For $i = 1$ to N:
1. Draw ϑ from a density $q(\vartheta|\theta^{i-1})$.
2. Set $\theta^i = \vartheta$ with probability

$$\alpha(\vartheta|\theta^{i-1}) = \min\left\{1, \frac{p(Y|\vartheta)p(\vartheta)/q(\vartheta|\theta^{i-1})}{p(Y|\theta^{i-1})p(\theta^{i-1}))/q(\theta^{i-1}|\vartheta)}\right\}$$

and $\theta^i = \theta^{i-1}$ otherwise.

Because $p(\theta|Y) \propto p(Y|\theta)p(\theta)$ we can replace the posterior densities in the calculation of the acceptance probabilities $\alpha(\vartheta|\theta^{i-1})$ with the product of the likelihood and prior, which does not require the evaluation of the marginal data density $p(Y)$.

12.2.2 Random-Walk Metropolis–Hastings Algorithm

The most widely used MH algorithm for DSGE model applications is the *random walk MH* (RWMH) algorithm. The basic version of this algorithm uses a normal distribution centered at the previous θ^i draw as the proposal density:

$$\vartheta|\theta^i \sim N\left(\theta^i, c^2\hat{\Sigma}\right). \tag{213}$$

Given the symmetric nature of the proposal distribution, the acceptance probability becomes

$$\alpha = \min\left\{\frac{p(\vartheta|Y)}{p(\theta^{i-1}|Y)}, 1\right\}.$$

A draw, ϑ, is accepted with probability one if the posterior at ϑ has a higher value than the posterior at θ^{i-1}. The probability of acceptance decreases as the posterior at the candidate value decreases relative to the current posterior.

To implement the RWMH, the user needs to specify c, and $\hat{\Sigma}$. The proposal variance controls the relative variances and correlations in the proposal distribution. The sampler can work very poorly if q is strongly at odds with the target distribution. A good choice

for $\hat{\Sigma}$ seeks to incorporate information from the posterior, to potentially capture the *a posteriori* correlations among parameters. Obtaining this information can be difficult. A popular approach, used in Schorfheide (2000), is to set $\hat{\Sigma}$ to be the negative of the inverse Hessian at the mode of the log posterior, $\hat{\theta}$, obtained by running a numerical optimization routine before running the MCMC algorithm. Using this as an estimate for the covariance of the posterior is attractive, because it can be viewed as a large sample approximation to the posterior covariance matrix.

Unfortunately, in many applications, the maximization of the posterior density is tedious and the numerical approximation of the Hessian may be inaccurate. These problems may arise if the posterior distribution is very nonelliptical and possibly multimodal, or if the likelihood function is replaced by a nondifferentiable particle filter approximation. In both cases, a (partially) adaptive approach may work well: First, generate a set of posterior draws based on a reasonable initial choice for $\hat{\Sigma}$, eg, the prior covariance matrix. Second, compute the sample covariance matrix from the first sequence of posterior draws and use it as $\hat{\Sigma}$ in a second run of the RWMH algorithm. In principle, the covariance matrix $\hat{\Sigma}$ can be adjusted more than once. However, $\hat{\Sigma}$ must be fixed eventually to guarantee the convergence of the posterior simulator. Samplers that constantly (or automatically) adjust $\hat{\Sigma}$ are known as adaptive samplers and require substantially more elaborate theoretical justifications.

12.2.3 Numerical Illustration

We generate a single sample of size $T = 80$ from the stylized DSGE model using the parameterization in Table 5. The DSGE model likelihood function is combined with the prior distribution in Table 7 to form a posterior distribution. Draws from this posterior distribution are generated using the RWMH described in the previous section. The chain is initialized with a draw from the prior distribution. The covariance matrix $\hat{\Sigma}$ is based on the negative inverse Hessian at the mode. The scaling constant c is set equal to 0.075, which leads to an acceptance rate for proposed draws of 0.55.

The top panels of Fig. 32 depict the sequences of posterior draws of the Calvo parameter ζ_p^i and preference shock standard deviation σ_ϕ^i. It is apparent from the figure that the draws are serially correlated. The draws for the standard deviation are strongly contaminated by the initialization of the chain, but they eventually settle to a range of 0.8–1.1. The bottom panel depicts recursive means of the form

$$\bar{h}_{N|N_0} = \frac{1}{N - N_0} \sum_{i=N_0 + 1}^{N} h(\theta^i). \tag{214}$$

To remove the effect of the initialization of the Markov chain, it is common to drop the first N_0 draws from the computation of the posterior mean approximation. In the figure we set $N_0 = 7500$ and $N = 37,500$. Both recursive means eventually settle to a limit point.
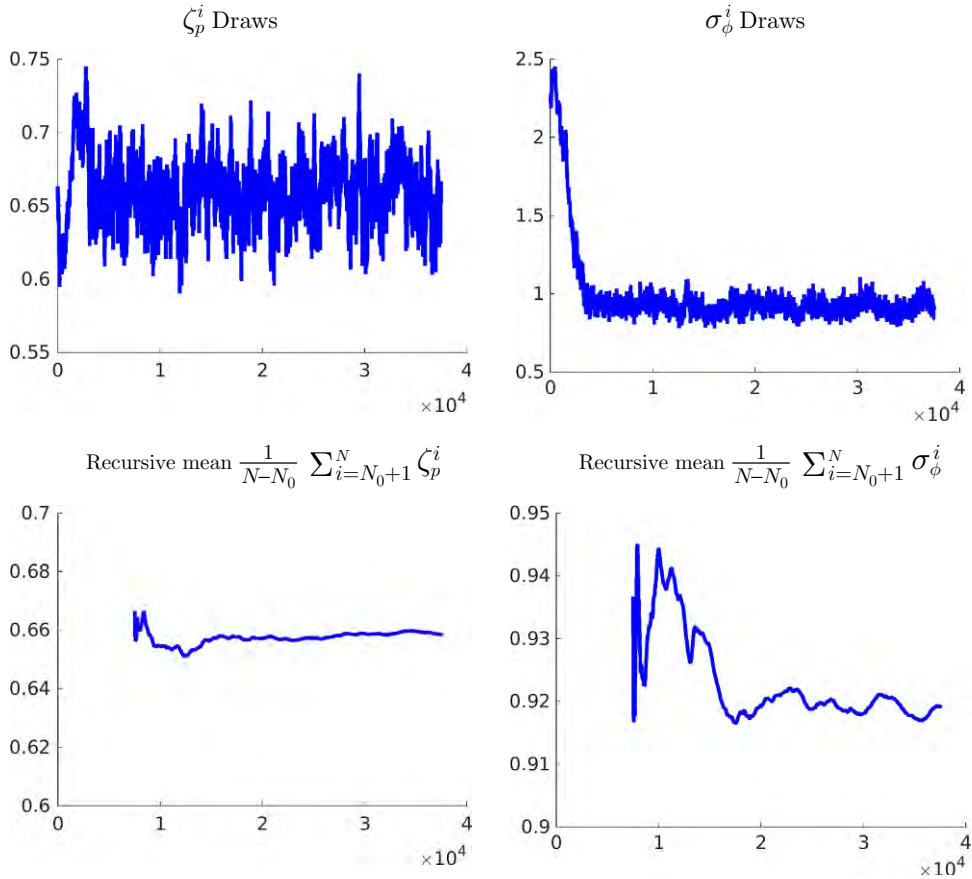
Fig. 32 Parameter draws from MH algorithm. *Notes*: The posterior is based on a simulated sample of observations of size $T = 80$. The *top panel* shows the sequence of parameter draws and the *bottom panel* shows recursive means.

The output of the algorithm is stochastic, which implies that running the algorithm repeatedly will generate different numerical results. Under suitable regularity conditions the recursive means satisfy a CLT. The easiest way to obtain a measure of numerical accuracy is to run the RWMH algorithm, say, fifty times using random starting points, and compute the sample variance of $\bar{h}_{N|N_0}$ across chains. Alternatively, one could compute a heteroskedasticity and autocorrelation consistent (HAC) standard error estimate for $\bar{h}_{N|N_0}$ based on the output of a single chain.

Fig. 33 depicts univariate prior and posterior densities, which are obtained by applying a standard kernel density estimator to draws from the prior and posterior distribution. In addition, one can also compute posterior credible sets based on the output of the
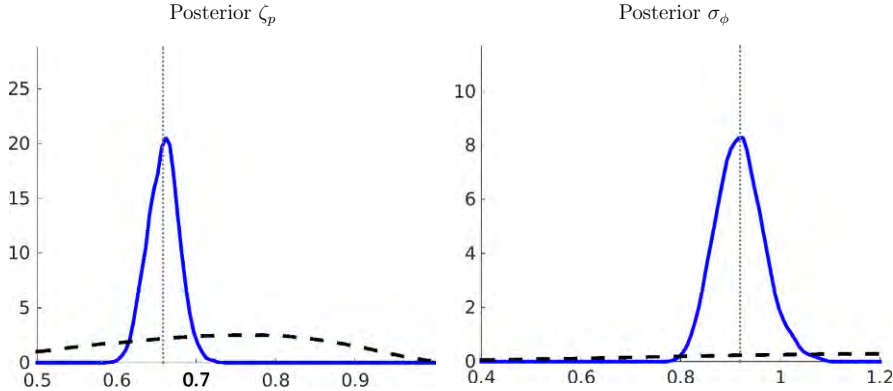
Fig. 33 Prior and posterior densities. *Notes:* The *dashed lines* represent the prior densities, whereas the *solid lines* correspond to the posterior densities of ζ_p and σ_ϕ. The posterior is based on a simulated sample of observations of size $T = 80$. We generate $N = 37,500$ draws from the posterior and drop the first $N_0 = 7,500$ draws.

posterior sampler. For a univariate parameter, the shortest credible set is given by the highest–posterior–density (HPD) set defined as

$$CS_{HPD}(Y) = \{\theta \mid p(\theta|Y) \geq \kappa_\alpha\}, \tag{215}$$

where κ_α is chosen to ensure that the credible set has the desired posterior coverage probability.

12.2.4 Blocking

Despite a careful choice of the proposal distribution $q(\cdot|\theta^{i-1})$, it is natural that the efficiency of the MH algorithm decreases as the dimension of the parameter vector θ increases. The success of the proposed random walk move decreases as the dimension d of the parameter space increases. One way to alleviate this problem is to break the parameter vector into blocks. Suppose the dimension of the parameter vector θ is d. A partition of the parameter space, B, is a collection of N_{blocks} sets of indices. These sets are mutually exclusive and collectively exhaustive. Call the subvectors that correspond to the index sets θ_b, $b = 1,\ldots,N_{blocks}$. In the context of a sequence of parameter draws, let θ_b^i refer to the b-th block of i-th draw of θ and let $\theta_{<b}^i$ refer to the i-th draw of all of the blocks before b and similarly for $\theta_{>b}^i$. Algorithm 8 describes a generic Block MH algorithm.

Algorithm 8 (Block MH Algorithm). Draw $\theta^0 \in \Theta$ and then for $i = 1$ to N:
1. Create a partition B^i of the parameter vector into N_{blocks} blocks $\theta_1,\ldots,\theta_{N_{blocks}}$ via some rule (perhaps probabilistic), unrelated to the current state of the Markov chain.
2. For $b = 1,\ldots,N_{blocks}$:

(a) Draw $\vartheta_b \sim q(\cdot|[\theta^i_{<b}, \theta^{i-1}_b, \theta^{i-1}_{\geq b}])$.

(b) With probability,

$$\alpha = \max\left\{\frac{p([\theta^i_{<b}, \vartheta_b, \theta^{i-1}_{>b}]|Y)q(\theta^{i-1}_b, |\theta^i_{<b}, \vartheta_b, \theta^{i-1}_{>b})}{p(\theta^i_{<b}, \theta^{i-1}_b, \theta^{i-1}_{>b}|Y)q(\vartheta_b|\theta^i_{<b}, \theta^{i-1}_b, \theta^{i-1}_{>b})}, 1\right\},$$

set $\theta^i_b = \vartheta_b$, otherwise set $\theta^i_b = \theta^{i-1}_b$.

In order to make the Block MH algorithm operational, the researcher has to decide how to allocate parameters to blocks in each iteration and how to choose the proposal distribution $q(\cdot|[\theta^i_{<b}, \theta^{i-1}_b, \theta^{i-1}_{>b}])$ for parameters of block b.

A good rule of thumb, however, is that we want the parameters *within* a block, say, θ^b, to be as correlated as possible, while we want the parameters between blocks, say, θ_b and θ_{-b}, to be as independent as possible, according to Robert and Casella (2004). Unfortunately, picking the "optimal" blocks to minimize dependence across blocks requires *a priori* knowledge about the posterior and is therefore often infeasible. Chib and Ramamurthy (2010) propose grouping parameters randomly. Essentially, the user specifies how many blocks to partition the parameter vector into and every iteration a new set of blocks is constructed. Key to the algorithm is that the block configuration be independent of the Markov chain. This is crucial for ensuring the convergence of the chain.

In order to tailor the block-specific proposal distributions, Chib and Ramamurthy (2010) advocate using an optimization routine—specifically, simulated annealing—to find the mode of the conditional posterior distribution. As in the RWMH-V algorithm, the variance of the proposal distribution is based on the inverse Hessian of the conditional log posterior density evaluated at the mode. Unfortunately, the tailoring requires many likelihood evaluations that slow down the algorithm and a simpler procedure, such as using marginal or conditional covariance matrices from an initial approximation of the joint posterior covariance matrix, might be computationally more efficient.

12.2.5 Marginal Likelihood Approximations

The computations thus far do not rely on the marginal likelihood $p(Y)$, which appears in the denominator of Bayes Theorem. Marginal likelihoods play an important role in assessing the relative fit of models because they are used to turn prior model probabilities into posterior probabilities. The most widely used marginal likelihood approximation in the DSGE model literature is the modified harmonic mean estimator proposed by Geweke (1999). This estimator is based on the identity

$$\int \frac{f(\theta)}{p(Y)}d\theta = \int \frac{f(\theta)}{p(Y|\theta)p(\theta)}p(\theta|Y)d\theta, \tag{216}$$

where $f(\theta)$ has the property that $\int f(\theta)d\theta = 1$. The identity is obtained by rewriting Bayes Theorem, multiplying both sides with $f(\theta)$ and integrating over θ. Realizing that the

left-hand side simplifies to $1/p(Y)$ and that the right-hand side can be approximated by a Monte Carlo average we obtain

$$\hat{p}_{HM}(Y) = \left[\frac{1}{N} \sum_{i=1}^{N} \frac{f(\theta^i)}{p(Y|\theta^i)p(\theta^i)} \right]^{-1}, \qquad (217)$$

where the θ_i's are drawn from the posterior $p(\theta|Y)$. The function $f(\theta)$ should be chosen to keep the variance of $f(\theta^i)/p(Y|\theta^i)p(\theta^i)$ small. Geweke (1999) recommends using for $f(\theta)$ a truncated normal approximation of the posterior distribution for θ that is computed from the output of the posterior sampler. Alternative methods to approximate the marginal likelihood are discussed in Chib and Jeliazkov (2001), Sims et al. (2008), and Ardia et al. (2012). An and Schorfheide (2007) and Herbst and Schorfheide (2015) provide accuracy comparisons of alternative methods.

12.2.6 Extensions

The basic estimation approach for linearized DSGE models has been extended in several dimensions. Typically, the parameter space is restricted to a subspace in which a linearized model has a unique nonexplosive rational expectations solution (determinacy). Lubik and Schorfheide (2004) relax this restriction and also consider the region of the parameter space in which the solution is indeterminate. By computing the posterior probability of parameter values associated with indeterminacy, they are able to conduct a posterior odds assessment of determinacy vs indeterminacy. Justiniano and Primiceri (2008) consider a linearized DSGE model with structural shocks that exhibit stochastic volatility and develop an MCMC algorithm for posterior inference. A further extension is provided by Curdia et al. (2014), who also allow for shocks that, conditional on the volatility process, have a fat-tailed student-t distribution to capture extreme events such as the Great Recession. Schorfheide (2005a) and Bianchi (2013) consider the estimation of linearized DSGE models with regime switching in the coefficients of the state-space representation.

Müller (2012) provides an elegant procedure to assess the robustness of posterior inference to shifts in the mean of the prior distribution. One of the attractive features of his procedure is that the robustness checks can be carried out without having to reestimate the DSGE model under alternative prior distributions. Koop et al. (2013) propose some diagnostics that allow users to determine the extent to which the likelihood function is informative about the DSGE model parameters. In a nutshell, the authors recommend examining whether the variance of marginal posterior distributions shrinks at the rate T^{-1} (in a stationary model) if the number of observations is increased in a simulation experiment.

12.2.7 Particle MCMC

We now turn to the estimation of fully nonlinear DSGE models. As discussed in Section 10, for nonlinear DSGE models the likelihood function has to be approximated

by a nonlinear filter. Embedding a particle filter approximation into an MCMC sampler leads to a so-called particle MCMC algorithm. We refer to the combination of a particle-filter approximated likelihood and the MH algorithm as a PFMH algorithm. This idea was first proposed for the estimation of nonlinear DSGE models by Fernández-Villaverde and Rubio-Ramírez (2007). The theory underlying the PFMH algorithm is developed in Andrieu et al. (2010). Flury and Shephard (2011) discuss non-DSGE applications of particle MCMC methods in econometrics. The modification of Algorithm 7 is surprisingly simple: one only has to replace the exact likelihood function $p(Y|\theta)$ with the particle filter approximation $\hat{p}(Y|\theta)$.

Algorithm 9 (PFMH Algorithm). For $i = 1$ to N:
1. Draw ϑ from a density $q(\vartheta|\theta^{i-1})$.
2. Set $\theta^i = \vartheta$ with probability

$$\alpha(\vartheta|\theta^{i-1}) = \min\left\{1, \; \frac{\hat{p}(Y|\vartheta)p(\vartheta)/q(\vartheta|\theta^{i-1})}{\hat{p}(Y|\theta^{i-1})p(\theta^{i-1})/q(\theta^{i-1}|\vartheta)}\right\}$$

and $\theta^i = \theta^{i-1}$ otherwise. The likelihood approximation $\hat{p}(Y|\vartheta)$ is computed using Algorithm 6.

The surprising implication of the theory developed in Andrieu et al. (2010) is that the distribution of draws generated by Algorithm 9 from the PFMH algorithm that replaces $p(Y|\theta)$ with $\hat{p}(Y|\theta)$ in fact does converge to the exact posterior. The replacement of the exact likelihood function by the particle-filter approximation generally increases the persistence of the Markov chain and makes Monte Carlo approximations less accurate; see Herbst and Schorfheide (2015) for numerical illustrations. Formally, the key requirement is that the particle-filter approximation provide an unbiased estimate of the likelihood function. In practice it has to be ensured that the variance of the numerical approximation is small relative to the expected magnitude of the differential between $p(Y|\theta^{i-1})$ and $p(Y|\vartheta)$ in an ideal version of the algorithm in which the likelihood could be evaluated exactly. Thus, before embedding the particle-filter approximation into a likelihood function, it is important to assess its accuracy for low- and high-likelihood parameter values.

12.3 SMC Methods

Sequential Monte Carlo (SMC) techniques to generate draws from posterior distributions of a static parameter θ are emerging as an attractive alternative to MCMC methods. SMC algorithms can be easily parallelized and, properly tuned, may produce more accurate approximations of posterior distributions than MCMC algorithms. Chopin (2002) showed how to adapt the particle filtering techniques discussed in Section 10.3 to conduct posterior inference for a static parameter vector. Textbook treatments of SMC algorithms can be found, for instance, in Liu (2001) and Cappé et al. (2005).

The first paper that applied SMC techniques to posterior inference in a small-scale DSGE models was Creal (2007). Herbst and Schorfheide (2014) develop the algorithm further, provide some convergence results for an adaptive version of the algorithm building on the theoretical analysis of Chopin (2004), and show that a properly tailored SMC algorithm delivers more reliable posterior inference for large-scale DSGE models with a multimodal posterior than the widely used RWMH-V algorithm. Creal (2012) provides a recent survey of SMC applications in econometrics. Durham and Geweke (2014) show how to parallelize a flexible and self-tuning SMC algorithm for the estimation of time series models on graphical processing units (GPU). The remainder of this section draws heavily from the more detailed exposition in Herbst and Schorfheide (2014, 2015).

SMC combines features of classic importance sampling and modern MCMC techniques. The starting point is the creation of a sequence of intermediate or bridge distributions $\{\pi_n(\theta)\}_{n=0}^{N_\phi}$ that converge to the target posterior distribution, ie, $\pi_{N_\phi}(\theta) = \pi(\theta)$. At any stage the posterior distribution $\pi_n(\theta)$ is represented by a swarm of particles $\{\theta_n^i, W_n^i\}_{i=1}^N$ in the sense that the Monte Carlo average

$$\bar{h}_{n,N} = \frac{1}{N} \sum_{i=1}^N W_n^i h(\theta_n^i) \overset{a.s.}{\to} \mathbb{E}_{\pi_n}[h(\theta_n)]. \tag{218}$$

The bridge distributions can be generated either by taking power transformations of the entire likelihood function, that is, $[p(Y|\theta)]^{\phi_n}$, where $\phi_n \uparrow 1$, or by adding observations to the likelihood function, that is, $p(Y_{1:t_n}|\theta)$, where $t_n \uparrow T$. We refer to the first approach as likelihood tempering and the second approach as data tempering. Formally, the sequences of bridge distributions are defined as (likelihood tempering)

$$\pi_n(\theta) = \frac{[p(Y|\theta)]^{\phi_n} p(\theta)}{\int [p(Y|\theta)]^{\phi_n} p(\theta) d\theta} \quad n = 0, \ldots, N_\phi, \quad \phi_n \uparrow 1, \tag{219}$$

and (data tempering, writing $t_n = \lfloor \phi_n T \rfloor$)

$$\pi_n^{(D)}(\theta) = \frac{p(Y_{1:\lfloor \phi_n T \rfloor}) p(\theta)}{\int p(Y_{1:\lfloor \phi_n T \rfloor}) p(\theta) d\theta} \quad n = 0, \ldots, N_\phi, \quad \phi_n \uparrow 1, \tag{220}$$

respectively. While data tempering is attractive in sequential applications, eg, real-time forecasting, likelihood tempering generally leads to more stable posterior simulators for two reasons: First, in the initial phase it is possible to add information that corresponds to a fraction of an observation. Second, if the latter part of the sample contains influential observations that drastically shift the posterior mass, data tempering may have difficulties adapting to the new information.

12.3.1 The SMC Algorithm

The algorithm can be initialized with draws from the prior density $p(\theta)$, provided the prior density is proper. For the prior in Table 7 it is possible to directly sample independent draws

θ_0^i from the marginal distributions of the DSGE model parameters. One can add an accept-reject step that eliminates parameter draws for which the linearized model does not have a unique stable rational expectations solution. The initial weights W_0^i can be set equal to one. We adopt the convention that the weights are normalized to sum to N.

The SMC algorithm proceeds iteratively from $n = 0$ to $n = N_\phi$. Starting from stage $n - 1$ particles $\{\theta_{n-1}^i, W_{n-1}^i\}_{i=1}^N$ each stage n of the algorithm consists of three steps: *correction*, that is, reweighting the stage $n - 1$ particles to reflect the density in iteration n; *selection*, that is, eliminating a highly uneven distribution of particle weights (degeneracy) by resampling the particles; and *mutation*, that is, propagating the particles forward using a Markov transition kernel to adapt the particle values to the stage n bridge density.

Algorithm 10 (Generic SMC Algorithm with Likelihood Tempering).

1. **Initialization.** $(\phi_0 = 0)$. Draw the initial particles from the prior: $\theta_0^i \overset{iidp}{\sim} p(\theta)$ and $W_0^i = 1$, $i = 1,\ldots,N$.

2. **Recursion.** For $n = 1,\ldots,N_\phi$,

   (a) Correction. Reweight the particles from stage $n - 1$ by defining the incremental weights

   $$\tilde{w}_n^i = [p(Y|\theta_{n-1}^i)]^{\phi_n - \phi_{n-1}} \tag{221}$$

   and the normalized weights

   $$\tilde{W}_n^i = \frac{\tilde{w}_n^i W_{n-1}^i}{\frac{1}{N}\sum_{i=1}^N \tilde{w}_n^i W_{n-1}^i}, \qquad i = 1,\ldots,N. \tag{222}$$

   (b) Selection (Optional). Resample the particles via multinomial resampling. Let $\{\hat{\theta}^i\}_{i=1}^N$ denote N *iid* draws from a multinomial distribution characterized by support points and weights $\{\theta_{n-1}^i, \tilde{W}_n^i\}_{i=1}^N$ and set $W_n^i = 1$.

   (c) Mutation. Propagate the particles $\{\hat{\theta}_i, W_n^i\}$ via N_{MH} steps of an MH algorithm with transition density $\theta_n^i \sim K_n(\theta_n|\hat{\theta}_n^i; \zeta_n)$ and stationary distribution $\pi_n(\theta)$. An approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$ is given by

   $$\bar{h}_{n,N} = \frac{1}{N}\sum_{i=1}^N h(\theta_n^i) W_n^i. \tag{223}$$

3. For $n = N_\phi$ $(\phi_{N_\phi} = 1)$ the final importance sampling approximation of $\mathbb{E}_\pi[h(\theta)]$ is given by:

   $$\bar{h}_{N_\phi,N} = \sum_{i=1}^N h(\theta_{N_\phi}^i) W_{N_\phi}^i. \tag{224}$$

The correction step is a classic importance sampling step, in which the particle weights are updated to reflect the stage n distribution $\pi_n(\theta)$. Because this step does not change the particle value, it is typically not necessary to reevaluate the likelihood function.

The selection step is optional. On the one hand, resampling adds noise to the Monte Carlo approximation, which is undesirable. On the other hand, it equalizes the particle weights, which increases the accuracy of subsequent importance sampling approximations. The decision of whether or not to resample is typically based on a threshold rule for the variance of the particle weights. As for the particle filter in Section 10.3, we can define an effective particle sample size as:

$$\widehat{ESS}_n = N \Big/ \left(\frac{1}{N} \sum_{i=1}^{N} (\tilde{W}_n^i)^2 \right) \tag{225}$$

and resample whenever $\widehat{ESS}_n$ is less that $N/2$ or $N/4$. In the description of Algorithm 10 we consider multinomial resampling. Other, more efficient resampling schemes are discussed, for instance, in the books by Liu (2001) or Cappé et al. (2005) (and references cited therein).

The mutation step changes the particle values. In the absence of the mutation step, the particle values would be restricted to the set of values drawn in the initial stage from the prior distribution. This would clearly be inefficient, because the prior distribution is a poor proposal distribution for the posterior in an importance sampling algorithm. As the algorithm cycles through the N_ϕ phases, the particle values successively adapt to the shape of the posterior distribution. The key feature of the transition kernel $K_n(\theta_n | \hat{\theta}_n; \zeta_n)$ is the invariance property:

$$\pi_n(\theta_n) = \int K_n(\theta_n | \hat{\theta}_n; \zeta_n) \pi_n(\hat{\theta}_n) d\hat{\theta}_n. \tag{226}$$

Thus, if $\hat{\theta}_n^i$ is a draw from π_n, then so is θ_n^i. The mutation step can be implemented by using one or more steps of the RWMH algorithm described in Section 12.2.2. The probability of mutating the particles can be increased by blocking or by iterating the RWMH algorithm over multiple steps. The vector ζ_n summarizes the tuning parameters, eg, c and $\hat{\Sigma}$ of the RWMH algorithm.

The SMC algorithm produces as a by-product an approximation of the marginal likelihood. It can be shown that

$$\hat{p}_{SMC}(Y) = \prod_{n=1}^{N_\phi} \left(\frac{1}{N} \sum_{i=1}^{N} \tilde{w}_n^i W_{n-1}^i \right)$$

converges almost surely to $p(Y)$ as the number of particles $N \to \infty$.

12.3.2 Tuning the SMC Algorithm

The implementation of the SMC algorithm requires the choice of several tuning constants. The most important choice is the number of particles N. As shown in Chopin (2004), Monte Carlo averages computed from the output of the SMC algorithm satisfy

a CLT as the number of particles increases to infinity. This means that the variance of the Monte Carlo approximation decreases at the rate $1/N$. The user has to determine the number of bridge distributions N_ϕ and the tempering schedule ϕ_n. Based on experiments with a small–scale DSGE model, Herbst and Schorfheide (2015) recommend a convex tempering schedule of the form $\phi_n = (n/N_\phi)^\lambda$ with $\lambda \approx 2$. Durham and Geweke (2014) recently developed a self-tuning algorithm that chooses the sequence ϕ_n adaptively as the algorithm cycles through the stages.

The mutation step requires the user to determine the number of MH steps N_{MH} and the number of parameter blocks. The increased probability of mutation raises the accuracy but unfortunately, the number of likelihood evaluations increases as well, which slows down the algorithm. The scaling constant c and the covariance matrix $\hat{\Sigma}$ can be easily chosen adaptively. Based on the MH rejection frequency, c can be adjusted to achieve a target rejection rate of approximately 25–40%. For $\hat{\Sigma}_n$ one can use an approximation of the posterior covariance matrix computed at the end of the stage n correction step.

To monitor the accuracy of the SMC approximations Durham and Geweke (2014) suggest creating H groups of N particles and setting up the algorithm so that there is no communication across groups. This leads to H Monte Carlo approximations of posterior moments of interest. The across–group standard deviation of within–group Monte Carlo averages provides a measure of numerical accuracy. Parallelization of the SMC algorithm is relatively straightforward because the mutation step and the computation of the incremental weights in the correction step can be carried out in parallel on multiple processors, each of which is assigned a group of particles. In principle, the exact likelihood function can be replaced by a particle-filter approximation, which leads to an SMC^2 algorithm, developed by Chopin et al. (2012) and discussed in more detail in the context of DSGE models in Herbst and Schorfheide (2015).

12.3.3 Numerical Illustration

We now illustrate the SMC model in the context of the stylized DSGE models. The setup is similar to the one in Section 12.2.3. We generate $T = 80$ observations using the parameters listed in Table 5 and use the prior distribution given in Table 7. The algorithm is configured as follows. We use $N = 2048$ particles and $N_\phi = 500$ tempering stages. We set $\lambda = 3$, meaning that we add very little information in the initial stages to ensure that the prior draws adapt to the shape of the posterior. We use one step of a single-block RWMH algorithm in the mutation step and choose c and $\hat{\Sigma}_n$ adaptively as described in Herbst and Schorfheide (2014). The target acceptance rate for the mutation step is 0.25. Based on the output of the SMC algorithm, we plot marginal bridge densities $\pi_n(\cdot)$ for the price stickiness parameter ζ_p and the shock standard deviation σ_ϕ in Fig. 34. The initial set of particles is drawn from the prior distribution. As ϕ_n increases to one, the distribution concentrates. The final stage approximates the posterior distribution.
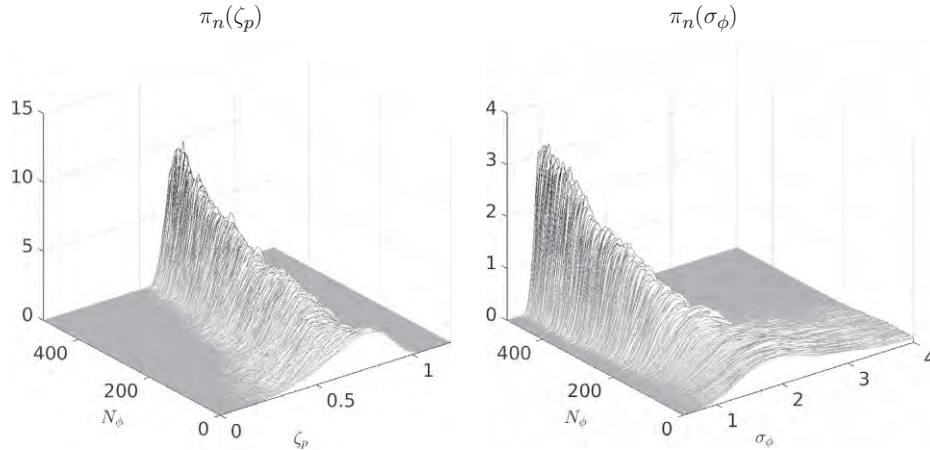
Fig. 34 SMC bridge densities. *Notes*: The posterior is based on a simulated sample of observations of size $T = 80$. The two panels show the sequence of posterior (bridge) densities $\pi_n(\cdot)$.

12.4 Model Diagnostics

DSGE models provide stylized representations of the macroeconomy. To examine whether a specific model is able to capture salient features of the data Y from an *a priori* perspective, prior predictive checks provide an attractive diagnostic. Prior (and posterior) predictive checks are discussed in general terms in the textbooks by Lancaster (2004) and Geweke (2005). The first application of a prior predictive check in the context of DSGE models is Canova (1994).

Let $Y^*_{1:T}$ be an artificial sample of length T. The predictive distribution for $Y^*_{1:T}$ based on the time t information set $\mathcal{F}_t$ is

$$p(Y^*_{1:T}|\mathcal{F}_t) = \int p(Y^*_{1:T}|\theta)p(\theta|\mathcal{F}_t)d\theta. \tag{227}$$

We used a slightly more general notation (to accommodate posterior predictive checks below) with the convention that $\mathcal{F}_0$ corresponds to prior information. The idea of a predictive check is to examine how far the actual realization $Y_{1:T}$ falls into the tail of the predictive distribution. If $Y_{1:T}$ corresponds to an unlikely tail event, then the model is regarded as poorly specified and should be adjusted before it is estimated.

In practice, the high-dimensional vector $Y_{1:T}$ is replaced by a lower-dimensional statistic $\mathcal{S}(Y_{1:T})$, eg, elements of the sample autocovariance matrix $vech(\hat{\Gamma}_{yy}(h))$, for which it is easier to calculate or visualize tail probabilities. While it is not possible to directly evaluate the predictive density of sample statistics, it is straightforward to generate draws. In the case of a prior predictive check, let $\{\theta^i\}^N_{i=1}$ be a sequence of parameter draws from the prior. For each draw, simulate the DSGE model, which leads to the trajectory

$Y_{1:T}^{*i}$. For each of the simulated trajectories, compute the sample statistic $\mathcal{S}(\cdot)$, which leads to a draw from the predictive density.

For a posterior predictive check one equates $\mathcal{F}_t$ with the sample $Y_{1:T}$. The posterior predictive check examines whether the estimated DSGE model captures the salient features of the sample. A DSGE model application can be found in Chang et al. (2007), who examine whether versions of an estimated stochastic growth model are able to capture the variance and the serial correlation of hours worked.

12.5 Limited Information Bayesian Inference

Bayesian inference requires a likelihood function $p(Y|\theta)$. However, as discussed in Section 11, many of the classical approaches to DSGE model estimation, eg, (generalized) methods of moments and impulse response function matching, do not utilize the likelihood function of the DSGE model, in part because there is some concern about misspecification. These methods are referred to as limited-information (instead of full information) techniques. This subsection provides a brief survey of Bayesian approaches to limited-information inference.

12.5.1 Single-Equation Estimation

Lubik and Schorfheide (2005) estimate monetary policy rules for small open economy models by augmenting the policy rule equation with a vector-autoregressive law of motion for the endogenous regressors, eg, the output gap and inflation in the case of our stylized model. This leads to a VAR for output, inflation, and interest rates, with cross-coefficient restrictions that are functions of the monetary policy rule parameters. The restricted VAR can be estimated with standard MCMC techniques. Compared to the estimation of a fully specified DSGE model, the limited-information approach robustifies the estimation of the policy rule equation against misspecification of the private sector's behavior. Kleibergen and Mavroeidis (2014) apply a similar technique to the estimation of a New Keynesian Phillips curve. Their work focuses on the specification of prior distributions that regularize the likelihood function in settings in which the sample only weakly identifies the parameters of interest, eg, the slope of the New Keynesian Phillips curve.

12.5.2 Inverting a Sampling Distribution

Suppose one knows the sampling distribution $p(\hat{\theta}|\theta)$ of an estimator $\hat{\theta}$. Then, instead of updating beliefs conditional on the observed sample Y, one could update the beliefs about θ based on the realization of $\hat{\theta}$:

$$p(\theta|\hat{\theta}) = \frac{p(\hat{\theta}|\theta)p(\theta)}{\int p(\hat{\theta}|\theta)p(\theta)}. \tag{228}$$

This idea dates back at least to Pratt et al. (1965) and is useful in situations in which a variety of different distributions for the sample Y lead to the same distribution of the estimator $\hat{\theta}$. The drawback of this approach is that a closed-form representation of the density $p(\hat{\theta}|\theta)$ is typically not available.

In practice one could use a simulation–based approximation of $p(\hat{\theta}|\theta)$, which is an idea set forth by Boos and Monahan (1986). Alternatively, one could replace the finite-sample distribution with a limit distribution, eg,

$$\sqrt{T}(\hat{\theta}_T - \theta_T)|\theta_T \Longrightarrow N(0, V(\theta)), \tag{229}$$

where the sequence of "true" parameters θ_T converges to θ. This approach is considered by Kwan (1999). In principle $\hat{\theta}_T$ could be any of the frequentist estimators studied in Section 11 for which we derived an asymptotic distribution, including the MD estimator, the IRF matching estimator, or the GMM estimator. However, in order for the resulting limited-information posterior to be meaningful, it is important that the convergence to the asymptotic distribution be uniform in θ, which requires (229) to hold for each sequence $\theta_T \rightarrow \theta$. A uniform convergence to a normal distribution is typically not attainable as θ_T approaches the boundary of the region of the parameter space in which the time series $Y_{1:T}$ is stationary.

Rather than making statements about the approximation of the limited-information posterior distribution $p(\theta|\hat{\theta})$, Müller (2013) adopts a decision-theoretic framework and shows that decisions based on the quasi-posterior that is obtained by inverting the limit distribution of $\hat{\theta}_T|\theta$ are asymptotically optimal (in the sense that they minimize expected loss) under fairly general conditions. Suppose that the likelihood function of a DSGE model is misspecified. In this case the textbook analysis of the ML estimator in Section 11.1 has to be adjusted as follows. The information matrix equality that ensures that $\| -\nabla_\theta^2 \ell_T(\theta|Y) - \mathcal{I}(\theta_0) \|$ converges to zero is no longer satisfied. If we let $D = \text{plim}_{T\rightarrow\infty} -\nabla_\theta^2 \ell_T(\theta|Y)$, then the asymptotic variance of the ML estimator takes the sandwich form $D\mathcal{I}(\theta_0)D'$. Under the limited-information approach coverage sets for individual DSGE model parameters would be computed based on the diagonal elements of $D\mathcal{I}(\theta_0)D'$, whereas under a full-information Bayesian approach with mis-specified likelihood function, the coverage sets would (asymptotically) be based on $\mathcal{I}^{-1}(\theta_0)$. Thus, the limited-information approach robustifies the coverage sets against model misspecification.

Instead of inverting a sampling distribution of an estimator, one could also invert the sampling distribution of some auxiliary sample statistic $\hat{\varphi}(Y)$. Not surprisingly, the main obstacle is the characterization of the distribution $\hat{\varphi}|\theta$. A collection of methods referred to as approximate Bayesian computations (ABC) use a simulation approximation of $p(\hat{\varphi}|\theta)$ and they could be viewed as a Bayesian version of indirect inference. These algorithms target

$$p_{ABC}^{\delta}(\theta, \hat{\varphi}^*|\hat{\varphi}) \propto p(\hat{\varphi}^*|\theta)p(\theta)\mathbb{I}\{\|\hat{\varphi}^* - \hat{\varphi}\| \leq \delta\}, \tag{230}$$

where $\hat{\varphi}$ refers to the auxiliary statistic computed from the observed data, $\hat{\varphi}^*$ is the auxiliary statistic computed from data simulated from the model conditional on a parameter θ, and δ is the level of tolerance for discrepancies between model-simulated and observed statistics. To date, there are few applications of ABC in econometrics. Forneron and Ng (2015) discuss the relationship between ABC and the simulated MD estimators introduced in Section 11.2 and Scalone (2015) explores a DSGE model application.

12.5.3 Limited-Information Likelihood Functions

Kim (2002) constructs a limited-information likelihood function from the objective function of an extremum estimator. For illustrative purposes we consider the GMM estimator discussed in Section 11.4, but the same idea can also be applied to the MD estimator and the IRF matching estimator. Suppose the data are generated under the probability measure $\mathbb{P}$ and at $\theta = \theta_0$ the following GMM moment condition is satisfied: $\mathbb{E}_{\mathbb{P}}[g(y_{t-p:t}|\theta_0)] = 0$. The sample objective function $Q_T(\theta|Y)$ for the resulting GMM estimator based on a weight matrix W was given in (207). Assuming uniform integrability of the sample objective function

$$\lim_{T \to \infty} \mathbb{E}_{\mathbb{P}}[Q_T(\theta_0|Y)] = r \tag{231}$$

where r is the number of overidentifying moment conditions (meaning the difference between the number of moments stacked in the vector $g(\cdot)$ and the number of elements of the parameter vector θ).

Let $\mathcal{P}(\theta)$ denote the collection of probability distributions that satisfy the moment conditions in the following sense:

$$\mathcal{P}(\theta) = \left\{ P \mid \lim_{T \to \infty} \mathbb{E}_P[TQ_T(\theta|Y)] = r \right\}. \tag{232}$$

$\mathcal{P}(\theta)$ cannot be used directly for likelihood-based inference because it comprises a collection of probability distributions indexed by θ. To obtain a unique distribution for each θ, Kim (2002) projects the "true" distribution $\mathbb{P}$ onto the set $\mathcal{P}(\theta)$ using the Kullback–Leibler discrepancy as the metric:

$$P^*(Y|\theta) = \operatorname{argmin}_{P \in \mathcal{P}(\theta)} \int \log(dP/d\mathbb{P})dP. \tag{233}$$

The solution takes the convenient form

$$p^*(Y|\theta) \propto \exp\left\{ -\frac{1}{2} Q_T(\theta|Y) \right\}, \tag{234}$$

where $p^*(Y|\theta) = dP/d\mathbb{P}$ is the Radon–Nikodym derivative of P with respect to $\mathbb{P}$.

Kim's (2002) results suggest that the frequentist objective functions of Sections 11.2–11.4 can be combined with a prior density and used for (limited information) Bayesian inference. The posterior mean

$$\hat{\theta} = \frac{\int \theta \exp\left\{-\frac{1}{2}Q_T(\theta|Y)\right\}p(\theta)d\theta}{\int \exp\left\{-\frac{1}{2}Q_T(\theta|Y)\right\}p(\theta)d\theta} \tag{235}$$

resembles the LT estimator discussed in Section 11.2. The main difference is that the LT estimator was interpreted from a frequentist perspective, whereas the quasi–posterior based on $p^*(Y|\theta)$ and statistics such as the posterior mean are meant to be interpreted from a Bayesian perspective. This idea has been recently exploited by Christiano et al. (2010) to propose a Bayesian IRF matching estimator. An application to an asset pricing model is presented in Gallant (2015) and an extension to models with latent variables is provided in Gallant et al. (2013). Inoue and Shintani (2014) show that the limited information marginal likelihood

$$p^*(Y|M) = \int p^*(Y|\theta, M)p(\theta)d\theta$$

can be used as a model selection criterion that asymptotically is able to select a correct model specification.

12.5.4 Nonparametric Likelihood Functions

There is also a literature on nonparametric likelihood functions that are restricted to satisfy model–implied moment conditions. Lazar (2003) and Schennach (2005) use empirical likelihood functions, which, roughly speaking assign probability p_t to observation y_t such that the likelihood function is written as $\prod_{t=1}^{T} p_t$, at least if the data are *iid*. One then imposes the side constraint $\sum_{t=1}^{T} p_t g(y_{t-p:t}|\theta) = 0$ and concentrates out p_t probabilities to obtain a profile objective function that only depends on θ. This method is designed for *iid* data and possible models in which $g(y_{t-p:t}|\theta)$ is a martingale difference sequence. Kitamura and Otsu (2011) propose to using a Dirichlet process to generate a prior for the distribution of $Y_{1:T}$ and then project this distribution on the set of distributions that satisfies the moment restrictions. Shin (2014) uses a Dirichlet process mixture and provides a time series extension.

13. CONCLUSION

Over the past two decades the development and application of solution and estimation methods for DSGE models have experienced tremendous growth. Part of this growth has been spurred by central banks, which have included DSGE models in their suites of

models used for forecasting and policy analysis. The rapid rise of computing power has enabled researchers to study more and more elaborate model specifications. As we have been writing this chapter, new methods have been developed and novel applications have been explored. While it is impossible to provide an exhaustive treatment of such a dynamic field, we hope that this chapter provides a thorough training for those who are interested in working in this area, offers a good overview of the state of the art as of 2015, and inspires innovative research that expands the frontier of knowledge.

ACKNOWLEDGMENTS

REFERENCES

Aldrich, E.M., 2014. GPU computing in economics. In: Schmedders, K., Judd, K.L. (Eds.), Handbook of Computational Economics, vol. 3. Elsevier, Amsterdam, pp. 557–598.

Aldrich, E.M., Kung, H., 2011. Computational methods for production-based asset pricing models with recursive utility. Economic Research Initiatives at Duke (ERID) Working Paper Series 87.

Aldrich, E.M., Fernández-Villaverde, J., Gallant, A.R., Rubio-Ramírez, J.F., 2011. Tapping the supercomputer under your desk: solving dynamic equilibrium models with graphics processors. J. Econ. Dyn. Control 35, 386–393.

Algan, Y., Allais, O., Den Haan, W.J., Rendahl, P., 2014. Solving and simulating models with heterogeneous agents and aggregate uncertainty. In: Schmedders, K., Judd, K.L. (Eds.), Handbook of Computational Economics, vol. 3. Elsevier, Amsterdam, pp. 277–324.

Altig, D., Christiano, L., Eichenbaum, M., Linde, J., 2011. Firm-specific capital, nominal rigidities and the business cycle. Rev. Econ. Dyn. 14 (2), 225–247. http://ideas.repec.org/a/red/issued/09-191.html.

Altug, S., 1989. Time-to-build and aggregate fluctuations: some new evidence. Int. Econ. Rev. 30 (4), 889–920.

Alvarez, F., Jermann, U.J., 2004. Using asset prices to measure the cost of business cycles. J. Polit. Econ. 112, 1223–1256.

An, S., Schorfheide, F., 2007. Bayesian analysis of DSGE models. Econ. Rev. 26 (2-4), 113–172. http://dx.doi.org/10.1080/07474930701220071.

Anderson, E.W., Hansen, L.P., McGrattan, E.R., Sargent, T.J., 1996. On the mechanics of forming and estimating dynamic linear economies. In: Amman, H.M., Kendrick, D.A., Rust, J. (Eds.), Handbook of Computational Economics, vol. 1. Elsevier, Amsterdam, pp. 171–252.

Andreasen, M.M., 2013. Non-linear DSGE models and the central difference Kalman filter. J. Appl. Econ. 28 (6), 929–955.

Andreasen, M.M., Fernández-Villaverde, J., Rubio-Ramírez, J.F., 2013. The pruned state-space system for nonlinear DSGE models: theory and empircal applications. NBER Working 18983.

Andrews, I., Mikusheva, A., 2015. Maximum likelihood inference in weakly identified DSGE models. Quant. Econ 6 (1), 123–152.

Andrieu, C., Doucet, A., Holenstein, R., 2010. Particle Markov chain Monte Carlo methods. J. R. Stat. Soc. Ser. B 72 (3), 269–342.

Ardia, D., Bastürk, N., Hoogerheide, L., van Dijk, H.K., 2012. A comparative study of Monte Carlo methods for efficient evaluation of marginal likelihood. Comput. Stat. Data Anal. 56 (11), 3398–3414.

Arulampalam, S., Maskell, S., Gordon, N., Clapp, T., 2002. A tutorial on particle filters for online nonlinear/non-gaussian Bayesian tracking. IEEE Trans. Signal Proc. 50 (2), 174–188.

Aruoba, S.B., Schorfheide, F., 2015. Inflation during and after the zero lower bound. Manuscript, University of Maryland.

Aruoba, S.B., Fernández-Villaverde, J., Rubio-Ramírez, J.F., 2006. Comparing solution methods for dynamic equilibrium economies. J. Econ. Dyn. Control 30 (12), 2477–2508.

Barthelmann, V., Novak, E., Ritter, K., 2000. High dimensional polynomial interpolation on sparse grids. Adv. Comput. Math. 12, 273–288.

Bastani, H., Guerrieri, L., 2008. On the application of automatic differentiation to the likelihood function for dynamic general equilibrium models. In: Bischof, C.H., Bücker, H.M., Hovland, P., Naumann, U., Utke, J. (Eds.), Advances in Automatic Differentiation: Lecture Notes in Computational Science and Engineering, vol. 64. Springer, pp. 303–313.

Bender, C.M., Orszag, S.A., 1999. Advanced Mathematical Methods for Scientists and Engineers: Asymptotic Methods and Perturbation Theory. Springer.

Benigno, P., Woodford, M., 2004. Optimal monetary and fiscal policy: a linear-quadratic approach. In: NBER Macroeconomics Annual 2003, vol. 18. MIT Press, Cambridge, MA, pp. 271–364.

Bernanke, B.S., Gertler, M., 1989. Agency costs, net worth, and business fluctuations. Am. Econ. Rev. 79 (1), 14–31.

Bernanke, B.S., Gertler, M., Gilchrist, S., 1999. The financial accelerator in a quantitative business cycle framework. In: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics, vol. 1. Elsevier, pp. 1341–1393.

Bianchi, F., 2013. Regime switches, agents' beliefs, and post-world war II U.S. macroeconomic dynamics. Rev. Econ. Stud. 80 (2), 463–490.

Blanchard, O.J., Kahn, C.M., 1980. The solution of linear difference models under rational expectations. Econometrica 48 (5), 1305–1312.

Bloom, N., 2009. The impact of uncertainty shocks. Econometrica 77, 623–685.

Bocola, L., 2015. The pass-through of sovereign risk. Manuscript, Northwestern University.

Boos, D.D., Monahan, J.F., 1986. Bootstrap methods using prior information. Biometrika 73 (1), 77–83.

Boyd, J.P., 2000. Chebyshev and Fourier Spectral Methods. Dover.

Boyd, J.P., Petschek, R.G., 2014. The relationships between Chebyshev, Legendre and Jacobi polynomials: the generic superiority of Chebyshev polynomials and three important exceptions. J. Sci. Comput. 59, 1–27.

Brenner, S., Scott, R., 2008. The Mathematical Theory of Finite Element Methods. Springer Verlag.

Brown, D.B., Smith, J.E., Peng, S., 2010. Information relaxations and duality in stochastic dynamic programs. Operat. Res. 58 (4), 785–801.

Brumm, J., Scheidegger, S., 2015. Using adaptive sparse grids to solve high-dimensional dynamic models. Manuscript, University of Zurich.

Bungartz, H.J., Griebel, M., 2004. Sparse grids. Acta Numer. 13, 147–269.

Burnside, C., Eichenbaum, M., Rebelo, S., 1993. Labor hoarding and the business cycle. J. Polit. Econ. 101 (2), 245–273. http://ideas.repec.org/a/ucp/jpolec/v101y1993i2p245-73.html.

Cai, Y., Judd, K.L., 2014. Advances in numerical dynamic programming and new applications. In: Schmedders, K., Judd, K.L. (Eds.), Handbook of Computational Economics, vol. 3. Elsevier, pp. 479–516.

Caldara, D., Fernández-Villaverde, J., Rubio-Ramírez, J.F., Yao, W., 2012. Computing DSGE models with recursive preferences and stochastic volatility. Rev. Econ. Dyn. 15, 188–206.

Canova, F., 1994. Statistical inference in calibrated models. J. Appl. Econ. 9, S123–S144.

Canova, F., 2007. Methods for Applied Macroeconomic Research. Princeton University Press.

Canova, F., 2014. Bridging cyclical DSGE models and the raw data. J. Monet. Econ. 67, 1–15. http://www.crei.cat/people/canova/pdf/%20files/dsge_trend.pdf.

Canova, F., De Nicoló, G., 2002. Monetary disturbances matter for business fluctuations in the G-7. J. Monet. Econ. 49 (4), 1131–1159. http://ideas.repec.org/a/ijc/ijcjou/y2007q4a4.html.

Canova, F., Ferroni, F., Matthes, C., 2014. Choosing the variables to estimate singular DSGE models. J. Appl. Econ. 29 (7), 1099–1117.

Cappé, O., Moulines, E., Ryden, T., 2005. Inference in Hidden Markov Models. Springer Verlag.

Cappé, O., Godsill, S.J., Moulines, E., 2007. An overview of existing methods and recent advances in sequential Monte Carlo. Proc. IEEE 95 (5), 899–924.

Carlstrom, C., Fuerst, T.S., 1997. Agency costs, net worth, and business fluctuations: a computable general equilibrium analysis. Am. Econ. Rev. 87, 893–910.

Chang, Y., Doh, T., Schorfheide, F., 2007. Non-stationary hours in a DSGE model. J. Money, Credit, Bank. 39 (6), 1357–1373.

Chari, V.V., Kehoe, P.J., McGrattan, E.R., 2008. Are structural VARs with long-run restrictions useful in developing business cycle theory? J. Monet. Econ. 55 (8), 1337–1352.

Chen, X., 2007. Large sample sieve estimation of semi-nonparametric models. In: Heckman, J.J., Leamer, E. E. (Eds.), Handbook of Econometrics, vol. 6. Elsevier, pp. 5549–5632.

Chernozhukov, V., Hong, H., 2003. An MCMC approach to classical estimation. J. Econ. 115, 293–346.

Chib, S., Greenberg, E., 1995. Understanding the metropolis-hastings algorithm. Am. Stat. 49, 327–335.

Chib, S., Jeliazkov, I., 2001. Marginal likelihoods from the metropolis hastings output. J. Am. Stat. Assoc. 96 (453), 270–281.

Chib, S., Ramamurthy, S., 2010. Tailored randomized block MCMC methods with application to DSGE models. J. Econ. 155 (1), 19–38.

Cho, J.O., Cooley, T.F., Kim, H.S.E., 2015. Business cycle uncertainty and economic welfare. Rev. Econ. Dyn. 18, 185–200.

Chopin, N., 2002. A sequential particle filter for static models. Biometrika 89 (3), 539–551.

Chopin, N., 2004. Central limit theorem for sequential Monte Carlo methods and its application to Bayesian inference. Ann. Stat. 32 (6), 2385–2411.

Chopin, N., Jacob, P.E., Papaspiliopoulos, O., 2012. SMC^2: an efficient algorithm for sequential analysis of state-space models. ArXiv:1101.1528.

Christiano, L.J., 1990. Linear-quadratic approximation and value-function iteration: a comparison. J. Bus. Econ. Stat. 8, 99–113.

Christiano, L.J., Eichenbaum, M., 1992. Current real-business-cycle theories and aggregate labor-market fluctuations. Am. Econ. Rev. 82 (3), 430–450. http://ideas.repec.org/a/aea/aecrev/v82y1992i3p430-50.html.

Christiano, L.J., Fisher, J.D.M., 2000. Algorithms for solving dynamic models with occasionally binding constraints. J. Econ. Dyn. Control 24, 1179–1232.

Christiano, L.J., Vigfusson, R.J., 2003. Maximum likelihood in the frequency domain: the importance of time-to-plan. J. Monet. Econ. 50 (4), 789–815. http://ideas.repec.org/a/eee/moneco/v50y2003i4p789-815.html.

Christiano, L.J., Eichenbaum, M., Evans, C.L., 1999. Monetary policy shocks: what have we learned and to what end. In: Taylor, J.B., Woodford, M. (Eds.), Handbook of Macroeconomics, vol. 1a. North Holland, Amsterdam, pp. 65–148.

Christiano, L.J., Eichenbaum, M., Evans, C.L., 2005. Nominal rigidities and the dynamic effects of a shock to monetary policy. J. Polit. Econ. 113 (1), 1–45.

Christiano, L.J., Eichenbaum, M., Vigfusson, R., 2007. Assessing structural VARs. In: Acemoglu, D., Rogoff, K., Woodford, M. (Eds.), NBER Macroeconomics Annual 2006, vol. 21. MIT Press, Cambridge, pp. 1–72.

Christiano, L.J., Trabandt, M., Walentin, K., 2010. Dsge models for monetary policy analysis. In: Friedman, B.M., Woodford, M. (Eds.), Handbook of Monetary Economics, vol. 3. Elsevier, pp. 285–367. http://ideas.repec.org/h/eee/monchp/3-07.html.

Christiano, L.J., Eihenbaum, M., Rebelo, S.T., 2011. When is the government spending multiplier large? J. Polit. Econ. 119 (1), 78–121.

Christiano, L.J., Motto, R., Rostagno, M., 2014. Risk shocks. Am. Econ. Rev. 104, 27–65.

Clough, R.W., 1960. The finite element method in plane stress analysis. In: Proceedings of the 2nd ASCE Conference on Electronic Computation.

Clough, R.W., Wilson, E.L., 1999. Early finite element research at Berkeley. Manuscript, University of California, Berkeley.

Cochrane, J.H., 1994. Shocks. Carnegie Rochester Conf. Ser. Publ. Pol. 41 (4), 295–364. http://ideas.repec.org/a/ijc/ijcjou/y2007q4a4.html.

Cochrane, J.H., 2011. Determinacy and identification with Taylor rules. J. Polit. Econ. 119 (3), 565–615. http://ideas.repec.org/a/ucp/jpolec/doi10.1086-660817.html.

Creal, D., 2007. Sequential Monte Carlo samplers for Bayesian DSGE models. Manuscript, Chicago Booth.

Creal, D., 2012. A survey of sequential Monte Carlo methods for economics and finance. Econ. Rev. 31 (3), 245–296.

Crisan, D., Rozovsky, B., (Eds.)2011. The Oxford Handbook of Nonlinear Filtering. Oxford University Press.

Curdia, V., Del Negro, M., Greenwald, D.L., 2014. Rare shocks, great recessions. J. Appl. Econ. 29 (7), 1031–1052.

DeJong, D.N., Dave, C., 2007. Structural Macroeconometrics. Princeton University Press.

DeJong, D.N., Ingram, B.F., Whiteman, C.H., 2000. A Bayesian approach to dynamic macroeconomics. J. Econ. 98 (2), 203–223.

Del Moral, P., 2004. Feynman-Kac Formulae. Springer Verlag.

Del Moral, P., 2013. Mean Field Simulation for Monte Carlo Integration. Chapman & Hall/CRC.

Del Negro, M., Schorfheide, F., 2004. Priors from general equilibrium models for VARs. Int. Econ. Rev. 45 (2), 643–673.

Del Negro, M., Schorfheide, F., 2008. Forming priors for DSGE models (and how it affects the assessment of nominal rigidities). J. Monet. Econ. 55 (7), 1191–1208. ISSN 0304-3932. http://dx.doi.org/10.1016/j.jmoneco.2008.09.006. http://www.sciencedirect.com/science/article/B6VBW-4TKPVGT-3/2/508d89fdb8eb927643250b7f36aab161.

Del Negro, M., Schorfheide, F., 2009. Monetary policy with potentially misspecified models. Am. Econ. Rev. 99 (4), 1415–1450. http://www.econ.upenn.edu/schorf/papers/mpol_p11.pdf.

Del Negro, M., Schorfheide, F., 2011. Bayesian macroeconometrics. In: van Dijk, H., Koop, G., Geweke, J. (Eds.), Handbook of Bayesian Econometrics. Oxford University Press, pp. 293–389.

Del Negro, M., Schorfheide, F., 2013. DSGE model-based forecasting. In: Elliott, G., Timmermann, A. (Eds.), Handbook of Economic Forecasting, vol. 2. North Holland, Amsterdam, pp. 57–140.

Del Negro, M., Schorfheide, F., Smets, F., Wouters, R., 2007. On the fit of new Keynesian models. J. Bus. Econ. Stat. 25 (2), 123–162.

Del Negro, M., Hasegawa, R., Schorfheide, F., 2014. Dynamic prediction pools: an investigation of financial frictions and forecasting performance. NBER Working Paper 20575.

Delvos, F.J., 1982. d-Variate Boolean interpolation. J. Approx. Theory 34, 99–114.

Demkowicz, L., 2007. Computing with hp-Adaptive Finite Elements, Volume 1. Chapman & Hall/CRC.

Den Haan, W.J., De Wind, J., 2012. Nonlinear and stable perturbation-based approximations. J. Econ. Dyn. Control 36, 1477–1497.

Den Haan, W.J., Marcet, A., 1990. Solving the stochastic growth model by parameterizing expectations. J. Bus. Econ. Stat. 8 (1), 31–34.

Den Haan, W.J., Marcet, A., 1994. Accuracy in simulations. Rev. Econ. Stud. 61, 3–17.

Díaz-Giménez, J., 1999. Linear-quadratic approximations: an introduction. In: Marimon, R., Scott, A. (Eds.), Computational Methods for the Study of Dynamic Economies. Oxford University Press.

Diebold, F.X., Ohanian, L.E., Berkowitz, J., 1998. Dynamic equilibrium economies: a framework for comparing models and data. Rev. Econ, Stud. 65 (3), 433–452.

Doucet, A., Johansen, A.M., 2011. A tutorial on particle filtering and smoothing: fifteen years later. In: Crisan, D., Rozovsky, B. (Eds.), Handook of Nonlinear Filtering. Oxford University Press.

Doucet, A., de Freitas, N., Gordon, N., 2001. Sequential Monte Carlo Methods in Practice. Springer Verlag.

Dridi, R., Guay, A., Renault, E., 2007. Indirect inference and calibration of dynamic stochastic general equilibrium models. J. Econ. 136 (2), 397–430.

Dufour, J.M., Khalaf, L., Kichian, M., 2013. Identification-robust analysis of DSGE and structural macroeconomic models. J. Monet. Econ. 60, 340–350.

Durbin, J., Koopman, S.J., 2001. Time Series Analysis by State Space Methods. Oxford University Press.

Durham, G., Geweke, J., 2014. Adaptive sequential posterior simulators for massively parallel computing environments. Adv. Econ. 34, 1–44.

Eggertsson, G.B., Woodford, M., 2003. The zero bound on interest rates and optimal monetary policy. Brook. Pap. Econ. Act. 34, 139–235.

Epstein, L.G., Zin, S.E., 1989. Substitution, risk aversion, and the temporal behavior of consumption and asset returns: a theoretical framework. Econometrica 57, 937–969.

Erdös, P., Turán, P., 1937. On interpolation I. Quadrature and mean convergence in the Lagrange interpolation. Ann. Math. 38, 142–155.

Fair, R.C., Taylor, J.B., 1983. Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models. Econometrica 51, 1169–1185.

Faust, J., 1998. The robustness of identified VAR conclusions about money. Carnegie Rochester Conf. Ser. Publ. Pol. 49 (4), 207–244. http://ideas.repec.org/a/ijc/ijcjou/y2007q4a4.html.

Fernández-Villaverde, J., 2010. Fiscal policy in a model with financial frictions. Am. Econ. Rev. Pap. Proc. 100, 35–40.

Fernández-Villaverde, J., Levintal, O., 2016. Solution methods for models with rare disasters. Manuscript, University of Pennsylvania.

Fernández-Villaverde, J., Rubio-Ramírez, J.F., 2004. Comparing dynamic equilibrium models to data: a Bayesian approach. J. Econ. 123 (1), 153–187.

Fernández-Villaverde, J., Rubio-Ramírez, J.F., 2006. Solving DSGE models with perturbation methods and a change of variables. J. Econ. Dyn. Control 30, 2509–2531.

Fernández-Villaverde, J., Rubio-Ramírez, J.F., 2007. Estimating macroeconomic models: a likelihood approach. Rev. Econ. Stud. 74 (4), 1059–1087.

Fernández-Villaverde, J., Rubio-Ramírez, J.F., 2008. How structural are structural parameters? In: Acemoglu, D., Rogoff, K., Woodford, M. (Eds.), NBER Macroeconomics Annual 2007, vol. 22. University of Chicago Press, Chicago, IL.

Fernández-Villaverde, J., Rubio-Ramírez, J.F., Santos, M.S.S., 2006. Convergence properties of the likelihood of computed dynamic models. Econometrica 74 (1), 93–119. http://dx.doi.org/10.1111/j.1468-0262.2006.00650.x.

Fernández-Villaverde, J., Rubio-Ramírez, J.F., Sargent, T.J., Watson, M.W., 2007. ABCs (and Ds) of understanding VARs. Am. Econ. Rev. 97 (3), 1021–1026.

Fernández-Villaverde, J., Guerrón-Quintana, P.A., Rubio-Ramírez, J.F., Uribe, M., 2011. Risk matters: the real effects of volatility shocks. Am. Econ. Rev. 101, 2530–2561.

Fernández-Villaverde, J., Guerrón-Quintana, P.A., Rubio-Ramírez, J.F., 2014. Supply-side policies and the zero lower bound. IMF Econ. Rev. 62, 248–260.

Fernández-Villaverde, J., Gordon, G., Guerrón-Quintana, P.A., Rubio-Ramírez, J.F., 2015a. Nonlinear adventures at the zero lower bound. J. Econ. Dyn. Control 57, 182–204.

Fernández-Villaverde, J., Guerrón-Quintana, P.A., Rubio-Ramírez, J.F., 2015b. Estimating dynamic equilibrium models with stochastic volatility. J. Econ. 185, 216–229.

Flury, T., Shephard, N., 2011. Bayesian inference based only on simulated likelihood: particle filter analysis of dynamic economic models. Econ. Theory 27, 933–956.

Fornberg, B., 1996. A Practical Guide to Pseudospectral Methods. Cambridge University Press.

Forneron, J.J., Ng, S., 2015. The ABC of simulation estimation with auxiliary statistics. Manuscript, Columbia University.

Gallant, A.R., 2015. Reflections on the probability space induced by moment conditions with implications for Bayesian inference. J. Fin. Econ. Forthcoming. http://jfec.oxfordjournals.org/content/early/2015/05/28/jjnec.nbv008.abstract.

Gallant, A.R., Giacomini, R., Ragusa, G., 2013. Generalized method of moments with latent variables. CEPR Discussion Papers DP9692.

Galor, O., 2007. Discrete Dynamical Systems. Springer.

Gaspar, J., Judd, K.L., 1997. Solving large-scale rational-expectations models. Macroecon. Dyn. 1, 45–75.

Geweke, J., 1999. Using simulation methods for Bayesian econometric models: inference, development, and communication. Econ. Rev. 18 (1), 1–126.

Geweke, J., 2005. Contemporary Bayesian Econometrics and Statistics. John Wiley & Sons, Inc.

Geweke, J., 2010. Complete and Incomplete Econometric Models. Princeton University Press, Princeton, NJ.

Geweke, J., Amisano, G., 2011. Optimal prediction pools. J. Econ. 164, 130–141.

Geweke, J., Amisano, G., 2012. Prediction with misspecified models. Am. Econ. Rev. Pap. Proc. 103 (3), 482–486.

Gordon, G., 2011. Computing dynamic heterogeneous-agent economies: tracking the distribution. PIER Working Paper 11-018, University of Pennsylvania.

Gorodnichenko, Y., Ng, S., 2010. Estimation of DSGE models when the data are persistent. J. Monet. Econ. 57 (3), 325–340.

Gourieroux, C., Monfort, A., Renault, E., 1993. Indirect inference. J. Appl. Econ. 8, S85–S118.

Gourieroux, C., Phillips, P.C.B., Yu, J., 2010. Indirect inference for dynamic panel models. J. Econ. 157 (1), 68–77.

Guerrón-Quintana, P.A., 2010. What you match does matter: the effects of observable variables on DSGE estimation. J. Appl. Econ. 25, 774–804.

Guerrón-Quintana, P.A., Inoue, A., Kilian, L., 2013. Frequentist inference in weakly identified dynamic stochastic general equilibrium models. Quant. Econ. 4, 197–229.

Guerrón-Quintana, P.A., Inoue, A., Kilian, L., 2014. Impulse response matching estimators for DSGE models. In: Center for Financial Studies (Frankfurt am Main): CFS working paper series, No. 498, CFS working paper seriesWirtschaftswissenschaften URL http://ssrn.com/abstract=2533453.

Guo, D., Wang, X., Chen, R., 2005. New sequential Monte Carlo methods for nonlinear dynamic systems. Stat. Comput. 15, 135–147.

Gust, C., Herbst, E., López-Salido, J.D., Smith, M.E., 2016. The empirical implications of the interest-rate lower bound. Federal Reserve Board.

Hamilton, J.D., 1994. Time Series Analysis. Princeton University Press.

Hansen, L.P., 1982. Large sample properties of generalized method of moments estimators. Econometrica 50 (4), 1029–1054. http://ideas.repec.org/a/ecm/emetrp/v50y1982i4p1029-54.html.

Hansen, G.D., Prescott, E.C., 1995. Recursive methods for computing equilibria of business cycle models. In: Cooley, T.F. (Ed.), Frontiers of Business Cycle Research. Princeton University Press, pp. 39–64.

Hansen, L.P., Sargent, T.J., 2013. Recursive Models of Dynamic Linear Economies. Princeton Press.

Hansen, L.P., Heaton, J.C., Li, N., 2008. Consumption strikes back? Measuring long-run risk. J. Polit. Econ. 116 (2), 260–302.

Herbst, E., Schorfheide, F., 2014. Sequential Monte Carlo sampling for DSGE models. J. Appl. Econ. 29 (7), 1073–1098.

Herbst, E., Schorfheide, F., 2015. Bayesian Estimation of DSGE Models. Princeton University Press.

Hnatkosvaka, V., Marmer, V., Tang, Y., 2012. Comparison of misspecified calibrated models: the minimum distance approach. J. Econ. 169 (1), 131–138.

Hughes, T.J.R., 2000. The Finite Element Method: Linear Static and Dynamic Finite Element Analysis. Dover.

Hurwicz, L., 1962. On the structural form of interdependent systems. In: Nagel, E., Tarski, A. (Eds.), Logic, Methodology and Philosophy of Science. Stanford University Press.

Ingram, B., Whiteman, C., 1994. Supplanting the minnesota prior-forecasting macroeconomic time series using real business cycle model priors. J. Monet. Econ. 49 (4), 1131–1159. http://ideas.repec.org/a/ijc/ijcjou/y2007q4a4.html.

Inoue, A., Shintani, M., 2014. Quasi-Bayesian model selection. Manuscript, Vanderbilt University.

Iskrev, N., 2010. Local identification of DSGE models. J. Monet. Econ. 2, 189–202. http://dx.doi.org/10.1016/j.jmoneco.2009.12.007.

Jin, H.H., Judd, K.L., 2002. Perturbation methods for general dynamic stochastic models. Manuscript, Hoover Institution.

Judd, K., 1998. Numerical Methods in Economics. MIT Press, Cambridge.

Judd, K.L., 1992. Projection methods for solving aggregate growth models. J. Econ. Theory 58, 410–452.

Judd, K.L., 2003. Perturbation methods with nonlinear changes of variables. Manuscript, Hoover Institution.

Judd, K.L., Guu, S.M., 1993. Perturbation solution methods for economic growth models. In: Varian, H. (Ed.), Economic and Financial Modeling with Mathematica. Springer Verlag, pp. 80–103.

Judd, K.L., Guu, S.M., 1997. Asymptotic methods for aggregate growth models. J. Econ. Dyn. Control 21, 1025–1042.

Judd, K.L., Guu, S.M., 2001. Asymptotic methods for asset market equilibrium analysis. Econ. Theory 18, 127–157.

Judd, K.L., Maliar, L., Maliar, S., 2011. How to solve dynamic stochastic models computing expectations just once. NBER Working Paper 17418.

Judd, K.L., Maliar, L., Maliar, S., 2011. Numerically stable and accurate stochastic simulation methods for solving dynamic models. Quant. Econ. 2, 173–210.

Judd, K.L., Maliar, L., Maliar, S., 2014. Lower bounds on approximation errors: testing the hypothesis that a numerical solution is accurate. Manuscript, Hoover Institution.

Judd, K.L., Maliar, L., Maliar, S., Valero, R., 2014. Smolyak method for solving dynamic economic models: Lagrange interpolation, anisotropic grid and adaptive domain. J. Econ. Dyn. Control 44, 92–123.

Justiniano, A., Primiceri, G.E., 2008. The time-varying volatility of macroeconomic fluctuations. Am. Econ. Rev. 98 (3), 604–641.

Kantas, N., Doucet, A., Singh, S., Maciejowski, J., Chopin, N., 2014. On particle methods for parameter estimation in state-space models. ArXiv Working Paper 1412.8659v1.

Kilian, L., 1998. Small-sample confidence intervals for impulse response functions. Rev. Econ. Stat. 80 (2), 218–230. http://dx.doi.org/10.1162/003465398557465.

Kilian, L., 1999. Finite-sample properties of percentile and percentile-t bootstrap confidence intervals for impulse responses. Rev. Econ. Stat. 81 (4), 652–660.

Kim, J.Y., 2002. Limited information likelihood and Bayesian analysis. J. Econ. 107 (1-2), 175–193. http://dx.doi.org/10.1016/S0304-4076(01)00119-1.

Kim, J., Kim, S.H., 2003. Spurious welfare reversals in international business cycle models. J. Int. Econ. 60, 471–500.

Kim, J., Kim, S.H., Schaumburg, E., Sims, C.A., 2008. Calculating and using second-order accurate solutions of discrete time dynamic equilibrium models. J. Econ. Dyn. Control 32, 3397–3414.

Kimball, M.S., 1990. Precautionary saving in the small and in the large. Econometrica 58, 53–73.

King, R.G., Watson, M.W., 1998. The solution of singular linear difference systems under rational expectations. Int. Econ. Rev. 39, 1015–1026.

King, R.G., Plosser, C.I., Rebelo, S., 1988. Production, growth, and business cycles: I the basic neoclassical model. J. Monet. Econ. 21 (2-3), 195–232.

King, R.G., Plosser, C.I., Rebelo, S.T., 2002. Production, growth and business cycles: technical appendix. Comput. Econ. 20, 87–116.

Kitamura, Y., Otsu, T., 2011. Bayesian analysis of moment condition models using nonparametric priors. Manuscript, Yale University and LSE.

Kleibergen, F., Mavroeidis, S., 2009. Weak instrument robust tests in GMM and the New Keynesian Phillips curve. J. Bus. Econ. Stat. 27 (3), 293–311. http://ideas.repec.org/a/bes/jnlbes/v27i3y2009p293-311.html.

Kleibergen, F., Mavroeidis, S., 2014. Identification issues in limited-information Bayesian analysis of structural macroeconomic models. J. Appl. Econ. 29, 1183–1209.

Klein, P., 2000. Using the generalized Schur form to solve a multivariate linear rational expectations model. J. Econ. Dyn. Control 24 (10), 1405–1423. http://dx.doi.org/10.1016/S0165-1889(99)00045-7.

Kociecki, A., Kolasa, M., 2015. Global identification of linearized DSGE models. Manuscript, Bank of Poland.

Kogan, L., Mitra, I., 2014. Accuracy verification for numerical solutions of equilibrium models. Manuscript, MIT.

Kollmann, R., 2015. Tractable latent state filtering for non-linear DSGE models using a second-order approximation and pruning. Comput. Econ. 45, 239–260.

Komunjer, I., Ng, S., 2011. Dynamic identification of DSGE models. Econometrica 79 (6), 1995–2032.

Koop, G., Pesaran, H.M., Potter, S.M., 1996. Impulse response analysis in nonlinear multivariate models. J. Econ. 74, 119–147.

Koop, G., Pesaran, H.M., Smith, R.P., 2013. On identification of Bayesian DSGE models. J. Bus. Econ. Stat. 31 (3), 300–314.

Kopecky, K.A., Suen, R.M.H., 2010. Finite state Markov-chain approximations to highly persistent processes. Rev. Econ. Dyn. 13, 701–714.

Kormilitsina, A., Nekipelov, D., 2012. Approximation properties of Laplace-type estimators. Adv. Econ. 28, 291–318.

Kormilitsina, A., Nekipelov, D., 2016. Consistent variance of the Laplace type estimators: application to DSGE models. Int. Econ. Rev. 57 (2), 603–622.

Krüger, D., Kubler, F., 2004. Computing equilibrium in OLG models with stochastic production. J. Econ. Dyn. Control 28, 1411–1436.

Krusell, P., Smith, A.A., 1998. Income and wealth heterogeneity in the macroeconomy. J. Polit. Econ. 106, 867–896.

Kwan, Y.K., 1999. Asymptotic Bayesian analysis based on a limited information estimator. J. Econ. 88, 99–121.

Kydland, F.E., Prescott, E.C., 1982. Time to build and aggregate fluctuations. Econometrica 50 (6), 1345–1370.

Lancaster, T., 2004. An Introduction to Modern Bayesian Econometrics. Blackwell Publishing.

Lanczos, C., 1938. Trigonometric interpolation of empirical and analytical functions. J. Math. Phys. 17, 123–199.

Lazar, N.A., 2003. Bayesian empirical likelihood. Biometrika 90 (2), 319–326.

Lee, B.S., Ingram, B.F., 1991. Simulation estimation of time-series models. J. Econ. 47 (2-3), 197–205. http://ideas.repec.org/a/eee/econom/v47y1991i2-3p197-205.html.

Leeper, E.M., 1991. Equilibria under 'active' and 'passive' monetary and fiscal policies. J. Monet. Econ. 27, 129–147.

Leeper, E.M., Sims, C.A., 1995. Toward a modern macroeconomic model usable for policy analysis. In: Fischer, S., Rotemberg, J.J. (Eds.), NBER Macroeconomics Annual 1994. MIT Press, Cambridge, pp. 81–118.

Leland, H.E., 1968. Saving and uncertainty: the precautionary demand for saving. Q. J. Econ. 82, 465–473.

Levin, A., Onatski, A., Williams, J.C., Williams, N., 2006. Monetary policy under uncertainty in micro-founded macroeconometric models. In: Gertler, M., Rogoff, K. (Eds.), NBER Macroeconomics Annual 2005, vol. 20. MIT Press, Cambridge, pp. 229–287. http://www.columbia.edu/%7Eao2027/LOWW.pdf.

Levintal, O., 2015a. Fifth-order perturbation solution to DSGE models. Manuscript, Interdisciplinary Center Herzliya.

Levintal, O., 2015b. Taylor projection: a new solution method to dynamic general equilibrium models. Manuscript, Interdisciplinary Center Herzliya.

Liu, J.S., 2001. Monte Carlo Strategies in Scientific Computing. Springer Verlag.

Lubik, T., Schorfheide, F., 2003. Computing sunspot equilibria in linear rational expectations models. J. Econ. Dyn. Control 28 (2), 273–285.

Lubik, T., Schorfheide, F., 2005. Do central banks respond to exchange rate movements? A structural investigation. J. Monet. Econ. 54 (4), 1069–1087.

Lubik, T., Schorfheide, F., 2006. A Bayesian look at the new open macroeconomics. NBER Macroeconomics Annual 2005.

Lubik, T.A., Schorfheide, F., 2004. Testing for indeterminacy: an application to U.S. monetary policy. Am. Econ. Rev. 94 (1), 190–217.

Lucas Jr., R.E., 1987. Models of Business Cycles. Basil Blackwell, Oxford.

Lütkepohl, H., 1990. Asymptotic distributions of impulse response functions and forecast error variance decompositions of vector autoregressive models. Rev. Econ. Stat. 72 (1), 116–125. http://ideas.repec.org/a/tpr/restat/v72y1990i1p116-25.html.

Maliar, L., Maliar, S., 2014. Numerical methods for large scale dynamic economic models. In: Schmedders, K., Judd, K.L. (Eds.), Handbook of Computational Economics. vol. 3. Elsevier, pp. 325–477.

Maliar, L., Maliar, S., 2015. Merging simulation and projection approaches to solve high-dimensional problems with an application to a New Keynesian model. Quant. Econ. 6, 1–47.

Maliar, L., Maliar, S., Judd, K.L., 2011. Solving the multi-country real business cycle model using Ergodic set methods. J. Econ. Dyn. Control 35, 207–228.

Maliar, L., Maliar, S., Taylor, J.B., Tsener, I., 2015. A tractable framework for analyzing a class of nonstationary Markov models. NBER Working Paper 21155.

Maliar, L., Maliar, S., Villemot, S., 2013. Taking perturbation to the accuracy frontier: a hybrid of local and global solutions. Comput. Econ. 42, 307–325.

Malik, S., Pitt, M.K., 2011. Particle filters for continuous likelihood evaluation and maximization. J. Econ. 165, 190–209.

Malin, B.A., Krüger, D., Kubler, F., 2011. Solving the multi-country real business cycle model using a Smolyak-collocation method. J. Econ. Dyn. Control 35, 229–239.

Marcet, A., Lorenzoni, G., 1999. Parameterized expectations approach: some practical issues. In: Marimon, R., Scott, A. (Eds.), Computational Methods for the Study of Dynamic Economies. Oxford University Press.

Marcet, A., Marshall, D.A., 1994. Solving nonlinear rational expectations models by parameterized expectations: convergence to stationary solutions. .

Marmer, V., Otsu, T., 2012. Optimal comparison of misspecified moment restriction models under a chosen measure of fit. J. Econ. 170 (2), 538–550.

Mason, J.C., Handscomb, D., 2003. Chebyshev Polynomials. CRC Press.

Mavroeidis, S., 2005. Identification issues in forward-looking models estimated by GMM, with an application to the Phillips curve. J. Money Credit Bank. 37 (3), 421–448. http://ideas.repec.org/a/mcb/jmoncb/v37y2005i3p421-48.html.

Mavroeidis, S., 2010. Monetary policy rules and macroeconomic stability: some new evidence. Am. Econ. Rev. 100 (1), 491–503. http://www.ingentaconnect.com/content/aea/aer/2010/00000100/00000001/art00018.

Mavroeidis, S., Plagborg-Moller, M., Stock, J.H., 2014. Empirical evidence on inflation expectations in the New Keynesian Phillips curve. J. Econ. Lit. 52 (1), 124–188. http://ideas.repec.org/a/aea/jeclit/v52y2014i1p124-88.html.

McGrattan, E.R., 1994. The macroeconomic effects of distortionary taxation. J. Monet. Econ. 33 (3), 573–601.

McGrattan, E.R., 1996. Solving the stochastic growth model with a finite element method. J. Econ. Dyn. Control 20, 19–42.

Mittnik, S., Zadrozny, P.A., 1993. Asymptotic distributions of impulse responses, step responses, and variance decompositions of estimated linear dynamic models. Econometrica 61 (4), 857–870. http://ideas.repec.org/a/ecm/emetrp/v61y1993i4p857-70.html.

Müller, U., 2012. Measuring prior sensitivity and prior informativeness in large Bayesian models. J. Monet. Econ. 59, 581–597.

Müller, U., 2013. Risk of Bayesian inference in misspecified models, and the sandwich covariance matrix. Econometrica 81 (5), 1805–1849.

Nishiyama, S., Smetters, K., 2014. Analyzing fiscal policies in a heterogeneous-agent overlapping-generations economy. In: Schmedders, K., Judd, K.L. (Eds.), Handbook of Computational Economics, vol. 3. Elsevier, pp. 117–160.

Nobile, F., Tempone, R., Webster, C.G., 2008. An anisotropic sparse grid stochastic collocation method for partial differential equations with random input data. SIAM J. Numer. Anal. 46, 2411–2442.

Nocedal, J., Wright, S.J., 2006. Numerical Optimization. Springer Verlag.

Otrok, C., 2001. On measuring the welfare costs of business cycles. J. Monet. Econ. 47 (1), 61–92.

Pakes, A., Pollard, D., 1989. Simulation and the asymptotics of optimization estimators. Econometrica 57 (5), 1027–1057. http://EconPapers.repec.org/RePEc:ecm:emetrp:v:57:y:1989:i:5:p:1027-57.

Parra-Alvarez, J.C., 2015. Solution methods and inference in continuous-time dynamic equilibrium economies. Aarhus University.

Pesavento, E., Rossi, B., 2007. Impulse response confidence intervals for persistent data: what have we learned? J. Econ. Dyn. Control 31 (7), 2398–2412. ISSN 0165-1889. http://dx.doi.org/10.1016/j.jedc.2006.07.006.

Phillips, P.C.B., 1998. Impulse response and forecast error variance asymptotics in nonstationary vars. J. Econ. 83 (1-2), 21–56. http://ideas.repec.org/a/eee/econom/v83y1998i1-2p21-56.html.

Phillips, P.C., Solo, V., 1992. Asymptotics for linear processes. Ann. Stat. 20 (2), 971–1001.

Piazzesi, M., Schneider, M., 2006. Equilibrium yield curves. NBER Macroeconomics Annual 2006.

Pitt, M.K., Silva, R.d.S., Giordani, P., Kohn, R., 2012. On some properties of Markov chain Monte Carlo simulation methods based on the particle filter. J. Econ. 171, 134–151.

Pratt, J.W., Raiffa, H., Schlaifer, R., 1965. Introduction to Statistical Decision Theory. Wiley, New York, NY.

Preston, B., Roca, M., 2007. Incomplete markets, heterogeneity and macroeconomic dynamics. NBER Working Paper 13260.

Priestley, M.B., 1981. Spectral Analysis and Time Series. Academic Press.

Qu, Z., 2014. Inference in dynamic stochastic general equilibrium models with possible weak identification. Quant. Econ. 5, 457–494.

Qu, Z., 2015. A composite likelihood framework for analyzing singular DSGE models. Manuscript, Boston University.

Qu, Z., Tkachenko, D., 2012. Identification and frequency domain quasi-maximum likelihood estimation of linearized DSGE models. Quant. Econ. 3, 95–132.

Qu, Z., Tkachenko, D., 2014. Local and global parameter identification in DSGE models: allowing for indeterminancy. Manuscript, Boston University.

Rabanal, P., Rubio-Ramírez, J.F., 2005. Comparing New Keynesian models of the business cycle: a Bayesian approach. J. Monet. Econ. 52 (6), 1151–1166.

Ramey, V.A., 2016. Macroeconomic shocks and their propagation. In: Taylor, J.B., Uhlig, H. (Eds.), Handbook of Macroeconomics, vol. 2A. Elsevier, Amsterdam, Netherlands, pp. 71–162.

Ríos-Rull, J.V., Schorfheide, F., Fuentes-Albero, C., Kryshko, M., Santaeulalia-Llopis, R., 2012. Methods versus substance: measuring the effects of technology shocks. J. Monet. Econ. 59 (8), 826–846.

Robert, C.P., Casella, G., 2004. Monte Carlo Statistical Methods. Springer.

Rossi, B., Pesavento, E., 2006. Small-sample confidence intervals for multivariate impulse response functions at long horizons. J. Appl. Econ. 21 (8), 1135–1155. http://ideas.repec.org/a/jae/japmet/v21y2006i8p1135-1155.html.

Rotemberg, J.J., Woodford, M., 1997. An optimization-based econometric framework for the evaluation of monetary policy. In: Bernanke, B.S., Rotemberg, J.J. (Eds.), NBER Macroeconomics Annual 1997. MIT Press, Cambridge.

Rouwenhorst, K.G., 1995. Asset pricing implications of equilibrium business cycle models. In: Cooley, T.F. (Ed.), Frontiers of Business Cycle Research. Princeton University Press, pp. 294–330.

Rudebusch, G., Swanson, E., 2011. Examining the bond premium puzzle with a DSGE model. J. Monet. Econ. 55.

Rudebusch, G., Swanson, E., 2012. The bond premium in a DSGE model with long-run real and nominal risks. Am. Econ. J. Macroecon. 4, 105–143.

Rudin, W., 1976. Principles of Mathematical Analysis. McGraw and Hill, New York, NY.

Ruge-Murcia, F., 2012. Estimating nonlinear DSGE models by the simulated method of moments: with an application to business cycles. J. Econ. Dyn. Control 36, 914938.

Ruge-Murcia, F., 2014. Indirect inference estimation of nonlinear dynamic general equilibrium models: with an application to asset pricing under skewness risk. McGill University, Working Paper.

Ruge-Murcia, F.J., 2007. Methods to estimate dynamic stochastic general equilibrium models. J. Econ. Dyn. Control 31 (8), 2599–2636. http://ideas.repec.org/a/eee/dyncon/v31y2007i8p2599-2636.html.

Rust, J., 1996. Numerical dynamic programming in economics. In: Amman, H.M., Kendrick, D.A., Rust, J. (Eds.), Handbook of Computational Economics, vol. 1. Elsevier, pp. 619–729.

Sala, L., 2015. DSGE models in the frequency domain. J. Appl. Econ. 30, 219–240. http://ideas.repec.org/p/igi/igierp/504.html.

Samuelson, P.A., 1970. The fundamental approximation theorem of portfolio analysis in terms of means, variances and higher moments. Rev. Econ. Stud. 37, 537–542.

Sandmo, A., 1970. The effect of uncertainty on saving decisions. Rev. Econ. Stud. 37, 353–360.

Santos, M.S., 1992. Differentiability and comparative analysis in discrete-time infinite-horizon optimization. J. Econ. Theory 57, 222–229.

Santos, M.S., 1993. On high-order differentiability of the policy function. Econ. Theory 2, 565–570.

Santos, M.S., 2000. Accuracy of numerical solutions using the Euler equation residuals. Econometrica 68, 1337–1402.

Santos, M.S., Peralta-Alva, A., 2005. Accuracy of simulations for stochastic dynamic models. Econometrica 73 (6), 1939–1976.http://EconPapers.repec.org/RePEc:ecm:emetrp:v:73:y:2005:i:6:p:1939-1976.

Santos, M.S., Peralta-Alva, A., 2014. Analysis of numerical errors. In: Schmedders, K., Judd, K.L. (Eds.), Handbook of Computational Economics, vol. 3. Elsevier, pp. 517–556.

Santos, M.S., Rust, J., 2004. Convergence properties of policy iteration. SIAM J. Control Optim. 42, 2094–2115.

Santos, M.S., Vigo-Aguiar, J., 1998. Analysis of a numerical dynamic programming algorithm applied to economic models. Econometrica 66, 409–426.

Scalone, V., 2015. Estimating non-linear DSGEs with approximate Bayesian computations. Manuscript, University of Rome La Sapienza.

Schennach, S.M., 2005. Bayesian exponential tilted empirical likelihood. Biometrika 92, 31–46.

Schmitt-Grohé, S., Uribe, M., 2004. Solving dynamic general equilibrium models using a second-order approximation to the policy function. J. Econ. Dyn. Control 28, 755–775.

Schorfheide, F., 2000. Loss function-based evaluation of DSGE models. J. Appl. Econ. 15, 645–670.

Schorfheide, F., 2005. Learning and monetary policy shifts. Rev. Econ. Dyn. 8 (2), 392–419.

Schorfheide, F., 2005. VAR forecasting under misspecification. J. Econ. 128 (1), 99–136.

Schorfheide, F., 2013. Estimation and evaluation of DSGE models: progress and challenges. In: Acemoglu, D., Arellano, M., Dekel, E. (Eds.), Advances in Economics and Econometrics: Tenth World Congress, vol. III. Cambridge University Press, pp. 184–230.

Schwarz, G., 1978. Estimating the dimension of a model. Ann. Stat. 6 (2), 461–464.

Shin, M., 2014. Bayesian GMM. PhD Thesis, University of Pennsylvania.

Sikorski, K., 1985. Optimal solution of nonlinear equations. J. Complex. 1, 197–209.

Simmonds, J.G., Mann, J.E.J., 1997. A First Look at Perturbation Theory. Dover.

Simon, H.A., 1956. Dynamic programming under uncertainty with a quadratic criterion function. Econometrica 24, 74–81.

Sims, C.A., 2002. Solving linear rational expectations models. Comput. Econ. 20, 1–20.

Sims, C.A., Waggoner, D., Zha, T., 2008. Methods for inference in large multiple-equation Markov-switching models. J. Econ. 146 (2), 255–274. http://www.frbatlanta.org/filelegacydocs/wp0622.pdf.

Smets, F., Wouters, R., 2003. An estimated dynamic stochastic general equilibrium model of the euro area. J. Eur. Econ. Assoc. 1 (5), 1123–1175.

Smets, F., Wouters, R., 2007. Shocks and frictions in US business cycles: a Bayesian DSGE approach. Am. Econ. Rev. 97, 586–608.

Smith Jr., A., 1993. Estimating nonlinear time-series models using simulated vector autoregressions. J. Appl. Econ. 8, S63–S84. http://ideas.repec.org/a/jae/japmet/v8y1993isps63-84.html.

Smolyak, S.A., 1963. Quadrature and interpolation formulas for tensor products of certain classes of functions. Sov. Math. 4, 240–243.

Solín, P., Segeth, K., Doležel, I., 2004. Higher-Order Finite Elements Method. Chapman & Hall/CRC.

Stachurski, J., Martin, V., 2008. Computing the distributions of economic models via simulation. Econometrica 76, 443–450.

Stock, J.H., Watson, M.W., 2001. Vector autoregressions. J. Econ. Perspect. 15 (4), 101–115. http://ideas.repec.org/a/ijc/ijcjou/y2007q4a4.html.

Stokey, N.L., Lucas Jr., R.E., Prescott, E.C., 1989. Recursive Methods in Economic Dynamics. Harvard University Press.

Swanson, E.T., Anderson, G.S., Levin, A.T., 2006. Higher-order perturbation solutions to dynamic, discrete-time rational expectations models. Federal Reserve Bank of San Francisco, Working Paper Series, 2006-01.

Tallarini, T.D.J., 2000. Risk-sensitive real business cycles. J. Monet. Econ. 45, 507–532.

Tauchen, G., 1986. Finite state Markov-chain approximations to univariate and vector autoregressions. Econ. Lett. 20, 177–181.

Theil, H., 1957. A note on certainty equivalence in dynamic planning. Econometrica 25, 346–349.

Thompson, J.F., Warsi, Z., Mastin, C.W., 1985. Numerical Grid Generation: Foundations and Applications. North-Holland, New York, NY.

Uhlig, H., 1999. A toolkit for analysing nonlinear dynamic stochastic models easily. In: Marimon, R., Scott, A. (Eds.), Computational Methods for the Study of Dynamic Economies. Oxford University Press, pp. 30–61.

Uhlig, H., 2005. What are the effects of monetary policy on output? Results from an agnostic identification procedure. J. Monet. Econ. 52 (2), 381–419.

van Binsbergen, J.H., Fernández-Villaverde, J., Koijen, R.S., Rubio-Ramírez, J.F., 2012. The term structure of interest rates in a DSGE model with recursive preferences. J. Monet. Econ. 59, 634–648.

Waggoner, D., Zha, T., 2012. Confronting model misspecification in macroeconomics. J. Econ. 171 (2), 167184.

White, H., 1994. Estimation, Inference, and Specification Analysis. Cambridge University Press.

Woodford, M., 2003. Optimal interest-rate smoothing. Rev. Econ. Stud. 70, 861–886.