Solution and Estimation Methods for DSGE Models

Jesus Fernandez-Villaverde
University of Pennsylvania

Juan Rubio-Ramirez
Duke University

Frank Schorfheide*
University of Pennsylvania

April 7, 2015

Prepared for
Handbook of Macroeconomics, Volume 2

Preliminary and Incomplete

*Correspondence: Frank Schorfheide: Department of Economics, 3718 Locust Walk, University of Pennsylvania, Philadelphia, PA 19104-6297. Email: schorf@ssc.upenn.edu. (...) Schorfheide gratefully acknowledges financial support from the National Science Foundation under Grant SES 1061358. (...)

Abstract

(TBW)

Contents

1 Introduction **1**

2 Solution Methods **1**

3 A General Framework **3**

3.1 The Stochastic Neoclassical Growth Model 4

3.2 A Value Function 5

3.3 Euler Equation 6

3.4 Conditional Expectations 8

3.5 The Way Forward 9

4 Perturbation **10**

4.1 An Example 12

4.2 Regular versus Singular Perturbations 13

4.3 The General Case 14

4.4 Steady State 15

4.5 Exogenous Stochastic Process 17

4.6 Solution of the Model 18

4.7 Perturbation 19

4.8 First-Order Approximation 20

4.9 Second-Order Approximation 23

4.10 Higher-Order Approximations 26

4.11 An Example 26

4.12 Prunning 32

4.13 Change of Variables 32

4.14 The Optimal Change of Variables 35

4.15 Perturbing the Value Function 36

5 Projection **47**

5.1 Relation with Econometrics 48

5.2 Spectral Techniques 49

5.3 Monomials 49

5.4 Trigonometric series 50

5.5 Orthoghonal polynomials of Jacobi type 50

5.6 Chebyshev Polynomials 51

5.7 Tensors 53

5.8 Complete Polynomials 53

5.9 Smolyak's Algorithm 54

5.10 Finite Elements 56

5.11 Choosing the Objective Function 57

5.12 Weight Function I: Least Squares 58

5.13 Weight Function II: Subdomain 59

5.14 Weight Function III: Moments 59

5.15 Weight Function V: Galerkin or Rayleigh-Ritz 60

6 Which Methods to Use? **61**

7 Analysis of Error **61**

8 Parallel Programming **62**

8.1 Example I: value function iteration 63

8.2 Example II: random walk Metropolis-Hastings 64

9 Confronting DSGE Models with Data **65**

9.1 A Simple DSGE Model and Its Solution 65

9.2 Model Implications 71

9.3 Empirical Analogues 80

9.4 Dealing with Trends 86

9.5 What Next? 88

10 The Likelihood Function **89**

10.1 A Generic Filter and Smoother 90

10.2 Likelihood Function for a Linearized DSGE Model 92

10.3 Likelihood Function for Nonlinear DSGE Models 93

11 Statistical Inference **97**

11.1 Identification 99

11.2 Frequentist Inference 101

11.3 Bayesian Inference 105

12 Frequentist Estimation Techniques **108**

12.1 Likelihood-Based Estimation 108

12.2 (Simulated) Method of Moments Estimation and Indirect Inference 112

12.3 Impulse Response Function Matching 116

12.4 GMM Estimation 118

13 Bayesian Estimation Techniques **120**

13.1 Prior Distributions 120

13.2 MCMC Methods 121

13.3 SMC Methods 128

13.4 Model Evaluation 131

13.5 Bayesian Inference At Work 131

13.6 Limited Information Bayesian Inference 132

14 Conclusion **134**

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, Plosser, and Rebelo (1988) as well as New Keynesian monetary models with numerous real and nominal frictions that are based on the work of Christiano, Eichenbaum, and Evans (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, 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. The first part of this chapter discusses model solution techniques, whereas the second part is devoted to model estimation and evaluation.

2 Solution Methods

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

However, numerical analysis and computer programming are not parts of the standard curriculum for economists either at the undergraduate or the graduate level. This educational gap has brought three problems. First, many macroeconomists have been reluctant to accept that analytic results are limited and that the cavalier assumptions sometimes taken to allow for closed-form solutions often confuse the discussion more than clarify it. While there is an important role for analytic result in special cases for building intuition, for understanding the logic of economic mechanisms, and for testing the performance of our 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 design of 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, obviously, not enough for real world advise. We need to gauge the magnitude of such nominal interest rate reduction. Similarly, saying that an increase in government spending increases output does not provide enough information to design a countercyclical fiscal package.

Second, the lack of familiarity with numerical analysis has led to the slow diffusion of best practices in solution methods and to a certain lack of interest in issues such as detailed assessments of numerical errors in computations and their possible implications. For example, welfare comparisons of different policies are highly dependent on the accuracy of our solution methods (see the examples of spurious welfare reversals in Kim and Kim, 2003). Similarly, the identification of parameter values may depend on the quality of the solution (see, Binsbergen *et al.*, 2013, and Fernández-Villaverde *et al.*, 2013, for several cases involving DSGE models with recursive preferences). Although much progress in the quality of computational work has been made in the last few years, there is still room for further improvements.

Third, even within the set of state-of-the-art solution methods, researchers have sometimes been unsure about the trade-offs (for example, in terms of speed versus accuracy) involved in the choosing among different algorithms.

In this first half of the chapter, we will cover some basic ideas about solution methods for DSGE models, discuss some of trade-offs in created by the alternative algorithms, introduce concepts related with the assessment of the accuracy of the solution, and briefly mention parallel programing and the promise it brings of opening the door to the solution of a much richer class of models. Throughout the chapter we will include a number of remarks with additional material for those readers willing to dig deeper into some of the issues we deal with.

Remark 1 (Initial Solution Methods)**.** In the interest of space, 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 mentioned some 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. Kydland and Prescott (1982) exploited the fact that their economy was Pareto optimal to solve the social planner's problem instead of the recursive equilibrium of their model. To do so, they

substituted the original social planner's problem by a linear quadratic approximation to it, that is, they wrote all the constraints as a linear function of the state variables and approximated the utility function with a second-order Taylor expansion. Then they took advantage that we have plenty of efficient methods to solve linear quadratic problems to find a fast solution to their problem. King, Plosser, and Rebelo (in the widely disseminated technical appendix, not published until 2002) 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 this method below by interpreting it as a first-order perturbation. Christiano (1990) applied value function iteration to solve directly for the social planner's problem of a stochastic neoclassical growth model.

Remark 2 (Discrete versus 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 and space considerations make dealing with both cases difficult. This, however, should not be a reason for forgetting about the recent advances in the computation of DSGE models in continuous time (see Fernández-Villaverde *et al.*, 2015) or for underestimating the analytic power of continuous time. Our position is that 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 formal mathematics, in particular with the well-developed study of functional equations.

Indeed, we can cast numerous problems in macroeconomics in the form of a functional equation.[1] Let us then define a functional equation more precisely. Let J^1 and J^2 be two

[1]Much of we have to say in this chapter is not, by any means, limited to macroeconomics. Similar problems (with similar notation and solution methods) appear in other fields such as finance, industrial organization, international finance, etc.

functional spaces, $\Omega \subseteq \mathbb{R}^l$ and let $\mathcal{H} : J^1 \to J^2$ be an operator between these two spaces. A *functional equation problem* is to find a function $d : \Omega \to \mathbb{R}^m$ (where Ω is the state space) such that:

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

From equation (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 that the zero in the reals.

Examples of problems in macroeconomics that can be easily 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, and then 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 own particular 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

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

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

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:

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

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 more general applications, this would not be possible and some care is required to stay within either the equilibrium problem or the social planner's problem.

3.2 A Value Function

Under standard technical conditions (Stokey, Lucas, and Prescott, 1989), we can transform the sequential problem defined by equations (2a)-(4) into a recursive problem in terms of a value function $V(k_t, z_t)$ for the social planner. 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 use 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

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

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 our problem in terms of a value function is convenient for several reasons. First is that we have many results about the properties of value functions and the decision rules associated with it (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 to tackle the dynamic programming problem. The second reason is that, as a default, we can use value function iteration (which we would explain in section xxx below), 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:

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

for all k_t and z_t, where

$$d = V(\cdot, \cdot)$$

we see how the operator $\mathcal{H}$, a rewriting of the Bellman operator, takes the value function $V(\cdot, \cdot)$ and obtains a zero. More precisely, equation (6) is an integral equation given the presence of the expectation operator. This can lead to some non-trivial measure theory considerations that we leave aside for the moment.

3.3 Euler Equation

We outlined above a several reasons why casting the problem in terms of a value function is attractive. Unfortunately, this formulation is often difficult. As soon as the model that we are dealing with does not satisfy the two fundamental welfare theorems, we cannot easily move between the social planner's problem and the competitive equilibrium and 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.[2]

An alternative is to work directly with the set of equilibrium conditions. These include the first-order conditions for households, firms, and if specified, government, budget constraints, market clearing conditions, and laws of motion for exogenous processes. Since at the core of this approach, we will have the Euler equations for the agents in the model that encode the core of optimal behavior (with the other conditions being somewhat mechanical), this approach is commonly known as solving the Euler equation (although sometimes also referred as solving the equilibrium conditions of the models). But, regardless of the name we use, the approach is extremely general and it allows 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 equations (2a)-(4) is:

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

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

[2]See, however, xxx for examples of how to recast a non-pareto optimal economy into the mold of an associated pareto-optimal problem.

$(g^2(k_t, z_t))$ given capital, k_t, and productivity, z_t, today. Then, we can rewrite the first-order condition as:

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

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

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

for all k_t and z_t. 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 $g(\cdot)$. Mapping equations (8) and (9) into our operator $\mathcal{H}$ is straightforward:

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

for all k_t and z_t,

where:

$$d = g(\cdot, \cdot).$$

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

An additional consideration that we need to take care of in this Euler equation approach 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:

$$\lim_{t \to \infty} \beta \frac{u'(c_{t+1})}{u'(c_0)} k_t = 0$$

is satisfied. We will describe below how we build our solution methods to ensure that this is, indeed, the case.

3.4 Conditional Expectations

Note that 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 be interested in finding 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, more likely, because solving for the conditional expectation may avoid problems associated with the decision rule. For example, we could enrich the basic stochastic neoclassical growth model with additional constraints (such as a non-negative investment constraint: $k_{t+1} \geq (1-\delta)k_t$) that may 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.[3]

To see how this would work, we can set up as our unknown function:

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

where we take advantage that $\mathbb{E}_t(\cdot)$ is itself another 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 $g(\cdot, \cdot)$, we can find:

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

and

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

[3]See Fernández-Villaverde *et al.* (2014) for an example of this principle. 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 very similar to that of a conditional expectation. Once these function have been found, the rest of the endogenous variables of the model, including the nominal interest rate, can be derived without additional approximation. In particular, the ZLB is always satisfied.

This shows that 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 $g(\cdot,\cdot)$. Similarly:

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

is the value of c_{t+1} implied by the recursive application of $g(\cdot,\cdot)$.

To solve for $g(\cdot,\cdot)$, we use its definition in equation (10):

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

and write:

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

where

$$d = g\left(k_t, z_t\right).$$

3.5 The Way Forward

In this section 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 illustrate 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, were not an exhaustive list and dozens of other cases can be constructed following the same ideas.

We will move now two study the two main families of solution methods for functional equation problems: perturbation and projection methods. Both families substitute 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 an 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, thus, examples of parameters. We will call "coefficients" the numerical terms in the numerical solution. While the "parameters" usually have a clear economic interpretation associated with them, the "coefficients" will, most of time, lack such interpretation.

Remark 3 (Structural Parameters?)**.** We are being careful at not adding the adjective "structural." Here we follow Hurwicz (1962), who defined "structural" as invariant to a class of policy interventions the researcher is interested in analyzing. Many parameters of interest may not "structural" in that 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 possible 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), determining the states is a step previous to the numerical solution of the model and therefore outside the purview of this chapter. The real goal of the numerical solution would be to determine, for a degree of approximation n, the value of the coefficients θ that get the approximated solution of the model as close as possible to the exact unknown solution under an appropriate metric.

4 Perturbation

Perturbation methods -sometimes also known as asymptotic methods- build approximate solution to a DSGE model by starting from the exact solution of a particular case of the

model or from the solution of a close, related model whose solution we have access to. In their more common incarnation in macroeconomics (but not necessarily the only possible one), perturbation builds higher-order Taylor series approximations to the solution of DSGE model around its deterministic steady state. With a long tradition in physics and other natural sciences, perturbation methods were popularized in economics by Judd and Guu (1993).[4]

Recently, perturbation methods have gained much popularity among researchers over the last decade because of three reasons. First, they are accurate. Perturbation methods find an approximate solution that is inherently local, that is, that is highly accurate around the point where we take the Taylor series expansion. However, researchers have documented that perturbation often display good global properties along a wide range of state variable values (see the evidence presented in Aruoba, Fernández-Villaverde, and Rubio-Ramírez, 2006, and Caldara *et al.*, 2012). Second, the results are intuitive and easily interpretable. For example, a second order expansion includes a term that corrects for the standard deviation of the shocks that drive the dynamics of the economy. This term, which captures precautionary behavior, breaks the certainty equivalence of linear approximations that makes it difficult to talk about welfare and risk in a linearized world. Third, thanks to software such as `Dynare` and `Dynare++` (developed by Michel Juillard and a team of collaborators), higher perturbations are easy to compute even for practitioners less familiar with numerical method.[5]

More concretely, perturbation methods solve the functional equation problem:

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

by specifying an approximation 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_i of the form:

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

where $\mathbf{x}_0$ is the point around which we build our perturbation solution, $\theta_{i,0}$ is a scalar, $\theta_{i,1}$ is a n-dimensional vector that stacks additional coefficients $\theta_{i.2} - \theta_{i,n+1}$, $\theta_{i,2}$ is a $n \times n$ matrix

[4]Perturbation methods were already widely used in physics in the 19th century, but they became a key tool in science with the development of quantum mechanics in the first half of the 20th century.

[5]`Dynare` (a toolbox for Matlab) and `Dynare++` (a stand-alone application) allow the researcher to write, in a concise and intuitive language, the equilibrium conditions of a DSGE model and find a perturbation solution to it, up to third-order in `Dynare` and to an arbitrary order in `Dynare++`. See `http://www.dynare.org/`.

that stacks additional coefficients $\theta_{i,n+2} + \theta_{i,n(n+1)+2}$, and so on up to an order j. As we will describe momentarily, the coefficients $\theta_{i,l}$'s are found using implicit-function theorems.

Note that the traditional linearization approach popularized by King, Plosser, and Rebelo (2002) delivers a solution of the form:

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

In other words, linearization is nothing more than a first-order perturbation and higher-order approximations generalize it 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 all that we know from perturbation theory. But the direction of influence also goes in the other direction: we can use much of what we already know about the linearized solution to DSGE models (such as how to efficiently solve for the coefficients $\theta_{i,0}$ and $\theta_{i,1}$) to confidently apply perturbation.

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. Loglinearization, for example, approximates the solution of the model in terms of the log-deviations of the variables with respect to their steady state. We will come back to this point later in the chapter.

Remark 5 (First-Oder Perturbation and Linear-Quadratic Approximations)**.** Kydland and Prescott (1982), equivalent under linear constraints.

Before getting into technical details of how to implement perturbation methods, a simple example will clarify many of the ideas we just introduced. Also, we will briefly distinguish between regular and singular perturbations.

4.1 An Example

Let us image that we want to compute $\sqrt{26}$ without the use of a hand calculator and without resorting to some of the traditional algorithms to do so. A possibility is to write:

$$\sqrt{26} = \sqrt{25 * 1.04} = \sqrt{25} * \sqrt{1.04} = 5 * \sqrt{1.04}.$$

Now, note that $\sqrt{1.04} \approx 1.02$. Therefore, we can find:

$$\sqrt{26} \approx 5 * 1.02 = 5.1,$$

an approximated solution that we were able to find with trivial computations. The numerical error between the exact solution, $\sqrt{26} = 5.099$ and our approximation:

$$e = \left| \frac{5.1 - 5.099}{5.099} \right| = 0.00019$$

might be sufficiently small for many (but not necessarily all) applications of interest.

This idea can be easily generalized. The square root of an arbitrary number x can be written as:

$$\sqrt{x} = \sqrt{y^2 * (1 + \sigma)} = y * \sqrt{(1 + \sigma)} \approx y * \left(1 + \sqrt{\sigma}\right)$$

where y is an integer whose square is close to x and $\sigma$is a perturbation parameter such that:

$$\sigma = \frac{x}{y^2} - 1.$$

When $x = y^2$, the perturbation parameter is zero, and the approximated solution is exact. When x and y^2 are not equal but close, we can build an approximated solution $y * (1 + \sqrt{\sigma})$, whose accuracy will fall as x and y^2 get away from each other.

This example illustrates, therefore, several important points. First, that we can easily build an approximated solution to an otherwise complicated problem with a high degree of accuracy. Second, that such solution is local and that its accuracy falls as we move away from the point where the solution holds exactly.

4.2 Regular versus Singular Perturbations

Perturbations of a problem can be either regular or singular. A regular perturbation is situation where a *small* change in the problem induces a *small* change in the solution. A typical example in economics is a standard New Keynesian model. A small change, for example, in the standard deviation of the monetary policy shock will lead to a small change in the properties of the equilibrium dynamics. A singular perturbation is a situation where a *small* change in the problem induces a *large* change in the solution. An example can be 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.

Most problems involving DSGE model will result in regular perturbations and, therefore, we will concentrate now our attention on them. But this does not need to be 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, this class of problems may become more common with DSGE models.

4.3 The General Case

We present first the general case of how to find a perturbation solution of a DSGE model by 1) using the equilibrium conditions to the model and 2) by finding a higher-order Taylor series approximations. 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.

We start by noticing that the set of equilibrium conditions of the model can be written as

$$\mathbb{E}_t \mathcal{H}(\mathbf{y}, \mathbf{y}', \mathbf{x}, \mathbf{x}') = 0, \tag{12}$$

where $\mathbf{y}$ is a $n_y \times 1$ vector of controls and $\mathbf{x}$ is a $n_x \times 1$ vector of states. We define $n = n_x + n_y$. 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 $\mathcal{H}(\cdot)$: for those equilibrium conditions without expectations, the conditional expectation operator will not have any impact. Moving $\mathbb{E}_t(\cdot)$ outside $\mathcal{H}(\cdot)$ will make some of the derivations below easier to follow. Note that, to save on space, we use 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 easy to see on which variables the perturbation will have a direct effect. Thus, we partition the state vector x_t as

$$\mathbf{x} = [\mathbf{x}_1;\, \mathbf{x}_2]^t.$$

where $\mathbf{x}_1$ is a $(n_x - n_\epsilon) \times 1$ vector of endogenous state variables and x_2 is a $n_\epsilon \times 1$ vector of exogenous state variables.

4.4 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:

$$\mathcal{H}(\bar{\mathbf{y}}, \bar{\mathbf{y}}, \bar{\mathbf{x}}, \bar{\mathbf{x}}) = 0. \tag{13}$$

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 non-linear equation solver.

Remark 6 (Simplifying the solution of $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$)**.** Finding the solution $(\bar{\mathbf{x}}, \bar{\mathbf{y}})$ 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 reduced the system from being one of n equations on n unknowns into a system of $n' < n$ equations on n' unknowns. Since the complexity of solving a nonlinear system of equations grows exponentially in the dimension of the problem (see Sikorski, 1985, for some 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:

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

where the notation is the same than in section x 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, at 1:

$$\bar{l} = 1.$$

Then, we can 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:

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

where $\overline{c}$ and $\overline{w}$ are the deterministic steady state values of consumption and wage respectively. Another way to think about this procedure is to realize that it is 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 the solution $(\bar{\mathbf{x}}, \overline{\mathbf{y}})$, other good practices include to pick units that make algebraic and numerical computations easier to handle. For example, we can pick units to make output $output_t$ equal to 1 in the deterministic steady state:

$$\overline{output} = 1.$$

Again, in the context of the stochastic neoclassical growth model, we will have:

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

If, in addition, we impose $\bar{l} = 1$ as before, we can find

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

and wages:

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

Then, we can go back to the intertemporal Euler equation:

$$\frac{1}{\overline{c}} = \frac{1}{\overline{c}}\beta\left(1 + \overline{r} - \delta\right)$$

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

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

Since:

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

we get:

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

and:

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

from which:

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

With this example, we have shown that two judicious choices of units ($\overline{l} = 1$ and $\overline{output} = 1$) renders 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 advise in this remark dramatically simplifies the task in a large class of situations.

Note that this $(\bar{\mathbf{x}}, \overline{\mathbf{y}})$ is different from a fixed-point $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ of (12):

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

because in the former case we eliminate the conditional expectation operator while in the later we do not. The vector $(\widehat{\mathbf{x}}, \widehat{\mathbf{y}})$ 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).

4.5 Exogenous Stochastic Process

For the exogenous stochastic variables, we specify a stochastic process of the form:

$$\mathbf{x}_2' = \mathbf{C}(\mathbf{x}_2) + \sigma \eta_\epsilon \epsilon' \tag{14}$$

where:

1. $\mathbf{C}(\cdot)$ is a potentially non-linear function. At this level of abstraction we are not imposing much structure on $\mathbf{C}(\cdot)$, but in concrete applications we will need to impose some additional constraints. For example, it is often convenient to assume that all the eigenvalues of the Hessian matrix of $\mathbf{C}(\cdot)$ evaluated at the non-stochastic steady state $(\bar{\mathbf{x}}, \overline{\mathbf{y}})$ lie within the unit circle.

2. ϵ' contains the n_ϵ exogenous zero-mean innovations. Initially, we only assume that ϵ_{t+1} is independent and identically distributed with finite second moments, meaning that no distributional assumption is imposed and the innovations may therefore be non-Gaussian. This is denoted by $\epsilon_{t+1} \sim IID(\mathbf{0}, \mathbf{I})$. Additional moment restrictions will be imposed as needed.

3. η_ϵ is a $n_\epsilon \times n_\epsilon$ matrix that determines the variances-covariances of the innovations,

4. $\sigma \geq 0$ is a perturbation parameter that scales η.

Remark 7 (Linearity of Innovations)**.** The assumption that innovations enter linearly in (14) may appear restrictive, but it is without loss of generality because the state vector can be extended to deal with non-linearities between $\mathbf{x}_2$ and ϵ'. A popular case is to introduce innovations with time-varying volatility (see Fernández-Villaverde and Rubio-Ramírez, 2007).

Often, it will be the case that $\mathbf{C}(\cdot)$ is linear, and we can write:

$$x_2' = Cx_2 + \sigma\eta_\epsilon\epsilon'$$

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

Note that in (14) we only have one perturbation parameter even if we have a model with many different sources of innovations. The matrix η takes account of relative sizes (and comovements) of the different innovations. If we set $\sigma = 0$, we have a deterministic model.

Also, we are departing from Samuelson (1970) and Jin and Judd (2002), who impose that the innovations to this stochastic process have a bounded support. By doing so, these authors avoid some problems with the stability of the simulations coming from the perturbation solution that we will discuss below. Instead, we will discuss pruning as a superior strategy to fix these problems.

4.6 Solution of the Model

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

$$\mathbf{y} = \mathbf{g}(\mathbf{x}; \sigma), \tag{15}$$

and for the state variables

$$\mathbf{x}' = \mathbf{h}\left(\mathbf{x};\sigma\right) + \sigma\eta\epsilon', \tag{16}$$

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: control 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 flexible to capture the dynamics of a very rich class of models. Also, note that by separate states $\mathbf{x}$ and the perturbation parameter σ by a semicolon to emphasize the difference between both elements of the functions $\mathbf{g}$ and $\mathbf{h}$.

The matrix η is of order $n_x \times n_\epsilon$ and is given by:

$$\eta = \left[\begin{array}{c} \emptyset \\ \eta_\epsilon \end{array} \right]$$

where the first n_x rows come from the fact that states today determine the endogenous states tomorrow and the last n_ϵ rows from the observation that the exogenous states tomorrow depend on the states today and the innovations tomorrow.

4.7 Perturbation

The goal of perturbation is to find an approximation of the functions $\mathbf{g}\left(\cdot\right)$ and $\mathbf{h}\left(\cdot\right)$ around the deterministic steady state, $\mathbf{x}_t = \bar{\mathbf{x}}$ and $\sigma = 0$. First, note by the definition of the deterministic steady state (13) we have that

$$\overline{\mathbf{y}} = \mathbf{g}(\bar{\mathbf{x}}; 0) \tag{17}$$

and

$$\bar{\mathbf{x}} = \mathbf{h}(\bar{\mathbf{x}}; 0). \tag{18}$$

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

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

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

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

where $F_{x_i^k \sigma^j}(\mathbf{x};\sigma)$ denotes 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 $(\mathbf{x};\sigma)$. We will exploit this fact repeatedly.

4.8 First-Order Approximation

A first-order approximation looks for approximations to $\mathbf{g}$ and $\mathbf{h}$ around $(\mathbf{x};\sigma) = (\bar{\mathbf{x}};0)$ of the form:

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

where $\mathbf{g}_{\mathbf{x}}(\cdot)$ and $\mathbf{h}_{\mathbf{x}}(\cdot)$ are the gradients of $\mathbf{g}$ and $\mathbf{h}$ respectively (where the gradient includes 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 equations (17) and (18), we can write

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

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

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

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, we need to introduce the tensor notation.

Remark 8 (Tensor Notation)**.** Tensors, commonly used in physics, keep notation while we perform a perturbation at a manageable level. An n^{th}-rank tensor in a 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 element of this matrix located at the intersection of the i-th row and α-th column.
3. Thus, $[\mathcal{H}_y]^i_\alpha [\mathbf{g}_x]^\alpha_\beta [\mathbf{h}_x]^\beta_j = \sum_{\alpha=1}^{n_y} \sum_{\beta=1}^{n_x} \frac{\partial \mathcal{H}^i}{\partial y^\alpha} \frac{\partial \mathbf{g}^\alpha}{\partial x^\beta} \frac{\partial \mathbf{h}^\beta}{\partial x^j}$.
4. $[\mathcal{H}_{y'y'}]^i_{\alpha\gamma}$:
   (a) $\mathcal{H}_{y'y'}$ is a three dimensional array with n rows, n_y columns, and n_y pages.
   (b) Then $[\mathcal{H}_{y'y'}]^i_{\alpha\gamma}$ denotes the element of $\mathcal{H}_{y'y'}$ located at the intersection of row i, column α and page γ.

Note how, in particular, tensor notation eliminates the need to keep track of the sum operators.

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

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

Note that 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. Then, 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 $\mathbf{h_x}(\bar{\mathbf{x}}; 0)$. We can solve with a standard quadratic matrix equation solver.

Remark 9 (Quadratic Equation Solvers)**.** Procedures to solve quadratic systems:

1. Blanchard and Kahn (1980).
2. Uhlig (1999).

3. Sims (2000).

4. Klein (2000).[to be completed].

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

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

Then:

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

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

$$\begin{aligned}
\mathbf{g}_\sigma &= \mathbf{0} \\
\mathbf{h}_\sigma &= \mathbf{0}
\end{aligned}$$

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

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

These equations embody a certainty equivalence property as defined by Simon (1956) and Theil (1957). 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 $\sigma = 0$). The intuition for this result is simple. Risk-aversion depends on the second derivative of the utility function (concave utility). However, Leland (1968) and Sandmo (1970) showed that a sufficient condition for risk aversion to translate into precautionary behavior is the convexity of marginal

utility. In other words, we need information from 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.

Drawbacks of certainty equivalence.

4.9 Second-Order Approximation

The second-order approximations to $\mathbf{g}$ around $(\mathbf{x};\sigma) = (\bar{\mathbf{x}};0)$ is

$$\begin{aligned}
[\mathbf{g}(\mathbf{x};\sigma)]^i &= [\mathbf{g}(\bar{\mathbf{x}};0)]^i + [\mathbf{g}_x(\bar{\mathbf{x}};0)]^i_a[(\mathbf{x}-\bar{\mathbf{x}})]_a + [\mathbf{g}_\sigma(\bar{\mathbf{x}};0)]^i[\sigma] \\
&+\frac{1}{2}[\mathbf{g}_{xx}(\bar{\mathbf{x}};0)]^i_{ab}[(\mathbf{x}-\bar{\mathbf{x}})]_a[(\mathbf{x}-\bar{\mathbf{x}})]_b \\
&+\frac{1}{2}[\mathbf{g}_{x\sigma}(\bar{\mathbf{x}};0)]^i_a[(\mathbf{x}-\bar{\mathbf{x}})]_a[\sigma] \\
&+\frac{1}{2}[\mathbf{g}_{\sigma x}(\bar{\mathbf{x}};0)]^i_a[(\mathbf{x}-\bar{\mathbf{x}})]_a[\sigma] \\
&+\frac{1}{2}[\mathbf{g}_{\sigma\sigma}(\bar{\mathbf{x}};0)]^i[\sigma][\sigma]
\end{aligned}$$

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

$$\begin{aligned}
[\mathbf{h}(\mathbf{x};\sigma)]^j &= [\mathbf{h}(\bar{\mathbf{x}};0)]^j + [\mathbf{h}_x(\bar{\mathbf{x}};0)]^j_a[(\mathbf{x}-\bar{\mathbf{x}})]_a + [\mathbf{h}_\sigma(\bar{\mathbf{x}};0)]^j[\sigma] \\
&+\frac{1}{2}[\mathbf{h}_{xx}(\bar{\mathbf{x}};0)]^j_{ab}[(\mathbf{x}-\bar{\mathbf{x}})]_a[(\mathbf{x}-\bar{\mathbf{x}})]_b \\
&+\frac{1}{2}[\mathbf{h}_{x\sigma}(\bar{\mathbf{x}};0)]^j_a[(\mathbf{x}-\bar{\mathbf{x}})]_a[\sigma] \\
&+\frac{1}{2}[\mathbf{h}_{\sigma x}(\bar{\mathbf{x}};0)]^j_a[(\mathbf{x}-\bar{\mathbf{x}})]_a[\sigma] \\
&+\frac{1}{2}[\mathbf{h}_{\sigma\sigma}(\bar{\mathbf{x}};0)]^j[\sigma][\sigma],
\end{aligned}$$

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

The unknowns of these expansions are $[\mathbf{g}_{xx}]^i_{ab}$, $[\mathbf{g}_{x\sigma}]^i_a$, $[\mathbf{g}_{\sigma x}]^i_a$, $[\mathbf{g}_{\sigma\sigma}]^i$, $[\mathbf{h}_{xx}]^j_{ab}$, $[\mathbf{h}_{x\sigma}]^j_a$, $[\mathbf{h}_{\sigma x}]^j_a$, $[\mathbf{h}_{\sigma\sigma}]^j$. These coefficients can be identified by taking the derivative of $F(x;\sigma)$ with respect to x and σ twice and evaluating them at $(x;\sigma) = (\bar{x};0)$. By the arguments provided earlier, these derivatives must be zero.

The question is now how to solve the system. First, we use $F_{xx}(\bar{\mathbf{x}};0)$ to identify $g_{xx}(\bar{\mathbf{x}};0)$ and $h_{xx}(\bar{\mathbf{x}};0)$:

$$
\begin{aligned}
&[F_{xx}(\bar{\mathbf{x}};0)]^i_{jk} = \\
&\left([\mathcal{H}_{y'y'}]^i_{\alpha\gamma}[g_x]^\gamma_\delta[h_x]^\delta_k + [\mathcal{H}_{y'y}]^i_{\alpha\gamma}[g_x]^\gamma_k + [\mathcal{H}_{y'x'}]^i_{\alpha\delta}[h_x]^\delta_k + [\mathcal{H}_{y'x}]^i_{\alpha k}\right)[g_x]^\alpha_\beta[h_x]^\beta_j \\
&+[\mathcal{H}_{y'}]^i_\alpha[g_{xx}]^\alpha_{\beta\delta}[h_x]^\delta_k[h_x]^\beta_j + [\mathcal{H}_{y'}]^i_\alpha[g_x]^\alpha_\beta[h_{xx}]^\beta_{jk} \\
&+\left([\mathcal{H}_{yy'}]^i_{\alpha\gamma}[g_x]^\gamma_\delta[h_x]^\delta_k + [\mathcal{H}_{yy}]^i_{\alpha\gamma}[g_x]^\gamma_k + [\mathcal{H}_{yx'}]^i_{\alpha\delta}[h_x]^\delta_k + [\mathcal{H}_{yx}]^i_{\alpha k}\right)[g_x]^\alpha_j \\
&+[\mathcal{H}_y]^i_\alpha[g_{xx}]^\alpha_{jk} \\
&+\left([\mathcal{H}_{x'y'}]^i_{\beta\gamma}[g_x]^\gamma_\delta[h_x]^\delta_k + [\mathcal{H}_{x'y}]^i_{\beta\gamma}[g_x]^\gamma_k + [\mathcal{H}_{x'x'}]^i_{\beta\delta}[h_x]^\delta_k + [\mathcal{H}_{x'x}]^i_{\beta k}\right)[h_x]^\beta_j \\
&+[\mathcal{H}_{x'}]^i_\beta[h_{xx}]^\beta_{jk} \\
&+[\mathcal{H}_{xy'}]^i_{j\gamma}[g_x]^\gamma_\delta[h_x]^\delta_k + [\mathcal{H}_{xy}]^i_{j\gamma}[g_x]^\gamma_k + [\mathcal{H}_{xx'}]^i_{j\delta}[h_x]^\delta_k + [\mathcal{H}_{xx}]^i_{jk} = 0; \\
&i = 1,\ldots n, \quad j,k,\beta,\delta = 1,\ldots n_x; \quad \alpha,\gamma = 1,\ldots n_y.
\end{aligned}
$$

We know the derivatives of $\mathcal{H}$. We also know the first derivatives of g and h evaluated at $(\mathbf{y},\mathbf{y}',\mathbf{x},\mathbf{x}') = (\bar{\mathbf{y}},\bar{\mathbf{y}},\bar{\mathbf{x}},\bar{\mathbf{x}})$. Hence, the above expression represents a system of $n \times n_x \times n_x$ linear equations in then $n \times n_x \times n_x$ unknowns elements of g_{xx} and h_{xx}.

Similarly, $g_{\sigma\sigma}$ and $h_{\sigma\sigma}$ can be obtained by solving:

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

a system of n linear equations in the n unknowns given by the elements of $g_{\sigma\sigma}$ and $h_{\sigma\sigma}$.

Note the cross derivatives $g_{x\sigma}$ and $h_{x\sigma}$ are zero when evaluated at $(\bar{x},0)$.

- Why? Write the system $F_{\sigma x}(\bar{x};0) = 0$ taking into account that all terms containing either g_σ or h_σ are zero at $(\bar{x},0)$.

- Then:

$$[F_{\sigma x}(\bar{x};0)]^i_j = [\mathcal{H}_{y'}]^i_\alpha [g_x]^\alpha_\beta [h_{\sigma x}]^\beta_j + [\mathcal{H}_{y'}]^i_\alpha [g_{\sigma x}]^\alpha_\gamma [h_x]^\gamma_j +$$
$$[\mathcal{H}_y]^i_\alpha [g_{\sigma x}]^\alpha_j + [\mathcal{H}_{x'}]^i_\beta [h_{\sigma x}]^\beta_j = 0;$$
$$i = 1, \ldots n; \quad \alpha = 1, \ldots, n_y; \quad \beta, \gamma, j = 1, \ldots, n_x.$$

- This is a system of $n \times n_x$ equations in the $n \times n_x$ unknowns given by the elements of $g_{\sigma x}$ and $h_{\sigma x}$.
- The system is homogeneous in the unknowns.
- Thus, if a unique solution exists, it is given by:

$$\begin{aligned} g_{\sigma x} &= 0 \\ h_{\sigma x} &= 0 \end{aligned}$$

Structure of the Solution

- The perturbation solution of the model satisfies:

$$\begin{aligned} g_\sigma(\bar{x};0) &= 0 \\ h_\sigma(\bar{x};0) &= 0 \\ g_{x\sigma}(\bar{x};0) &= 0 \\ h_{x\sigma}(\bar{x};0) &= 0 \end{aligned}$$

- Standard deviation only appears in:
  1. A constant term given by $\frac{1}{2} g_{\sigma\sigma} \sigma^2$ for the control vector y_t.
  2. The first $n_x - n_\epsilon$ elements of $\frac{1}{2} h_{\sigma\sigma} \sigma^2$.
- Correction for risk.
- Quadratic terms in endogenous state vector x_1.
- Those terms capture non-linear behavior.

4.10 Higher-Order Approximations

- We can iterate this procedure as many times as we want.
- We can obtain n-th order approximations.
- Problems:
  1. Existence of higher order derivatives (Santos, 1992).
  2. Numerical instabilities.
  3. Computational costs.

4.11 An Example

Since the previous derivations are somewhat abstract, 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 equations (2a)-(4). The only difference is that, to make the algebra easier to follow, we assume that:

$$u(c) = \log c.$$

The equilibrium conditions of the model are then:

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

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

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

The decision rules we are searching for are:

$$c_t = c(k_t, z_t; \sigma)$$

and

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

- One particular case the model has a closed form solution: $\delta = 1$.
- Why? Because, the income and the substitution effect from a productivity shock cancel each other:

1. Labor is constant.
2. Consumption is a fixed fraction of income.

- Not very realistic, but we are trying to learn here.

Solution

- Equilibrium conditions with $\delta = 1$.

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

- By "guess and verify"

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

- How can you check? Plug the solution in the equilibrium conditions.

Another Way to Solve the Problem

- Now let us suppose that you missed the lecture when "guess and verify" was explained.
- You need to compute the RBC.
- What you are searching for? A decision rule for consumption:

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

and another one for capital:

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

Note that our d is just the stack of $c(k_t, z_t)$ and $k(k_t, z_t)$.

Equilibrium Conditions

- We substitute in the equilibrium conditions the budget constraint and the law of motion for technology.
- And we write the decision rules explicitly as function of the states.
- Then, we have:

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

- Hence, we want to transform the problem.
- Which perturbation parameter? Standard deviation σ.
- Why σ? Discrete versus continuous time.
- Set $\sigma = 0$ $\Rightarrow$deterministic model, $z_t = 0$ and $e^{z_t} = 1$.
- We know how to solve the deterministic steady state.

A Parametrized Decision Rule

- We search for decision rule:

$$c_t = c\left(k_t, z_t; \sigma\right)$$

  and

$$k_{t+1} = k\left(k_t, z_t; \sigma\right)$$

- Note new parameter σ.
- We are building a local approximation around $\sigma = 0$.

Taylor's Theorem

- Equilibrium conditions:

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

- We will take derivatives with respect to k_t, z_t, and σ.
- Apply Taylor's theorem to build solution around deterministic steady state. How?

Asymptotic Expansion I

$$\begin{aligned} c_t &= c\left(k_t, z_t; \sigma\right)|_{k,0,0} = c\left(k, 0; 0\right) \\ &+ c_k\left(k, 0; 0\right)\left(k_t - k\right) + c_z\left(k, 0; 0\right) z_t + c_\sigma\left(k, 0; 0\right)\sigma \\ &+ \frac{1}{2} c_{kk}\left(k, 0; 0\right)\left(k_t - k\right)^2 + \frac{1}{2} c_{kz}\left(k, 0; 0\right)\left(k_t - k\right) z_t \\ &+ \frac{1}{2} c_{k\sigma}\left(k, 0; 0\right)\left(k_t - k\right)\sigma + \frac{1}{2} c_{zk}\left(k, 0; 0\right) z_t\left(k_t - k\right) \\ &+ \frac{1}{2} c_{zz}\left(k, 0; 0\right) z_t^2 + \frac{1}{2} c_{z\sigma}\left(k, 0; 0\right) z_t \sigma \\ &+ \frac{1}{2} c_{\sigma k}\left(k, 0; 0\right)\sigma\left(k_t - k\right) + \frac{1}{2} c_{\sigma z}\left(k, 0; 0\right)\sigma z_t \\ &+ \frac{1}{2} c_{\sigma^2}\left(k, 0; 0\right)\sigma^2 + ... \end{aligned}$$

Asymptotic Expansion II

$$\begin{aligned} k_{t+1} &= k\left(k_t, z_t; \sigma\right)|_{k,0,0} = k\left(k, 0; 0\right) \\ &+ k_k\left(k, 0; 0\right)\left(k_t - k\right) + k_z\left(k, 0; 0\right) z_t + k_\sigma\left(k, 0; 0\right)\sigma \\ &+ \frac{1}{2} k_{kk}\left(k, 0; 0\right)\left(k_t - k\right)^2 + \frac{1}{2} k_{kz}\left(k, 0; 0\right)\left(k_t - k\right) z_t \\ &+ \frac{1}{2} k_{k\sigma}\left(k, 0; 0\right)\left(k_t - k\right)\sigma + \frac{1}{2} k_{zk}\left(k, 0; 0\right) z_t\left(k_t - k\right) \\ &+ \frac{1}{2} k_{zz}\left(k, 0; 0\right) z_t^2 + \frac{1}{2} k_{z\sigma}\left(k, 0; 0\right) z_t \sigma \\ &+ \frac{1}{2} k_{\sigma k}\left(k, 0; 0\right)\sigma\left(k_t - k\right) + \frac{1}{2} k_{\sigma z}\left(k, 0; 0\right)\sigma z_t \\ &+ \frac{1}{2} k_{\sigma^2}\left(k, 0; 0\right)\sigma^2 + ... \end{aligned}$$

Comment on Notation

- From now on, to save on notation, I will write

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

- Note that:

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

- I 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.

Zero-Order Approximation

- First, we evaluate $\sigma = 0$:

$$F\left(k_t, 0; 0\right) = 0$$

- Steady state:

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

or,

$$1 = \alpha \beta k^{\alpha - 1}$$

- Then:

$$c = c\left(k, 0; 0\right) = (\alpha\beta)^{\frac{\alpha}{1-\alpha}} - (\alpha\beta)^{\frac{1}{1-\alpha}}$$
$$k = k\left(k, 0; 0\right) = (\alpha\beta)^{\frac{1}{1-\alpha}}$$

First-Order Approximation

- We take derivatives of $F\left(k_t, z_t; \sigma\right)$ around $k, 0$, and 0.

- With respect to k_t:

$$F_k\left(k, 0; 0\right) = 0$$

- With respect to z_t:

$$F_z\left(k,0;0\right)=0$$

- With respect to σ:

$$F_\sigma\left(k,0;0\right)=0$$

Solving the System I

- Remember that:

$$F\left(k_t,z_t;\sigma\right)=$$
$$\mathcal{H}\left(c\left(k_t,z_t;\sigma\right),c\left(k\left(k_t,z_t;\sigma\right),z_{t+1};\sigma\right),k_t,k\left(k_t,z_t;\sigma\right),z_t;\sigma\right)=0$$

- Because $F\left(k_t,z_t;\sigma\right)$ must be equal to zero for any possible values of k_t,z_t, and σ, the derivatives of any order of F must also be zero.

- Then:

$$\begin{aligned}F_k\left(k,0;0\right)&=\mathcal{H}_1c_k+\mathcal{H}_2c_kk_k+\mathcal{H}_3+\mathcal{H}_4k_k=0\\F_z\left(k,0;0\right)&=\mathcal{H}_1c_z+\mathcal{H}_2\left(c_kk_z+c_k\rho\right)+\mathcal{H}_4k_z+\mathcal{H}_5=0\\F_\sigma\left(k,0;0\right)&=\mathcal{H}_1c_\sigma+\mathcal{H}_2\left(c_kk_\sigma+c_\sigma\right)+\mathcal{H}_4k_\sigma+\mathcal{H}_6=0\end{aligned}$$

Solving the System II

- Note that:

$$\begin{aligned}F_k\left(k,0;0\right)&=\mathcal{H}_1c_k+\mathcal{H}_2c_kk_k+\mathcal{H}_3+\mathcal{H}_4k_k=0\\F_z\left(k,0;0\right)&=\mathcal{H}_1c_z+\mathcal{H}_2\left(c_kk_z+c_k\rho\right)+\mathcal{H}_4k_z+\mathcal{H}_5=0\end{aligned}$$

  is a quadratic system of four equations on four unknowns: c_k, c_z, k_k, and k_z.

- All of them equivalent.

- Why quadratic? Stable and unstable manifold.

Solving the System III

- Also, note that:

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

  is a linear, and homogeneous system in c_σ and k_σ.

- Hence:

$$c_\sigma = k_\sigma = 0$$

- This means the system is certainty equivalent.

4.12 Prunning

Although the higher-order approximations that we just describe 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 **?** and **?** for illustrative examples). The presence of 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.

4.13 Change of Variables

Here we can cite the idea of changing variables (Fernández-Villaverde and Rubio-Ramírez, 2006). Instead of writing a Taylor expansion in terms of a variable x:

$$f(x) \simeq f(a) + f'(a)(x-a) + H.O.T.$$

we can write it in terms of a transformed variable $Y(x)$:

$$g(y) = h(f(X(y))) = g(b) + g'(b)(Y(x) - b) + H.O.T.$$

where $b = Y(a)$ and $X(y)$ is the inverse of $Y(x)$. By picking the right change of variables, we can significantly increase the accuracy of the perturbation. A common example of change

of variables (although rarely thought of in this way) is to loglinearize instead of linearizing in levels.

A first order perturbation produces an approximated policy function in levels of the form:[6]

$$\begin{aligned}(k' - k_0) &= a_1 (k - k_0) + b_1 z \\ (l - l_0) &= c_1 (k - k_0) + d_1 z\end{aligned}$$

where k and z are the current states of the economy, l_0 is steady state value for labor and where for convenience we have dropped the subscript p where no ambiguity exists.[7] Analogously a loglinear approximation of the policy function will take the form:

$$\begin{aligned}\log k' - \log k_0 &= a_2 (\log k - \log k_0) + b_2 z \\ \log l - \log l_0 &= c_2 (\log k - \log k_0) + d_2 z\end{aligned}$$

or in equivalent notation:

$$\begin{aligned}\widehat{k'} &= a_2 \widehat{k} + b_2 z \\ \widehat{l} &= c_2 \widehat{k} + d_2 z\end{aligned}$$

where $\widehat{x} = \log x - \log x_0$ 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 follow Judd's (2003) notation and write the linear system in levels as:

$$\begin{aligned}k'_p(k, z, \sigma) &= f^1(k, z, \sigma) = f^1(k_0, 0, 0) + f^1_1(k_0, 0, 0)(k - k_0) + f^1_2(k_0, 0, 0) z \\ l_p(k, z, \sigma) &= f^2(k, z, \sigma) = f^2(k_0, 0, 0) + f^2_1(k_0, 0, 0)(k - k_0) + f^2_2(k_0, 0, 0) z\end{aligned}$$

where:

[6]See Uhlig (1999) for details. Remember that this solution is the same as the one generated by a Linear Quadratic approximation of the utility function (Kydland and Prescott, 1982), the Eigenvalue Decomposition (Blanchard and Kahn, 1980 and King, Plosser and Rebelo, 2002), Generalized Schur Decomposition (Klein, 2000) or the QZ decomposition (Sims, 2002b) among others. Subject to applicability, all methods need to find the same policy functions since the linear space approximating a nonlinear space is unique.

[7]It can be shown that the coefficients on σ are zero in the first order perturbation.

$f^1(k_0,0,0)=k_0$	$f_1^1(k_0,0,0)=a_1$	$f_2^1(k_0,0,0)=b_1$
$f^2(k_0,0,0)=l_0$	$f_1^2(k_0,0,0)=c_1$	$f_2^2(k_0,0,0)=d_1$

Second we propose the changes of variables:

$h^1=\log f^1$	$Y^1(x)=\log x^1$	$X^1=\exp y^1$
$h^2=\log f^2$	$Y^2=x^2$	$X^2=y^2$

Judd's (2003) formulae for this particular example imply:

$$\begin{pmatrix} \log k'(\log k,z) \\ \log l(\log k,z) \end{pmatrix} = g(\log k,z) = \begin{pmatrix} \log f^1(k_0,0,0) \\ \log f^2(k_0,0,0) \end{pmatrix} +$$

$$\begin{pmatrix} \log k-\log k_0 & z-z_0 \\ \log k-\log k_0 & z-z_0 \end{pmatrix} \begin{pmatrix} \frac{1}{k_0}\begin{pmatrix} f_1^1(k_0,0,0) & f_2^1(k_0,0,0) \end{pmatrix}\begin{pmatrix} k_0 \\ 1 \end{pmatrix} \\ \frac{1}{l_0}\begin{pmatrix} f_1^2(k_0,0,0) & f_2^2(k_0,0,0) \end{pmatrix}\begin{pmatrix} k_0 \\ 1 \end{pmatrix} \end{pmatrix},$$

and thus:

$$\begin{aligned} \log k'-\log k_0 &= f_1^1(k_0,0,0)(\log k-\log k_0)+\frac{1}{k_0}f_2^1(k_0,0,0)z \\ \log l-\log l_0 &= \frac{k_0}{l_0}f_1^2(k_0,0,0)(\log k-\log k_0)+\frac{1}{l_0}f_2^2(k_0,0,0)z \end{aligned}$$

We equating coefficients we obtain a nice and simple closed-form relation between the parameters of both representations:[8]

[8] An alternative heuristic argument that delivers the same result is as follows. Take the system

$$\begin{aligned} (k'-k_0) &= a_1(k-k_0)+b_1z \\ (l-l_0) &= c_1(k-k_0)+d_1z \end{aligned}$$

and divide on both sides by the steady state value of the control variable:

$$\begin{aligned} \frac{k'-k_0}{k_0} &= a_1\frac{k-k_0}{k_0}+\frac{1}{k_0}b_1z \\ \frac{l-l_0}{l_0} &= c_1\frac{k-k_0}{l_0}+\frac{1}{l_0}d_1z \end{aligned}$$

and noticing that $\frac{x'-x_0}{x_0}\simeq\log x-\log x_0$ we get back the same relation that the one presented in the paper. Of course our argument is more general and does not depend on an additional approximation.

$a_2 = a_1$	$b_2 = \frac{1}{k_0} b_1$
$c_2 = \frac{k_0}{l_0} c_1$	$d_2 = \frac{1}{l_0} d_1$

Note that we have not used any assumption on the utility or production functions except that they satisfy the general technical conditions of the neoclassical growth model. Also moving from one coefficient set to the other one is an operation that only involves k_0 and l_0, values that we need to find anyway to compute the linearized version in levels. Therefore, once you have the linear solution, obtaining the loglinear one is immediate.

4.14 The Optimal Change of Variables

In the last section we showed how to find a loglinear approximation to the solution of the neoclassical growth model directly from its linear representation. Now we use the same approach to generalize our result to encompass the relationship between any power function approximation and the linear coefficients of the policy function. Also we search for the optimal change of variable inside this class of power functions and we report how the Euler equation errors improve with respect to the linear representation.

Before we argued that some practitioners have defended the use of loglinearizations to capture some of the nonlinearities in the data. This practice can be push one step ahead. We can generalize the log function into a general class of power function of the form:

$$\begin{aligned} k_p' \left(k, z; \gamma, \zeta, \mu, \varphi\right)^{\gamma} - k_0^{\gamma} &= a_3 \left(k^{\zeta} - k_0^{\zeta}\right) + b_3 z^{\varphi} \\ l_p \left(k, z; \gamma, \zeta, \mu, \varphi\right)^{\mu} - l_0^{\mu} &= c_3 \left(k^{\zeta} - k_0^{\zeta}\right) + d_3 z^{\varphi} \end{aligned}$$

with $\varphi \geq 1$. The last constraint assures that we will have real values for the power z^{φ}.

This class of functions is attractive because it provides a lot of flexibility in shapes with few free parameters while including the log transformation as the limit case when the coefficients γ, ζ and μ tend to zero and φ is equal to 1. Also a similar power function with only two parameters is proposed by Judd (2003) in a simple optimal growth model without leisure and stochastic perturbations. His finding of notable improvements in the accuracy of the solution when he optimally selects the value of these parameters is suggestive of the advantages of using this parametric family.

The changes of variables for this family of functions are given by :

$h^1 = (f^1)^\gamma$	$Y^1 = (x^1)^\zeta$	$X_1 = (y^1)^{\frac{1}{\zeta}}$
$h^2 = (f^1)^\mu$	$Y_2 = (x^2)^\varphi$	$X_2 = (y^2)^{\frac{1}{\varphi}}$

Following the same reasoning than in the previous section we derive a form for the system in term of the original coefficients:

$$\begin{aligned} k_p' \left(k, z; \gamma, \zeta, \mu, \varphi\right)^\gamma - k_0^\gamma &= \frac{\gamma}{\zeta} k_0^{\gamma-\zeta} a_1 \left(k^\zeta - k_0^\zeta\right) + \frac{\gamma}{\varphi} k_0^{\gamma-1} b_1 z^\varphi \\ l_p \left(k, z; \gamma, \zeta, \mu, \varphi\right)^\mu - l_0^\mu &= \frac{\mu}{\zeta} l_0^{\mu-1} k_0^{1-\zeta} c_1 \left(k^\zeta - k_0^\zeta\right) + \frac{\mu}{\varphi} l_0^{\mu-1} d_1 z^\varphi \end{aligned}$$

Therefore, the relation of between the new and the old coefficients is again very simple to compute:

$a_3 = \frac{\gamma}{\zeta} k_0^{\gamma-\zeta} a_1$	$b_3 = \frac{\gamma}{\varphi} k_0^{\gamma-1} b_1$
$c_3 = \frac{\mu}{\zeta} l_0^{\mu-1} k_0^{1-\zeta} c_1$	$d_3 = \frac{\mu}{\varphi} l_0^{\mu-1} d_1$

As we pointed out before when γ, ζ and μ tend to zero and φ is equal to 1 we get back the transformation derived in the previous section to move from the linear into the loglinear solution of the model.

4.15 Perturbing the Value Function

- We worked with the equilibrium conditions of the model.
- Sometimes we may want to perform a perturbation on the value function formulation of the problem.
- Possible reasons:
  1. Gain insight.
  2. Difficulty in using equilibrium conditions.
  3. Evaluate welfare.
  4. Initial guess for VFI.
- More general point: we can perturb any operator problem that we find useful.

Basic Problem

- Imagine that we have:

$$V\left(k_t, z_t\right) = \max_{c_t}\left[\left(1-\beta\right)\frac{c_t^{1-\gamma}}{1-\gamma} + \beta\mathbb{E}_t V\left(k_{t+1}, z_{t+1}\right)\right]$$
$$\text{s.t. } c_t + k_{t+1} = e^{z_t}k_t^{\theta} + \left(1-\delta\right)k_t$$
$$z_t = \lambda z_{t-1} + \sigma\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}\left(0,1\right)$$

- Write it as:

$$V\left(k_t, z_t; \chi\right) = \max_{c_t}\left[\left(1-\beta\right)\frac{c_t^{1-\gamma}}{1-\gamma} + \beta\mathbb{E}_t V\left(k_{t+1}, z_{t+1}; \chi\right)\right]$$
$$\text{s.t. } c_t + k_{t+1} = e^{z_t}k_t^{\theta} + \left(1-\delta\right)k_t$$
$$z_t = \lambda z_{t-1} + \chi\sigma\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}\left(0,1\right)$$

Alternative

- Another way to write the value function is:

$$V\left(k_t, z_t; \chi\right) =$$
$$\max_{c_t}\left[\begin{array}{c}\left(1-\beta\right)\frac{c_t^{1-\gamma}}{1-\gamma} + \\ \beta\mathbb{E}_t V\left(e^{z_t}k_t^{\theta} + \left(1-\delta\right)k_t - c_t, \lambda z_t + \chi\sigma\varepsilon_{t+1}; \chi\right)\end{array}\right]$$

- This form makes the dependences in the next period states explicit.

- The solution of this problem is value function $V\left(k_t, z_t; \chi\right)$ and a policy function for consumption $c\left(k_t, z_t; \chi\right)$.

Expanding the Value Function

The second-order Taylor approximation of the value function around the deterministic steady state $\left(k_{ss}, 0; 0\right)$ is:

$$V\left(k_t, z_t; \chi\right) \simeq$$
$$V_{ss} + V_{1,ss}\left(k_t - k_{ss}\right) + V_{2,ss}z_t + V_{3,ss}\chi$$
$$+\frac{1}{2}V_{11,ss}\left(k_t - k_{ss}\right)^2 + \frac{1}{2}V_{12,ss}\left(k_t - k_{ss}\right)z_t + \frac{1}{2}V_{13,ss}\left(k_t - k_{ss}\right)\chi$$
$$+\frac{1}{2}V_{21,ss}z_t\left(k_t - k_{ss}\right) + \frac{1}{2}V_{22,ss}z_t^2 + \frac{1}{2}V_{23,ss}z_t\chi$$
$$+\frac{1}{2}V_{31,ss}\chi\left(k_t - k_{ss}\right) + \frac{1}{2}V_{32,ss}\chi z_t + \frac{1}{2}V_{33,ss}\chi^2$$

where

$$\begin{aligned} V_{ss} &= V(k_{ss}, 0; 0) \\ V_{i,ss} &= V_i(k_{ss}, 0; 0) \text{ for } i = \{1, 2, 3\} \\ V_{ij,ss} &= V_{ij}(k_{ss}, 0; 0) \text{ for } i, j = \{1, 2, 3\} \end{aligned}$$

- By certainty equivalence, we will show below that:

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

- Taking advantage of the equality of cross-derivatives, and setting $\chi = 1$, which is just a normalization:

$$\begin{aligned} V(k_t, z_t; 1) \simeq\ & V_{ss} + V_{1,ss}(k_t - k_{ss}) + V_{2,ss} z_t \\ & + \frac{1}{2} V_{11,ss}(k_t - k_{ss})^2 + \frac{1}{2} V_{22,ss} z_{tt}^2 \\ & + V_{12,ss}(k_t - k_{ss}) z + \frac{1}{2} V_{33,ss} \end{aligned}$$

- Note that $V_{33,ss} \neq 0$, a difference from the standard linear-quadratic approximation to the utility functions.

Expanding the Consumption Function

- The policy function for consumption can be expanded as:

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

where:

$$\begin{aligned} c_{1,ss} &= c_1(k_{ss}, 0; 0) \\ c_{2,ss} &= c_2(k_{ss}, 0; 0) \\ c_{3,ss} &= c_3(k_{ss}, 0; 0) \end{aligned}$$

- Since the first derivatives of the consumption function only depend on the first and second derivatives of the value function, we must have that $c_{3,ss} = 0$ (precautionary consumption depends on the third derivative of the value function, Kimball, 1990).

Linear Components of the Value Function

- To find the linear 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 χ.
- Notation:
  1. $V_{i,t}$: derivative of the value function with respect to its i-th argument, evaluated in $(k_t, z_t; \chi)$.
  2. $V_{i,ss}$: derivative evaluated in the steady state, $(k_{ss}, 0; 0)$.
  3. We follow the same notation for higher-order (cross-) derivatives.

Derivatives

- Derivative with respect to c_t:

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

- Derivative with respect to k_t:

$$V_{1,t} = \beta \mathbb{E}_t V_{1,t+1} \left(\theta e^{z_t} k_t^{\theta-1} + 1 - \delta\right)$$

- Derivative with respect to z_t:

$$V_{2,t} = \beta \mathbb{E}_t \left[V_{1,t+1} e^{z_t} k_t^{\theta} + V_{2,t+1}\lambda\right]$$

- Derivative with respect to χ:

$$V_{3,t} = \beta \mathbb{E}_t \left[V_{2,t+1}\sigma\varepsilon_{t+1} + V_{3,t+1}\right]$$

- In the last three derivatives, we apply the envelope theorem to eliminate the derivatives of consumption with respect to k_t, z_t, and χ.

System of Equations I

Now, we have the system:

$$c_t + k_{t+1} = e^{z_t} k_t^{\theta} + (1-\delta) k_t$$
$$V(k_t, z_t; \chi) = (1-\beta) \frac{c_t^{1-\gamma}}{1-\gamma} + \beta \mathbb{E}_t V(k_{t+1}, z_{t+1}; \chi)$$
$$(1-\beta) c_t^{-\gamma} - \beta \mathbb{E}_t V_{1,t+1} = 0$$
$$V_{1,t} = \beta \mathbb{E}_t V_{1,t+1} \left(\theta e^{z_t} k_t^{\theta-1} + 1 - \delta\right)$$
$$V_{2,t} = \beta \mathbb{E}_t \left[V_{1,t+1} e^{z_t} k_t^{\theta} + V_{2,t+1} \lambda\right]$$
$$V_{3,t} = \beta \mathbb{E}_t \left[V_{2,t+1} \sigma \varepsilon_{t+1} + V_{3,t+1}\right]$$
$$z_t = \lambda z_{t-1} + \chi \sigma \varepsilon_t$$

If we set $\chi = 0$ and compute the steady state, we get a system of six equations on six unknowns, c_{ss}, k_{ss}, V_{ss}, $V_{1,ss}$, $V_{2,ss}$, and $V_{3,ss}$:

$$c_{ss} + \delta k_{ss} = k_{ss}^{\theta}$$
$$V_{ss} = (1-\beta) \frac{c_{ss}^{1-\gamma}}{1-\gamma} + \beta V_{ss}$$
$$(1-\beta) c_{ss}^{-\gamma} - \beta V_{1,ss} = 0$$
$$V_{1,ss} = \beta V_{1,ss} \left(\theta k_{ss}^{\theta-1} + 1 - \delta\right)$$
$$V_{2,ss} = \beta \left[V_{1,ss} k_{ss}^{\theta} + V_{2,ss} \lambda\right]$$
$$V_{3,ss} = \beta V_{3,ss}$$

- From the last equation: $V_{3,ss} = 0$.
- From the second equation: $V_{ss} = \frac{c_{ss}^{1-\gamma}}{1-\gamma}$.
- From the third equation: $V_{1,ss} = \frac{1-\beta}{\beta} c_{ss}^{-\gamma}$.
- After cancelling redundant terms:

$$c_{ss} + \delta k_{ss} = k_{ss}^{\theta}$$
$$1 = \beta \left(\theta k_{ss}^{\theta-1} + 1 - \delta\right)$$
$$V_{2,ss} = \beta \left[V_{1,ss} k_{ss}^{\theta} + V_{2,ss} \lambda\right]$$

- Then:

$$k_{ss} = \left[\frac{1}{\theta}\left(\frac{1}{\beta} - 1 + \delta\right)\right]^{\frac{1}{\theta-1}}$$
$$c_{ss} = k_{ss}^{\theta} - \delta k_{ss}$$
$$V_{2,ss} = \frac{1-\beta}{1-\beta\lambda} k_{ss}^{\theta} c_{ss}^{-\gamma}$$

- $V_{1,ss} > 0$ and $V_{2,ss} > 0$, as predicted by theory.

Quadratic Components of the Value Function

From the previous derivations, we have:

$$(1-\beta)\, c\left(k_t, z_t; \chi\right)^{-\gamma} - \beta \mathbb{E}_t V_{1,t+1} = 0$$
$$V_{1,t} = \beta \mathbb{E}_t V_{1,t+1}\left(\theta e^{z_t} k_t^{\theta-1} + 1 - \delta\right)$$
$$V_{2,t} = \beta \mathbb{E}_t\left[V_{1,t+1} e^{z_t} k_t^{\theta} + V_{2,t+1}\lambda\right]$$
$$V_{3,t} = \beta \mathbb{E}_t\left[V_{2,t+1}\sigma\varepsilon_{t+1} + V_{3,t+1}\right]$$

where:

$$\begin{aligned} k_{t+1} &= e^{z_t} k_t^{\theta} + (1-\delta)\, k_t - c\left(k_t, z_t; \chi\right) \\ z_t &= \lambda z_{t-1} + \chi\sigma\varepsilon_t,\ \varepsilon_t \sim \mathcal{N}(0,1) \end{aligned}$$

- We will now take derivatives of each of the four equations with respect to k_t, z_t, and χ.
- We will take advantage of the equality of cross derivatives.
- The envelope theorem does not hold anymore (we are taking derivatives of the derivatives of the value function).

First Equation I

We have:

$$(1-\beta)\, c\left(k_t, z_t; \chi\right)^{-\gamma} - \beta \mathbb{E}_t V_{1,t+1} = 0$$

- Derivative with respect to k_t:

$$\begin{aligned}&-(1-\beta)\gamma c\left(k_t, z_t; \chi\right)^{-\gamma-1} c_{1,t}\\ &-\beta \mathbb{E}_t\left[V_{11,t+1}\left(e^{z_t}\theta k_t^{\theta-1}+1-\delta-c_{1,t}\right)\right]=0\end{aligned}$$

  In steady state:

$$\left(\beta V_{11,ss}-(1-\beta)\gamma c_{ss}^{-\gamma-1}\right)c_{1,ss}=\beta\left[V_{11,ss}\left(\theta k_{ss}^{\theta-1}+1-\delta\right)\right]$$

  or

$$c_{1,ss}=\frac{V_{11,ss}}{\beta V_{11,ss}-(1-\beta)\gamma c_{ss}^{-\gamma-1}}$$

  where we have used that $1=\beta\left(\theta k_{ss}^{\theta-1}+1-\delta\right)$.

- Derivative with respect to z_t:

$$\begin{aligned}&-(1-\beta)\gamma c\left(k_t, z_t; \chi\right)^{-\gamma-1} c_{2,t}\\ &-\beta \mathbb{E}_t\left(V_{11,t+1}\left(e^{z_t}k_t^{\theta}-c_{2,t}\right)+V_{12,t+1}\lambda\right)=0\end{aligned}$$

  In steady state:

$$\left(\beta V_{11,ss}-(1-\beta)\gamma c_{ss}^{-\gamma-1}\right)c_{2,ss}=\beta\left(V_{11,ss}k_t^{\theta}+V_{12,ss}\lambda\right)$$

  or

$$c_{2,ss}=\frac{\beta}{\beta V_{11,ss}-(1-\beta)\gamma c_{ss}^{-\gamma-1}}\left(V_{11,ss}k_{ss}^{\theta}+V_{12,ss}\lambda\right)$$

- Derivative with respect to χ:

$$\begin{aligned}&-(1-\beta)\gamma c\left(k_t, z_t; \chi\right)^{-\gamma-1} c_{3,t}\\ &-\beta \mathbb{E}_t\left(-V_{11,t+1}c_{3,t}+V_{12,t+1}\sigma\varepsilon_{t+1}+V_{13,t+1}\right)=0\end{aligned}$$

  In steady state:

$$\left(\beta V_{11,ss}-(1-\beta)\gamma c_{ss}^{-\gamma-1}\right)c_{3,ss}=\beta V_{13,ss}$$

  or

$$c_{3,ss}=\frac{\beta}{\left(\beta V_{11,ss}-(1-\beta)\gamma c_{ss}^{-\gamma-1}\right)}V_{13,ss}$$

Second Equation I

We have:

$$V_{1,t} = \beta \mathbb{E}_t V_{1,t+1} \left(\theta e^{z_t} k_t^{\theta-1} + 1 - \delta \right)$$

- Derivative with respect to k_t:

$$V_{11,t} = \beta \mathbb{E}_t \left[\begin{array}{c} V_{11,t+1} \left(\theta e^{z_t} k_t^{\theta-1} + 1 - \delta - c_{1,t} \right) \left(\theta e^{z_t} k_t^{\theta-1} + 1 - \delta \right) \\ + V_{1,t+1} \theta \left(\theta - 1 \right) e^{z_t} k_t^{\theta-2} \end{array} \right]$$

  In steady state:

$$V_{11,ss} = \left[V_{11,ss} \left(\frac{1}{\beta} - c_{1,ss} \right) + \beta V_{1,ss} \theta \left(\theta - 1 \right) k_{ss}^{\theta-2} \right]$$

  or

$$V_{11,ss} = \frac{\beta}{1 - \frac{1}{\beta} + c_{1,ss}} V_{1,ss} \theta \left(\theta - 1 \right) k_{ss}^{\theta-2}$$

- Derivative with respect to z_t:

$$V_{12,t} = \beta \mathbb{E}_t \left[\begin{array}{c} V_{11,t+1} \left(e^{z_t} k_t^{\theta} - c_{2,t} \right) \left(\theta e^{z_t} k_t^{\theta-1} + 1 - \delta \right) \\ + V_{12,t+1} \lambda \left(\theta e^{z_t} k_t^{\theta-1} + 1 - \delta \right) + V_{1,t+1} \theta e^{z_t} k_t^{\theta-1} \end{array} \right]$$

  In steady state:

$$V_{12,ss} = V_{11,ss} \left(k_{ss}^{\theta} - c_{2,ss} \right) + V_{12,ss} \lambda + \beta V_{1,ss} \theta k_t^{\theta-1}$$

  or

$$V_{12,ss} = \frac{1}{1 - \lambda} \left[V_{11,ss} \left(k_{ss}^{\theta} - c_{2,ss} \right) + \beta V_{1,ss} \theta k_{ss}^{\theta-1} \right]$$

- Derivative with respect to χ:

$$V_{13,t} = \beta \mathbb{E}_t \left[- V_{11,t+1} c_{3,t} + V_{12,t+1} \sigma \varepsilon_{t+1} + V_{13,t+1} \right]$$

  In steady state,

$$\begin{aligned} V_{13,ss} &= \beta \left[- V_{11,ss} c_{3,ss} + V_{13,ss} \right] \Rightarrow \\ V_{13,ss} &= \frac{\beta}{\beta - 1} V_{11,ss} c_{3,ss} \end{aligned}$$

  but since we know that:

$$c_{3,ss} = \frac{\beta}{\left(\beta V_{11,ss} - \left(1 - \beta \right) \gamma c_{ss}^{-\gamma-1} \right)} V_{13,ss}$$

  the two equations can only hold simultaneously if $V_{13,ss} = c_{3,ss} = 0$.

Third Equation I

We have

$$V_{2,t} = \beta \mathbb{E}_t \left[V_{1,t+1} e^{z_t} k_t^\theta + V_{2,t+1} \lambda \right]$$

- Derivative with respect to z_t:

$$V_{22,t} = \beta \mathbb{E}_t \left[\begin{array}{c} V_{11,t+1} \left(e^{z_t} k_t^\theta - c_{2,t} \right) e^{z_t} k_t^\theta + V_{12,t+1} \lambda e^{z_t} k_t^\theta \\ + V_{1,t+1} e^{z_t} k_t^\theta + V_{21,t+1} \lambda \left(e^{z_t} k_t^\theta - c_{2,t} \right) + V_{22,t+1} \lambda^2 \end{array} \right]$$

In steady state:

$$\begin{aligned} V_{22,t} &= \beta \left[\begin{array}{c} V_{11,ss} \left(k_t^\theta - c_{2,ss} \right) k_{ss}^\theta + V_{12,ss} \lambda k_{ss}^\theta + V_{1,ss} k_{ss}^\theta \\ + V_{21,ss} \lambda \left(k_{ss}^\theta - c_{2,ss} \right) + V_{22,ss} \lambda^2 \end{array} \right] \Rightarrow \\ V_{22,ss} &= \frac{\beta}{1 - \beta \lambda^2} \left[\begin{array}{c} V_{11,ss} \left(k_t^\theta - c_{2,ss} \right) k_{ss}^\theta + 2 V_{12,ss} \lambda k_{ss}^\theta \\ + V_{1,ss} k_{ss}^\theta - V_{12,ss} \lambda c_{2,ss} \end{array} \right] \end{aligned}$$

where we have used $V_{12,ss} = V_{21,ss}$.

- Derivative with respect to χ:

$$V_{23,t} = \beta \mathbb{E}_t \left[\begin{array}{c} -V_{11,t+1} e^{z_t} k_t^\theta c_{3,t} + V_{12,t+1} e^{z_t} k_t^\theta \sigma \varepsilon_{t+1} + V_{13,t+1} e^{z_t} k_t^\theta \\ - V_{21,t+1} \lambda c_{3,t} + V_{22,t+1} \lambda \sigma \varepsilon_{t+1} + V_{23,t+1} \lambda \end{array} \right]$$

In steady state:

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

Fourth Equation

We have

$$V_{3,t} = \beta \mathbb{E}_t \left[V_{2,t+1} \sigma \varepsilon_{t+1} + V_{3,t+1} \right].$$

- Derivative with respect to χ:

$$V_{33,t} = \beta \mathbb{E}_t \left[\begin{array}{c} -V_{21,t+1} c_{3,t} \sigma \varepsilon_{t+1} + V_{22,t+1} \sigma^2 \varepsilon_{t+1}^2 + V_{23,t+1} \sigma \varepsilon_{t+1} \\ - V_{31,t+1} c_{3,t} + V_{32,t+1} \sigma \varepsilon_{t+1} + V_{33,t+1} \end{array} \right]$$

In steady state:

$$V_{33,ss} = \frac{\beta}{1-\beta} V_{22,ss}$$

System I

$$c_{1,ss} = \frac{V_{11,ss}}{\beta V_{11,ss} - (1-\beta)\gamma c_{ss}^{-\gamma-1}}$$

$$c_{2,ss} = \frac{\beta}{\beta V_{11,ss} - (1-\beta)\gamma c_{ss}^{-\gamma-1}} \left(V_{11,ss} k_{ss}^{\theta} + V_{12,ss}\lambda \right)$$

$$V_{11,ss} = \frac{\beta}{1 - \frac{1}{\beta} + c_{1,ss}} V_{1,ss}\theta(\theta-1) k_{ss}^{\theta-2}$$

$$V_{12,ss} = \frac{1}{1-\lambda} \left[V_{11,ss} \left(k_{ss}^{\theta} - c_{2,ss} \right) + \beta V_{1,ss} \theta k_{ss}^{\theta-1} \right]$$

$$V_{22,ss} = \frac{\beta}{1-\beta\lambda^2} \left[\begin{array}{c} V_{11,ss} \left(k_t^{\theta} - c_{2,ss} \right) k_{ss}^{\theta} + 2V_{12,ss}\lambda k_{ss}^{\theta} \\ + V_{1,ss} k_{ss}^{\theta} - V_{12,ss}\lambda c_{2,ss} \end{array} \right]$$

$$V_{33,ss} = \frac{\beta}{1-\beta} \sigma^2 V_{22,ss}$$

plus $c_{3,ss} = V_{13,ss} = V_{23,ss} = 0$.

System II

- This is a system of nonlinear equations.
- However, it has a recursive structure.
- By substituting variables that we already know, we can find $V_{11,ss}$.
- Then, using this results and by plugging $c_{2,ss}$, we have a system of two equations, on two unknowns, $V_{12,ss}$ and $V_{22,ss}$.
- Once the system is solved, we can find $c_{1,ss}$, $c_{2,ss}$, and $V_{33,ss}$ directly.

The Welfare Cost of the Business Cycle

- An advantage of performing the perturbation on the value function is that we have evaluation of welfare readily available.

- Note that at the deterministic steady state, we have:

$$V\left(k_{ss}, 0; \chi\right) \simeq V_{ss} + \frac{1}{2} V_{33,ss}$$

- Hence $\frac{1}{2} V_{33,ss}$ is a measure of the welfare cost of the business cycle.
- Note that this quantity is not necessarily negative. Indeed, it may well be positive in many models, like in a RBC with leisure choice. See Cho and Cooley (2000).

Our Example

- We know that $V_{ss} = \frac{c_{ss}^{1-\gamma}}{1-\gamma}$.
- Then, we can compute the decrease in consumption τ that will make the household indifferent between consuming $(1-\tau) c_{ss}$ units per period with certainty or c_t units with uncertainty.
- To do so, note that:

$$\begin{aligned} \frac{c_{ss}^{1-\gamma}}{1-\gamma} + \frac{1}{2} V_{33,ss} &= \frac{\left(c_{ss}\left(1-\tau\right)\right)^{1-\gamma}}{1-\gamma} \Rightarrow \\ \left(\left(1-\tau\right)^{1-\gamma} - 1\right) c_{ss}^{1-\gamma} &= \left(1-\gamma\right) \frac{1}{2} V_{33,ss} \end{aligned}$$

or

$$\tau = 1 - \left[1 + \frac{1-\gamma}{c_{ss}^{1-\gamma}} \frac{1}{2} V_{33,ss}\right]^{\frac{1}{1-\gamma}}$$

A Numerical Example

- We pick standard parameter values by setting

$$\beta = 0.99, \gamma = 2, \delta = 0.0294, \theta = 0.3, and \lambda = 0.95.$$

- Then, we get:

$$\begin{aligned} V\left(k_t, z_t; 1\right) &\simeq -0.54000 + 0.00295\left(k_t - k_{ss}\right) + 0.11684 z_t \\ &\quad -0.00007\left(k_t - k_{ss}\right)^2 - 0.00985 z_t^2 \\ &\quad -0.97508\sigma^2 - 0.00225\left(k_t - k_{ss}\right) z_t \\ c\left(k_t, z_t; \chi\right) &\simeq 1.85193 + 0.04220\left(k_t - k_{ss}\right) + 0.74318 z_t \end{aligned}$$

- DYNARE produces the same policy function by linearizing the equilibrium conditions of the problem.
- Also, the consumption equivalent of the welfare cost of the business cycle is 8.8475e-005, even lower than Lucas' (1987) original computation because of the smoothing possibilities implied by the presence of capital.
- Use as an initial guess for VFI.

5 Projection

We move now to explore projection, the second class of solution methods that we will cover in this chapter. Projection methods solve the functional equations of the form:

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

by specifying

$$d^j(\mathbf{x}, \theta) = \sum_{i=0}^{j} \theta_i \Psi_i(\mathbf{x}) \tag{19}$$

that is, by building a linear combination of basis function $\Psi_i(\mathbf{x})$ given coefficients θ_i. Note that we are building linear combinations, but that more general non-linear alternatives:

$$d^j(\mathbf{x}, \theta) = f(\Psi_i(\mathbf{x}), \theta)$$

for a known function $f(\cdot, \theta)$ are possible. However, the theory for more general approximations are less well developed than the one with linear combinations and, in any case, it is probably more pedagogical to start with the linear combination case. The fact that we are handling linear combinations also means that, in general, we will have with the same number of coefficients θ_i than basis functions $\Psi_i(\mathbf{x})$.

Inspection of equation (19) reveals that, to build the function $d^j(\mathbf{x}, \theta)$, we need to pick a basis $\{\Psi_i(x)\}_{i=0}^{\infty}$ and "project" $\mathcal{H}(\cdot)$ against that basis to find the θ_i's. Different choices of the basis function and on the projection metric will imply different projection methods, often known in the literature with their own names.

Basic Algorithm of projection goes is:

1. Define j known linearly independent functions $\psi_i : \Omega \to \mathbb{R}^m$ where $n < \infty$. We call the $\psi_1(\cdot), \psi_2(\cdot), ..., \psi_n(\cdot)$ the *basis functions*.
2. Define a vector of coefficients $\theta = [\theta_1, \theta_2, ..., \theta_n]$.
3. Define a combination of the basis functions and the θ's:
$$d^j(\cdot|\theta) = \sum_{i=0}^{j} \theta_i \psi_n(\cdot)$$
4. Plug $d^n(\cdot|\theta)$ into $H(\cdot)$ to find the *residual equation*:
$$R(\cdot|\theta) = \mathcal{H}(d^n(\cdot|\theta))$$
5. Find the value of $\widehat{\theta}$ that make the residual equation as close to $\mathbf{0}$ as possible given some objective function $\rho : J^1 \times J^1 \to J^2$:
$$\widehat{\theta} = \arg\min_{\theta \in \Re^n} \rho(R(\cdot|\theta), \mathbf{0})$$

5.1 Relation with Econometrics

- Looks a lot like OLS. Explore this similarity later in more detail.
- Also with semi-nonparametric methods as Sieves.
- Compare with:
    1. Policy iteration.
    2. Parameterized Expectations.

Two Issues

We need to decide:

1. Which basis we use?
    (a) Pick a global basis⇒spectral methods.
    (b) Pick a local basis⇒finite elements methods.
2. How do we "project"?

Different choices in 1 and 2 will result in slightly different projection methods.

5.2 Spectral Techniques

Spectral techniques were introduced in economics by Judd (19992). Spectral techniques use basis functions that are nonzero and smooth almost everywhere in Ω. The main advantage of this class of basis functions is their simplicity: building and working with the approximation would be particularly simple. The main disadvantage of spectral basis is that they would have a hard time capturing local behavior. Gibbs phenomenon.

5.2.1 Spectral Basis

We introduce now some of the spectral basis. First, we will deal with the unidimensional case where there is only one state variable. In the next subsection, we will deal with the more general case of an arbitrary number of state variables.

5.3 Monomials

A first possible basis is the monomials $c, x, x^2, x^3, ...$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 assures completeness of our approximation in the L^1 norm.

Unfortunately, monomials suffer from two severe problems. First, monomials are (nearly) multicollinear. To illustrate this point, we plot in figure x, the graphs of x^{10} with x^{11} for $x \in [0.5, 1.5]$. In the figure we can see how the both functions have a very similar shape. As we add higher monomials, the new components of the solution do not allow to significantly diminish the distance between the exact function we want to approximate and the computed approximation.[9]

The second problem of monomials, which we can also see in figure x, is that they vary considerably in size, leading to scaling problems and accumulation of numerical errors.

[9]A simple case where this problem is particularly clear is when the operator $\mathcal{H}(\cdot)$ is linear. In that situation, the solution of the projection involves inversion of matrices. When the basis functions are similar, the condition number of these matrices (the ratio of the largest and smallest absolute eigenvalues) are too high. Just the six first monomials can generate conditions numbers of 10^{10}. In fact, the matrix of the least square 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 monomials illustrate that we want to search for an orthogonal basis that has a bounded variation in range. Orthogonality will imply that, when we add a new element of the basis to the approximation (i.e., when we go from order j to order $j+1$), the newest element brings a sufficiently different behavior as to capture features of the unknown function d not well approximated before.

5.4 Trigonometric series

A second possible basis is the trigonometric series

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

Trigonometric series are particularly well-suited to approximate periodic functions. Unfortunately, economic problems are rarely periodic. Furthermore, the periodic approximations to nonperiodic functions suffer from the Gibbs phenomenon, requiring many terms to achieve good numerical performance (the rate of convergence to the true solution as $n \to \infty$ is only $O\left(n\right)$).

5.5 Orthoghonal polynomials of Jacobi type

Orthoghonal polynomials of Jacobi (also known as hypergeometric) type are a flexible class. The Jacobi polynomial of degree n, $P_n^{\alpha,\beta}\left(x\right)$ for $\alpha, \beta > -1$, is defined by the orthogonality condition of the form:

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

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

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

that the general n term is given by:

$$2^{-n}\sum_{k=0}^{n}\left(\begin{array}{c} n+\alpha \\ k \end{array}\right)\left(\begin{array}{c} n+\beta \\ n-k \end{array}\right)(x-1)^{n-k}(x+1)^{k}$$

Recursively:

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

The two most important cases of Jacobi polynomials are the Legendre, where $\alpha = \beta = -\frac{1}{2}$, and the Chebyshev: $\alpha = \beta = 0$.

5.6 Chebyshev Polynomials

Chebyshev polynomials are one of the most common tools of applied mathematics. See, for example, Boyd (2001) and Fornberg (1998). Their popularity is easily explained if we consider their advantages:

1. Numerous simple close-form expressions for the Chebyshev polynomials are available. Thus, the researcher can easily move from one representation to another according to her convenience. We will present several of them below.

2. The change between the coefficients of a Chebyshev expansion of a function and the values of the function at the Chebyshev nodes are quickly performed by the cosine transform.

3. Chebyshev polynomials are more robust than their alternatives for interpolation.

4. Chebyshev polynomials are bounded between $[-1, 1]$ while Legendre polynomials are not, offering a better performance close to the boundaries of the problems.

5. Chebyshev polynomials are smooth functions.

6. Several theorems bound the errors for Chebyshev polynomials interpolations.

We have a number of alternative ways to define Chebyshev polynomials. The most common definition is recursive, with $T_0(x) = 1$, $T_1(x) = x$, and the general $n+1$-th order polynomial given by:

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

Applying this recursive definition, the first few polynomials are then 1, x, $2x^2 - 1$, $4x^3 - 3x$, $8x^4 - 8x^2 + 1$, etc... In figure x, we plot the Chebysev polynomials of order 0 to 5.

The n zeros of the polynomial $T_n(x_k) = 0$ are given by:

$$x_k = \cos \frac{2k-1}{2n}\pi, k = 1, ..., n.$$

The fact that the Chebyshev polynomials of order n has n zeros will be useful below. Also, note that this zeros are clustered quadratically towards ± 1.

As we mentioned above, there is number of explicit and equivalent definitions for the Chebyshev polynomials:

$$\begin{aligned}
T_n(x) &= \cos(n \arccos x) \\
&= \frac{1}{2}\left(z^n + \frac{1}{z^n}\right) \text{ where } \frac{1}{2}\left(z + \frac{1}{z}\right) = x \\
&= \frac{1}{2}\left(\left(x + \left(x^2-1\right)^{0.5}\right)^n + \left(x - \left(x^2-1\right)^{0.5}\right)^n\right) \\
&= \frac{1}{2}\sum_{k=0}^{[n/2]} (-1)^k \frac{(n-k-1)!}{k!\,(n-2k)!}(2x)^{n-2k} \\
&= \frac{(-1)^n \pi^{0.5}}{2^n \Gamma\left(n+\frac{1}{2}\right)}\left(1-x^2\right)^{0.5} \frac{d^n}{dx^n}\left(\left(1-x^2\right)^{n-\frac{1}{2}}\right)
\end{aligned}$$

A few remarks deserve to be highlighted. First, the domain of the Chebyshev polynomials is $[-1, 1]$. Since our state space is, in general, different, we use a linear translation from $[a, b]$ into $[-1, 1]$:

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

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

$$\frac{1}{\left(1-x^2\right)^{0.5}}.$$

Chebyshev Interpolation Theorem: 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.

5.6.1 Multidimensional Problems

- Chebyshev polynomials are defined on $[-1, 1]$.
- However, most problems in economics are multidimensional.
- How do we generalize the basis?
- Curse of dimensionality.

5.7 Tensors

- Assume we want to approximate $F : [-1, 1]^d \to \mathbb{R}$.
- Let T_j denote the Chebyshev polynomial of degree $j = 0, 1, .., \kappa$.
- We can approximate F with tensor product of Chebyshev polynomials of degree κ:

$$\hat{F}(x) = \sum_{n_1=0}^{\kappa} \cdots \sum_{n_d=0}^{\kappa} \xi_{n_1,...,n_d} T_{n_1}(x_1) \cdots T_{n_d}(x_d)$$

- Beyond simplicity, an advantage of the tensor basis is that if the one-dimensional basis is orthogonal in a norm, the tensor basis is orthogonal in the product norm.
- Disadvantage: number of elements increases exponentially. We end up having terms $x_1^\kappa x_2^\kappa \cdots x_d^\kappa$, total number of $(\kappa + 1)^d$.

5.8 Complete Polynomials

- Solution: eliminate some elements of the tensor in such a way that there is not much numerical degradation.
- Judd and Gaspar (1997): Use complete polynomials instead

$$\mathcal{P}_\kappa^d \equiv \left\{ x_1^{i_1} \cdots x_d^{i_d} \text{ with } \sum_{l=1}^{d} i_l \leq \kappa,\ 0 \leq i_1, ..., i_d \right\}$$

- Advantage: much smaller number of terms, no terms of order $d\kappa$ to evaluate.
- Disadvantage: still too many elements.

5.9 Smolyak's Algorithm

- Define $m_1 = 1$ and $m_i = 2^{i-1} + 1$, $i = 2, ...$
- Define $G^i = \{x_1^i, ..., x_{m_i}^i\} \subset [-1, 1]$ as the set of the extrema of the Chebyshev polynomials

$$x_j^i = -cos\left(\frac{\pi(j-1)}{m_i - 1}\right) \quad j = 1, ..., m_i$$

  with $G^1 = \{0\}$. It is crucial that $G^i \subset G^{i+1}$, $\forall i = 1, 2, \ldots$
- Example:

$$\begin{aligned} i &= 1, m_i = 1, \mathcal{G}^i = \{0\} \\ i &= 2, m_i = 3, \mathcal{G}^i = \{-1, 0, 1\} \\ i &= 3, m_i = 5, \mathcal{G}^i = \{-1, -\cos\left(\frac{\pi}{4}\right), 0, -\cos\left(\frac{3\pi}{4}\right), 1\} \end{aligned}$$

- For $q > d$, define a sparse grid

$$\mathcal{H}(q, d) = \bigcup_{q-d+1 \leq |i| \leq q} (\mathcal{G}^{i_1} \times ... \times \mathcal{G}^{i_d}),$$

  where $|i| = i_1 + \ldots + i_d$. The number q defines the size of the grid and thus the precision of the approximation.
- For example, let $q = d + 2 = 5$:

$$\mathcal{H}(5, 3) = \bigcup_{3 \leq |i| \leq 5} (\mathcal{G}^{i_1} \times ... \times \mathcal{G}^{i_d}).$$

$$\begin{array}{lll} \mathcal{G}^3 \times \mathcal{G}^1 \times \mathcal{G}^1, & \mathcal{G}^1 \times \mathcal{G}^3 \times \mathcal{G}^1, & \mathcal{G}^1 \times \mathcal{G}^1 \times \mathcal{G}^3 \\ \mathcal{G}^2 \times \mathcal{G}^2 \times \mathcal{G}^1, & \mathcal{G}^2 \times \mathcal{G}^1 \times \mathcal{G}^2, & \mathcal{G}^1 \times \mathcal{G}^2 \times \mathcal{G}^2 \\ \mathcal{G}^2 \times \mathcal{G}^1 \times \mathcal{G}^1, & \mathcal{G}^1 \times \mathcal{G}^2 \times \mathcal{G}^1, & \mathcal{G}^1 \times \mathcal{G}^1 \times \mathcal{G}^2 \\ \mathcal{G}^1 \times \mathcal{G}^1 \times \mathcal{G}^1 & & \end{array}$$

- Number of points for $q = d + 2$

$$1 + 4d + 4\frac{d(d-1)}{2}$$

- Largest number of points along one dimension

$$\begin{aligned} i &= q - d + 1 \\ m_i &= 2^{q-d} + 1 \end{aligned}$$

- Rectangular grid

$$\left[2^{q-d} + 1\right]^d$$

- Key: with rectangular grid, the number of grid points increases exponentially in the number of dimensions. With the Smolyak algorithm number of points increases polynomially in the number of dimensions.

Size of the Grid for $q = d + 2$			
d	$2^{q-d} + 1$	$\#\mathcal{H}(q,d)$	$\left[2^{q-d} + 1\right]^d$
2	5	13	25
3	5	25	125
4	5	41	625
5	5	61	$3,125$
12	5	313	$244,140,625$

- For one dimension denote the interpolating Chebyshev polynomials as

$$\mathcal{U}^i(x^i) = \sum_{j=1}^{m_i} \xi_j^i T_j(x^i)$$

and the d-dimensional tensor product by $U^{i_1} \otimes ... \otimes U^{i_d}(x)$.

- For $q > d$, approximating function (Smolyak's algorithm) given by

$$\mathcal{A}(q,d)(x) = \sum_{q-d+1 \leq |i| \leq q} (-1)^{q-|i|} \begin{pmatrix} d-1 \\ q-|i| \end{pmatrix} (\mathcal{U}^{i_1} \otimes \ldots \otimes \mathcal{U}^{i_d})(x)$$

- Method is (almost) optimal within the set of polynomial approximations (Barthelmann, Novak and Ritter, 1999).

- Method is universal, that is, almost optimal for many different function spaces.

Boyd's Moral Principal

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.

5.10 Finite Elements

- Standard Reference: McGrattan (1999).
- Bound the domain Ω in small of the state variables.
- Partition Ω in small in nonintersecting elements.
- These small sections are called elements.
- The boundaries of the elements are called nodes.

Partition into Elements

- Elements may be of unequal size.
- We can have small elements in the areas of Ω where the economy will spend most of the time while just a few, big size elements will cover wide areas of the state space infrequently visited.
- Also, through elements, we can easily handle issues like kinks or constraints.
- There is a whole area of research concentrated on the optimal generation of an element grid. See Thomson, Warsi, and Mastin (1985).

Structure

- Choose a basis for the policy functions in each element.

- Since the elements are small, a linear basis is often good enough:

$$\psi_i(k) = \begin{cases} \frac{x - x_{i-1}}{x_i - x_{i-1}} & \text{if } x \in [x_{i-1}, x_i] \\ \frac{x_{i+1} - x}{x_{i+1} - x_i} & \text{if } k \in [x_i, x_{i+1}] \\ 0 \text{ elsewhere} \end{cases}$$

- Plug the policy function in the Equilibrium Conditions and find the unknown coefficients.
- Paste it together to ensure continuity.
- Why is this an smart strategy?
- Advantages: we will need to invert an sparse matrix.
- When should be choose this strategy? speed of computation versus accuracy.

The literature distinguish among three different refinements:

1. *h-refinement*: this scheme subdivides each element into smaller elements to improve resolution uniformly over the domain.
2. *r-refinement*: this scheme subdivides each element only in those regions where there are high nonlinearities.
3. *p-refinement*: this scheme increases the order of the approximation in each element. If the order of the expansion is high enough, we will generate in that way an hybrid of finite and spectral methods knows as spectral elements.

5.11 Choosing the Objective Function

- The most common answer to the second question is given by a *weighted residual*.
- That is why often projection methods are also called weighted residual methods
- This set of techniques propose to get the residual close to **0** in the weighted integral sense.

- Given some weight functions $\phi_i : \Omega \to \mathbb{R}^m$:

$$\rho\left(R\left(\cdot\left|\,\theta\right.\right),\mathbf{0}\right)=\begin{cases} 0 \text{ if } \int_\Omega \phi_i\left(x\right) R\left(\cdot\left|\,\theta\right.\right) dx=\mathbf{0},\ i=0,..,n \\ 1 \text{ otherwise}\end{cases}$$

- Then the problem is to choose the θ that solve the system of equations:

$$\int_\Omega \phi_i\left(x\right) R\left(\cdot\left|\,\theta\right.\right) dx=\mathbf{0},\ i=0,..,n$$

Remarks

- With the approximation of d by some functions ψ_i and the definition of some weight functions $\phi_i\left(\cdot\right)$, we have transform a rather intractable functional equation problem into the standard nonlinear equations system!
- The solution of this system can be found using standard methods, as a Newton for relatively small problems or a conjugate gradient for bigger ones.
- Issue: we have different choices for an weight function:

5.12 Weight Function I: Least Squares

- $\phi_i\left(x\right)=\frac{\partial R\left(x\left|\theta\right.\right)}{\partial\theta_i}$.
- This choice is motivated by the solution of the variational problem:

$$\min_\theta \int_\Omega R^2\left(\cdot\left|\,\theta\right.\right) dx$$

with first order condition:

$$\int_\Omega \frac{\partial R\left(x\left|\,\theta\right.\right)}{\partial\theta_i} R\left(\cdot\left|\,\theta\right.\right) dx=\mathbf{0},\ i=0,..,n$$

- Variational problem is mathematically equivalent to a standard regression problem in econometrics.
- OLS or NLLS are regression against a manifold spanned by the observations.
- Least Squares always generates symmetric matrices even if the operator $\mathcal{H}$ is not self-adjoint.

- Symmetric matrices are convenient theoretically (they simplify the proofs) and computationally (there are algorithms that exploit their structure to increase speed and decrease memory requirements).
- However, least squares may lead to ill-conditioning and systems of equations complicated to solve numerically.

5.13 Weight Function II: Subdomain

- We divide the domain Ω in n subdomains Ω_i and define the n step functions:

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

- This choice is then equivalent to solve the system:

$$\int_{\Omega_i} R(\cdot|\theta)\, dx = \mathbf{0},\ i = 0, .., n$$

5.14 Weight Function III: Moments

- Take $\{0, x, x^2, ..., x^{n-1}\}$ and compute the first n periods of the residual function:

$$\int_{\Omega_i} x^i R(\cdot|\theta)\, dx = \mathbf{0},\ i = 0, .., n$$

- This approach, widely used in engineering works well for a low n (2 or 3).
- However, for higher orders, its numerical performance is very low: high orders of x are highly collinear and arise serious rounding error problems.
- Hence, moments are to be avoided as weight functions.

Weight Function III: Collocation or Pseudospectral or Method of Selected Points

- $\phi_i(x) = \delta(x - x_i)$ where δ is the dirac delta function and x_i are the collocation points.
- This method implies that the residual function is zero at the n collocation points.

- Simple to compute since the integral only needs to be evaluated in one point. Specially attractive when dealing with strong nonlinearities.

- A systematic way to pick collocation points is to use a density function:

$$\mu_\gamma(x) = \frac{\Gamma\left(\frac{3}{2} - \gamma\right)}{(1-x^2)^\gamma \pi^{\frac{1}{2}} \Gamma(1-\gamma)} \quad \gamma < 1$$

  and find the collocation points as the x_j, $j = 0, ..., n$ solutions to:

$$\int_{-1}^{x_j} \mu_\gamma(x)\, dx = \frac{j}{n}$$

- For $\gamma = 0$, the density function implies equispaced points.

Weight Function IV: Orthogonal Collocation

- Variation of the collocation method:

  1. Basis functions are a set of orthogonal polynomials.
  2. Collocation points given by the roots of the $n - th$ polynomial.

- When we use Chebyshev polynomials, their roots are the collocation points implied by $\mu_{\frac{1}{2}}(x)$ and their clustering can be shown to be optimal as $n \to \infty$.

- Surprisingly good performance of orthogonal collocation methods.

5.15 Weight Function V: Galerkin or Rayleigh-Ritz

- $\phi_i(x) = \psi_i(x)$ with a linear approximating function $\sum_{i=0}^n \theta_i \psi_i(x)$.

- Then:

$$\int_\Omega \psi_i(x)\, H\left(\sum_{i=0}^n \theta_i \psi_i(x)\right) dx = \mathbf{0},\ i = 0, .., n$$

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

- Galerkin is a 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:

$$\lim_{n\to\infty} \sum_{i=0}^{n} \theta_i \psi_i (\cdot) = d (\cdot)$$

- Experience suggests that a Galerkin approximation of order n is as accurate as a Pseudospectral $n+1$ or $n+2$ expansion.

6 Which Methods to Use?

Description of trade-offs....

7 Analysis of Error

A key 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 a challenging task since the exact solution is unknown. However, different methods have been proposed to evaluate the error.

- As with projection, it is important to study the Euler equation errors.

Once we know we can improve errors:

- 1. Adding additional functions in the basis.
  2. Refine the elements.
- Multigrid schemes.

8 Parallel Programming

- Moore's Law (1965): transistor density of semiconductor chips would double roughly every 18 months.
- Problems when transistor size falls by a factor x:
  1. Electricity consumption goes up by x^4.
  2. Heat goes up.
  3. Manufacturing costs go up.
- Inherent limits on serial machines imposed by the speed of light (30 cm/nanosecond) and transmission limit of copper wire (9 cm/nanosecond): virtually impossible to build a serial Teraflop machine with current approach.
- Furthermore, real bottleneck is often memory access (RAM latency has only improved around 10% a year).
- Alternative: having more processors!
- Main idea⇒divide and conquer:
  1. Numerical computation.
  2. Data handling (`MapReduce` and `Hadoop`).
- Two issues:
  1. Algorithms.
  2. Coding.
  3.

When do we parallelize?

- Scalability:
  1. Strongly scalable: problems that are inherently easy to parallelize.

2. Weakly scalable: problems that are not.

- Granularity:

1. Coarse: more computation than communication.
2. Fine: more communication.

- Overheads and load balancing.
- Whether or not the problem is easy to parallelize may depend on the way you set it up.
- Taking advantage of your architecture.
- Trade off between speed up and coding time.
- Debugging and profiling may be challenging.
- You will need a good IDE, debugger, and profiler.

8.1 Example I: value function iteration

$$V(k) = \max_{k'} \{u(c) + \beta V(k')\}$$
$$c = k^{\alpha} + (1-\delta)k - k'$$

1. We have a grid of capital with 100 points, $k \in [k_1, k_2, ..., k_{100}]$.
2. We have a current guess $V^n(k)$.
3. We can send the problem:

$$\max_{k'} \{u(c) + \beta V^n(k')\}$$
$$c = k_1^{\alpha} + (1-\delta)k_1 - k'$$

   to processor 1 to get $V^{n+1}(k_1)$.

4. We can send similar problem for each k to each processor.
5. When all processors are done, we gather the $V^{n+1}(k_1)$ back.

8.2 Example II: random walk Metropolis-Hastings

- Get a draw from an arbitrary distribution $P(\cdot)$

$$\theta \sim P(\cdot)$$

- How?

  1. Given a state of the chain θ_{n-1}, we generate a proposal:

$$\theta^* = \theta_{n-1} + \lambda\varepsilon,\ \varepsilon \sim \mathcal{N}(0,1)$$

  2. We compute:

$$\alpha = \min\left\{1, \frac{P(\theta^*)}{P(\theta_{n-1})}\right\}$$

  3. We set:

$$\begin{aligned}\theta_n &= \theta^*\ w.p.\ \alpha \\ \theta_n &= \theta_{n-1}\ w.p.\ 1-\alpha\end{aligned}$$

- Problem: to generate θ^* we need to θ_{n-1}.
- No obvious fix (parallel chains violate the asymptotic properties of the chain).
- Limitations
- Amdahl's Law: the speedup of a program using multiple processors in parallel computing is limited by the time needed for the sequential fraction of the program.
- Overheads:

  1. Cost of starting a thread or process.
  2. Cost of communicating shared data.
  3. Cost of synchronizing.

- Load imbalance: often we use less than 10% of capability of a machine.
- ENDOFJESUS: 10160 words.

9 Confronting DSGE Models with Data

9.1 A Simple DSGE Model and Its Solution

Throughout the second part of this chapter we consider a simplified New Keynesian DSGE model in its log-linearized form. 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, Eichenbaum, and Evans (2005) and Smets and Wouters (2003). The specific version presented below is taken from Del Negro, Schorfheide, Smets, and Wouters (2007) and obtained by imposing a several parameter restrictions. As is, the model presented below is not suitable to be confronted with actual data, but it can be solved analytically, which is tremendously useful for discussing estimation and inference. We subsequently provide a brief summary of the model that focuses on the key equilibrium conditions.

The model economy consists of households, intermediate goods producers, final goods producers, a monetary policy authority, and a fiscal authority. Macroeconomic fluctuations are driven 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 will derive a law of motion for aggregate output X_t, real wages W_t, gross inflation π_t, and a gross nominal interest rate R_t. We assume that productivity Z_t in the economy is evolving exogenously according to a random walk with drift:

$$\ln Z_t = \ln \gamma + \ln Z_{t-1} + z_t, \quad z_t = \rho_z z_{t-1} + \sigma_z \epsilon_{z,t} \tag{20}$$

The productivity process Z_t induces a stochastic trend in output and real wages. To facilitate the model solution, it is useful to introduce the detrended variables $x_t = X_t/Z_t$ and $w_t = W_t/Z_t$. In terms of the detrended variables, the model has the following steady state:

$$x = x^*, \quad w = lsh = \frac{1}{1+\lambda}, \quad \pi = \pi^*, \quad R = \pi^* \frac{\gamma}{\beta}. \tag{21}$$

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 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 was the growth rate of technology. Under the

assumption that the production technology is linear in labor and labor is the only factor of production, the steady state labor share equals the steady state of detrended wages.

We now present the equilibrium conditions in log-linearized form. Let $\hat{x} = \ln(x_t/x)$, $\hat{w}_t = \ln(w_t/w)$, $\hat{\pi}_t = \ln(\pi_t/\pi)$, and $\hat{R}_t = \ln(R_t/R)$. We assume that labor is the only factor of production and that all output is consumed. The consumption Euler equation of the households takes the form

$$\hat{x}_t = \mathbb{E}_{t+1}[\hat{x}_{t+1}] - \left(\hat{R}_t - \mathbb{E}[\hat{\pi}_{t+1}]\right) + \mathbb{E}_t[z_{t+1}]. \tag{22}$$

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 the following labor supply equation:

$$\hat{w}_t = (1+\nu)\hat{x}_t + \phi_t, \tag{23}$$

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

$$\phi_t = \rho_\phi \phi_{t-1} + \sigma_\phi \epsilon_{\phi,t}. \tag{24}$$

The intermediate goods producers hire labor from the household 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 log-linearization around steady state aggregate output, the production function becomes

$$\hat{x}_t(j) = \hat{L}_t(j). \tag{25}$$

Nominal price rigidity is introduced via the Calvo mechanism. In each period firm j is unable to re-optimize 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 to the so-called New Keynesian Phillips curve, which can be expressed as

$$\hat{\pi}_t = \beta \mathbb{E}_t[\hat{\pi}_{t+1}] + \kappa_p(\hat{w}_t + \lambda_t), \quad \kappa_p = \frac{(1-\zeta_p\beta)(1-\zeta_p)}{\zeta_p}, \tag{26}$$

where β is the households' discount factor and λ_t can be interpreted as a price mark-up shock, which exogenously evolves according to

$$\lambda_t = \rho_\lambda \lambda_{t-1} + \sigma_\lambda \epsilon_{\lambda,t}. \tag{27}$$

It is possibe 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

$$\widehat{lsh}_t = \hat{w}_t. \tag{28}$$

Finally, the central bank sets the nominal interest rate according to the feedback rule

$$\hat{R}_t = \psi \hat{\pi}_t + \sigma_R \epsilon_{R,t} \quad \psi = 1/\beta. \tag{29}$$

We abstract from interest rate smoothing and the fact that central banks typically also react to some measure of real activity, e.g., 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 called 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.

To solve the model, first eliminate the nominal interest rate from the consumption Euler equation using (29):

$$\hat{x}_t = \mathbb{E}_{t+1}[\hat{x}_{t+1}] - \left(\frac{1}{\beta}\pi_t + \sigma_R \epsilon_{R,t} - \mathbb{E}[\hat{\pi}_{t+1}]\right) + \mathbb{E}_t[z_{t+1}]. \tag{30}$$

Now notice that the New Keynesian Phillips curve can be rewritten as

$$\frac{1}{\beta}\hat{\pi}_t - \mathbb{E}_t[\hat{\pi}_{t+1}] = \frac{\kappa_p}{\beta}\big((1+\nu)\hat{x}_t + \phi_t + \lambda_t\big). \tag{31}$$

Here we replaced wages $\hat{w}_t$ by the right-hand-side of (23). Substituting (31) into (30) and rearranging terms leads to the following expectational difference equation for output $\hat{x}_t$

$$\hat{x}_t = \psi_p \mathbb{E}[\hat{x}_{t+1}] - \frac{\kappa_p \psi_p}{\beta}(\phi_t + \lambda_t) + \psi_p \mathbb{E}_t[z_{t+1}] - \psi_p \sigma_R \epsilon_{R,t}, \tag{32}$$

where $0 \leq \psi_p \leq 1$ is given by

$$\psi_p = \left(1 + \frac{\kappa_p}{\beta(1+\nu)}\right)^{-1}.$$

Solving the expectational difference equation forward yields

$$\hat{x}_t = -\psi_p \sigma_R \epsilon_{R,t} + \sum_{j=0}^{\infty} \psi_p^j \mathbb{E}_t \left[-\frac{\kappa_p \psi_p}{\beta}(\phi_{t+j} + \lambda_{t+j}) + \psi_p z_{t+1+j} \right]. \tag{33}$$

Using the laws of motion of the exogenous shock processes in (20), (24), and (27), and assuming that the innovations ϵ_t are Martingale difference sequences, we deduce that output in percentage deviations from steady state evolves according to

$$\hat{x}_t = -\frac{\kappa_p \psi_p/\beta}{1 - \psi_p \rho_\phi}\phi_t - \frac{\kappa_p \psi_p/\beta}{1 - \psi_p \rho_\lambda}\lambda_t + \frac{\rho_z \psi_p}{1 - \psi_p \rho_z} z_t - \psi_p \sigma_R \epsilon_{R,t}. \tag{34}$$

After having determined the law of motion for output, we now solve for wages, inflation, and nominal interest rates. Using (23) and (28) we deduce that the labor share is determined according to

$$\widehat{lsh}_t = \left[1 - \frac{(1+\nu)\kappa_p \psi_p/\beta}{1 - \psi_p \rho_\phi}\right]\phi_t - \frac{(1+\nu)\kappa_p \psi_p/\beta}{1 - \psi_p \rho_\lambda}\lambda_t + \frac{(1+\nu)\rho_z \psi_p}{1 - \psi_p \rho_z} z_t - (1+\nu)\psi_p \sigma_R \epsilon_{R,t}. \tag{35}$$

To obtain the law of motion of inflation, we have to solve the New Keynesian Phillips curve (26) forward:

$$\hat{\pi}_t = \sum_{j=0}^{\infty} \beta^j \kappa_p \mathbb{E}_t[\widehat{lsh}_{t+j} + \lambda_{t+j}]. \tag{36}$$

Using (27) and (35) to evaluate the conditional expectations in (36), we obtain

$$\begin{aligned} \hat{\pi}_t &= \left[1 - \frac{(1+\nu)\kappa_p \psi_p/\beta}{1 - \psi_p \rho_\phi}\right] \frac{\kappa_p}{1 - \beta \rho_\phi}\phi_t + \left[1 - \frac{(1+\nu)\kappa_p \psi_p/\beta}{1 - \psi_p \rho_\lambda}\right] \frac{\kappa_p}{1 - \beta \rho_\lambda}\lambda_t \\ &+ \frac{\kappa_p (1+\nu)\rho_z \psi_p}{(1 - \psi_p \rho_z)(1 - \beta \rho_z)} z_t - \kappa_p (1+\nu)\psi_p \sigma_R \epsilon_{R,t}. \end{aligned} \tag{37}$$

Finally, combining (38) with the monetary policy rule (29) yields the solution for the nominal interest rate

$$\begin{aligned} \hat{R}_t &= \frac{1}{\beta}\left[1 - \frac{(1+\nu)\kappa_p \psi_p/\beta}{1 - \psi_p \rho_\phi}\right] \frac{\kappa_p}{1 - \beta \rho_\phi}\phi_t + \frac{1}{\beta}\left[1 - \frac{(1+\nu)\kappa_p \psi_p/\beta}{1 - \psi_p \rho_\lambda}\right] \frac{\kappa_p}{1 - \beta \rho_\lambda}\lambda_t \\ &+ \frac{\kappa_p (1+\nu)\rho_z \psi_p}{\beta(1 - \psi_p \rho_z)(1 - \beta \rho_z)} z_t + (1 - \kappa_p (1+\nu)\psi_p/\beta)\sigma_R \epsilon_{R,t}. \end{aligned} \tag{38}$$

The DSGE model solution for the detrended variables comprises (34), (35), (38), and (39).

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. If the observables consist of output growth, the labor share, net inflation rates and net interest rates, then the measurement equations take the form

$$\begin{aligned} \ln(X_t/X_{t-1}) &= \hat{x}_t - \hat{x}_{t-1} + z_t + \ln\gamma \\ \ln(lsh_t) &= \widehat{lsh}_t + \ln(lsh) \\ \ln \pi_t &= \hat{\pi}_t + \ln \pi^* \\ \ln R_t &= \hat{R}_t + \ln(\pi^*\gamma/\beta) \end{aligned} \tag{39}$$

The DSGE model solution has the form of a generic state space model. Define the vector of 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

$$\theta = [\beta, \gamma, \lambda, \pi^*, \zeta_p, \nu, \rho_\phi, \rho_\lambda, \rho_z, \sigma_\phi, \sigma_\lambda, \sigma_z, \sigma_R]'. \tag{40}$$

For now 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

$$s_t = \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t, \tag{41}$$

where $\epsilon_t = [\epsilon_{\phi,t}, \epsilon_{\lambda,t}, \epsilon_{z,t}, \epsilon_{R,t}]'$. The coefficient matrices $\Phi_1(\theta)$ and $\Phi_\epsilon(\theta)$ are determined by (20), (24), (27), the identity $\epsilon_{R,t} = \epsilon_{R,t}$, and a lagged version of (34). If we define the vector of observables as

$$y_t = M_y'[\ln(X_t/X_{t-1}), \ln lsh_t, \ln \pi_t, \ln R_t]', \tag{42}$$

where M_y' is a matrix that selects rows of the vector $[\ln(X_t/X_{t-1}), \ln lsh_t, \ln \pi_t, \ln R_t]'$ then the measurement equation can be written as

$$y_t = \Psi_0(\theta) + \Psi_1(\theta)s_t. \tag{43}$$

The coefficient matrices $\Psi_0(\theta)$ and $\Psi_1(\theta)$ can be obtained from (39), the equilibrium law of motion for the detrended model variables given by (34), (35), (38), and (39). They are summarized in Table 1.

Table 1: System Matrics for DSGE Model

State space representation:

$$
\begin{aligned}
y_t &= \Psi_0(\theta) + \Psi_1(\theta) s_t \\
s_t &= \Phi_1(\theta) s_{t-1} + \Phi_\epsilon(\theta) \epsilon_t
\end{aligned}
$$

System matrices:

$$
\Phi_0(\theta) = M_y' \begin{bmatrix} \ln \gamma \\ \ln(lsh) \\ \ln \pi^* \\ \ln(\pi^* \gamma/\beta) \end{bmatrix}, \quad \Phi_{\hat{x}\phi} = \frac{\kappa_p \psi_p/\beta}{1-\psi_p \rho_\phi}, \quad \Phi_{\hat{x}\lambda} = \frac{\kappa_p \psi_p/\beta}{1-\psi_p \rho_\lambda}, \quad \Phi_{\hat{x}z} = \frac{\rho_z \psi_p}{1-\psi_p \rho_z}
$$

$$
\Phi_1(\theta) = M_y' \begin{bmatrix}
-\Phi_{\hat{x}\phi} & -\Phi_{\hat{x}\lambda} & \Phi_{\hat{x}z}+1 & -\psi_p & -1 \\
1-(1+\nu)\Phi_{\hat{x}\phi} & -(1+\nu)\Phi_{\hat{x}\lambda} & (1+\nu)\Phi_{\hat{x}z} & -(1+\nu)\psi_p & 0 \\
(1-(1+\nu)\Phi_{\hat{x}\phi})\frac{\kappa_p}{1-\beta\rho_\phi} & (1-(1+\nu)\Phi_{\hat{x}\lambda})\frac{\kappa_p}{1-\beta\rho_\lambda} & (1+\nu)\Phi_{\hat{x}z}\frac{\kappa_p}{1-\beta\rho_z} & -\kappa_p(1+\nu)\psi_p & 0 \\
(1-(1+\nu)\Phi_{\hat{x}\phi})\frac{\kappa_p/\beta}{1-\beta\rho_\phi} & (1-(1+\nu)\Phi_{\hat{x}\lambda})\frac{\kappa_p/\beta}{1-\beta\rho_\lambda} & (1+\nu)\Phi_{\hat{x}z}\frac{\kappa_p/\beta}{1-\beta\rho_z} & (1-\kappa_p(1+\nu)\psi_p/\beta) & 0
\end{bmatrix}
$$

$$
\Psi_1(\theta) = \begin{bmatrix}
\rho_\phi & 0 & 0 & 0 & 0 \\
0 & \rho_\lambda & 0 & 0 & 0 \\
0 & 0 & \rho_z & 0 & 0 \\
0 & 0 & 0 & 0 & 0 \\
-\Phi_{\hat{x}\phi} & -\Phi_{\hat{x}\lambda} & \Phi_{\hat{x}z} & -\psi_p & 0
\end{bmatrix}, \; \psi_\epsilon(\theta) = \begin{bmatrix}
\sigma_\phi & 0 & 0 & 0 \\
0 & \sigma_\lambda & 0 & 0 \\
0 & 0 & \sigma_z & 0 \\
0 & 0 & 0 & \sigma_R \\
0 & 0 & 0 & 0
\end{bmatrix}
$$

M_y' is a $n_y \times 4$ selection matrix that selects rows of Φ_0 and Φ_1.

The state-space representation of the DSGE model given by (41) and (43) provide the basis for the subsequent econometric analysis. We conclude this section with four remarks. First, it is useful 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 (41) and (43). The economic state variables of our simple model are all exogenous. A richer DSGE model typically includes endogenous variables such as the capital stock or lagged inflation and interest rates as additional state variables. 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 $\ln Z_t/\gamma^t$ to s_t, and accounting

Table 2: DSGE Model Parameters Used for Illustration

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

for the deterministing trend component $(\ln \gamma)t$ in log output in the measurement equation. Third, the measurement equation (43) could be augmented by measurement errors. Fourth, if the DSGE model is solved with a nonlinear solution techniques, then, depending on how exactly the state vector s_t is defined, the state-transition equation (41), the measurement equation (43), or both are nonlinear. We will discuss how the various econometric procedures need to be adjusted for the presence of nonlinearities throughout the subsequent sections.

9.2 Model Implications

Once we specify a distribution for the vector ϵ_t the probability distribution of the DSGE model variables is fully determined. For the sake of concreteness, we assume for now that

$$\epsilon_t \sim iidN(0, I), \tag{44}$$

where I denotes the identity matrix. Recall that the innovation standard deviations were absorbed into the definition of the matrix $\Phi_\epsilon(\theta)$ in (43). Based on the probabilistic structure of the DSGE model we can derive a number of implications from the DSGE model that can later on be used to construct estimators of the parameter vector θ and evaluate the fit of the model.

9.2.1 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 (41) and (43). We will assume that the largest (in absolute value) eigenvalue of the matrix $\Phi_1(\theta)$ in (41) is less than one. It can be verified that this assumption is consistent with commonly imposed restrictions on the DSGE model parameters:

$$0 \leq \beta < 1, \quad \kappa > 0, \quad 1 + \nu \geq 0, \quad 0 \leq \rho_\phi < 1, \quad 0 \leq \rho_\lambda < 1, \quad 0 \leq \rho_z < 1.$$

This implies that the VAR(1) law of motion for s_t is covariance stationary (provided the innovation sequence is covariance stationary and the process has been initialized in the infinite past). Using the notation

$$\Gamma_{yy}(h) = \mathbb{E}[y_t y_{t-h}], \quad \Gamma_{ss}(h) = \mathbb{E}[s_t s_{t-h}], \quad \Gamma_{ys}(h) = \mathbb{E}[y_t s'_{t-h}], \quad \text{and} \quad \mathbb{E}[\epsilon_t \epsilon'_t] = \Sigma_\epsilon,$$

we can express the autocovariance matrix of s_t as the solution to the following Riccati equation:[10]

$$\Gamma_{ss}(0) = \Phi_1 \Gamma_{ss}(0) \Phi'_1 + \Phi_\epsilon \Phi'_\epsilon. \tag{45}$$

Here we used $\mathbb{E}[\epsilon_t \epsilon'_t] = I$, see (44). Once the covariance matrix of s_t has been determined, it is straightforward to compute the autocovariance matrices for $h \neq 0$ according to

$$\Gamma_{ss}(h) = \Phi_1^h \Gamma_{ss}(0). \tag{46}$$

Finally, using the measurement equation 43, we deduce that

$$\Gamma_{yy}(h) = \Psi_1 \Gamma_{ss}(h) \Psi'_1, \quad \Gamma_{ys}(h) = \Psi_1 \Gamma_{ss}(h). \tag{47}$$

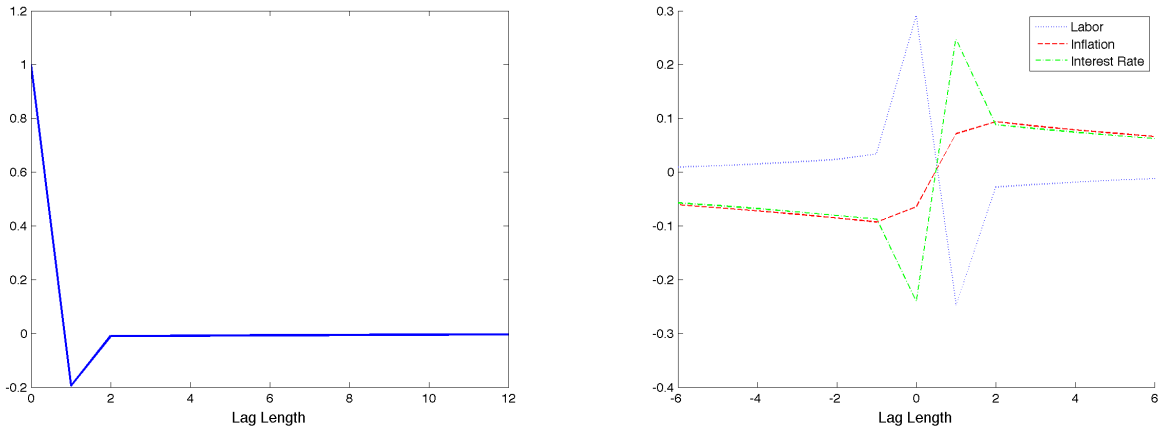
Correlations can be easily computed by normalizing the entries of the autocovariance matrices using the respective standard deviations.

Illustration: Figure 1 shows the model-implied autocorrelation function of output growth and the cross-correlations of output growth with the labor share, inflation, and interest rates. *** Elaborate ***

[10] Efficient numerical routines to solve Riccati equations are readily available in many software packages

Figure 1: Autocorrelations

$\mathrm{Corr}\big(\ln(X_t/X_{t-1}), \ln(X_{t-h}/X_{t-h-1})\big)$ $\mathrm{Corr}\big(\ln(X_t/X_{t-1}), \ln Z_{t-h}\big)$



Notes: y-axis: autocorrelation; x-axis displacement h.

The law of motion for the state vector s_t can also be expressed as the infinite-order vector moving average (MA) process

$$y_t = \Psi_0 + \Psi_1 \sum_{s=0}^{\infty} \Phi_1^s \Phi_\epsilon \epsilon_{t-s}. \tag{48}$$

Based on the moving average representation, it is straightforward to compute the h-step ahead forecast error, which is given by

$$e_{t|t-h} = y_t - \mathbb{E}_{t-s}[y_t] = \Psi_1 \sum_{s=0}^{h-1} \Phi_1^s \Phi_\epsilon \epsilon_{t-s}. \tag{49}$$

The h-step ahead forecast error covariance matrix is given by

$$\mathbb{E}[e_{t|t-h} e_{t|t-h}'] = \Psi_1 \left(\sum_{s=0}^{h-1} \Phi_1^s \Phi_\epsilon \Phi_\epsilon' \Phi_1^{s'} \right) \Psi_1' \quad \text{with} \quad \lim_{h \longrightarrow \infty} \mathbb{E}[e_{t|t-h} e_{t|t-h}'] = \Gamma_{ss}(0). \tag{50}$$

Provided that the innovations ϵ_t are uncorrelated, 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 I to zero. Then we can write

$$I = \sum_{i=1}^{n_\epsilon} I^{(i)}. \tag{51}$$

Moreover, we can express the contribution of shock j to the forecast error for y_t as

$$e_{t|t-h}^{(j)} = \Psi_1 \sum_{s=0}^{h-1} \Phi_1^s \Phi_\epsilon I^{(j)} \epsilon_{t-s}. \tag{52}$$

Thus, the contribution of shock j to the forecast error variance of observation $y_{i,t}$ is given by the ratio

$$\text{FEVD}(i,j,h) = \frac{\left[\Psi_1 \left(\sum_{s=0}^{h-1} \Phi_1^s \Phi_\epsilon I_{(j)} \Phi_\epsilon' \Phi_1^{s'}\right) \Psi_1'\right]_{ii}}{\left[\Psi_1 \left(\sum_{s=0}^{h-1} \Phi_1^s \Phi_\epsilon \Phi_\epsilon' \Phi_1^{s'}\right) \Psi_1'\right]_{ii}}, \tag{53}$$

where $[A]_{ij}$ denotes element (i,j) of a matrix A.

Illustration. Figure 2 shows the contribution of the four shocks to the forecast error variance of output growth, the labor share, inflation, and interest rates. *** elaborate ***

9.2.2 Spectrum

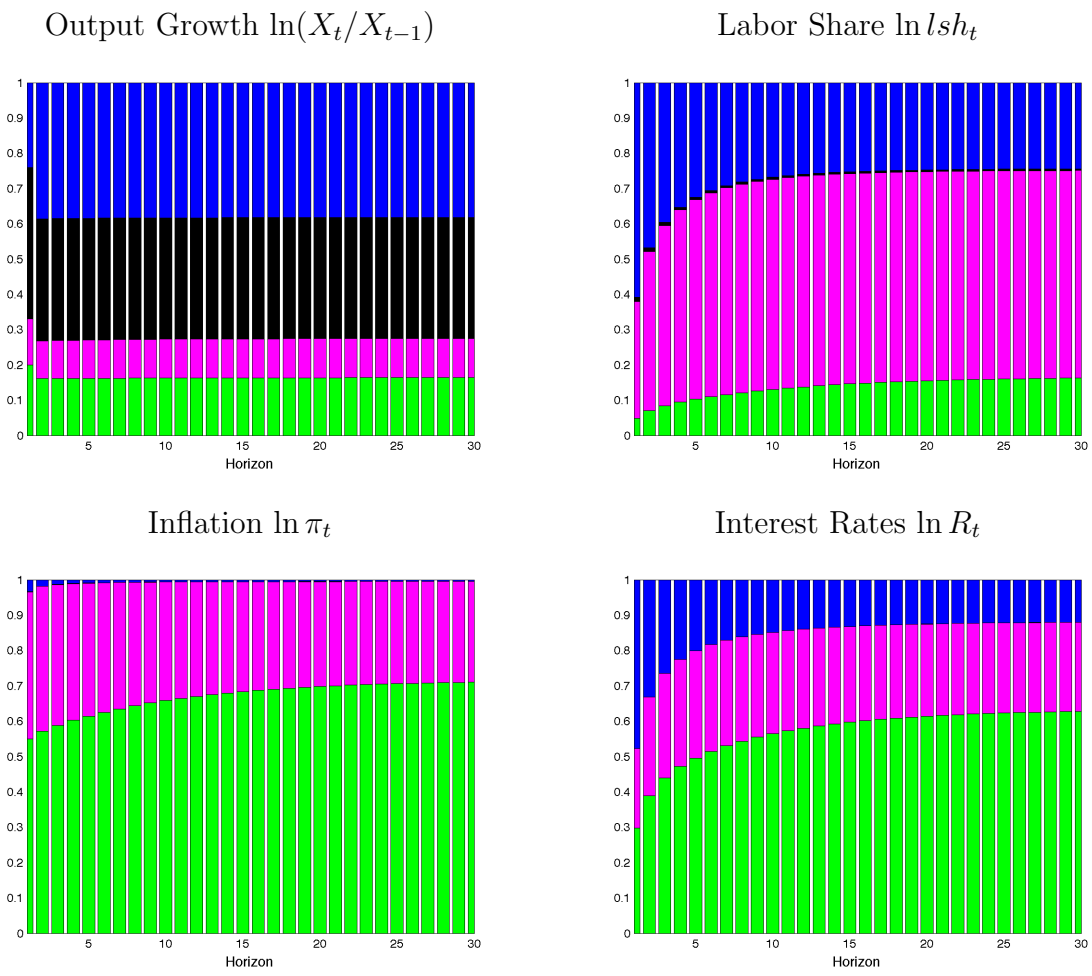
The analysis of the forecast error covariance matrix over different horizons h highlights that the uncertainty about macroeconomic outcomes implied by DSGE models varies with the horizon. Moreover, it is conceivable that some of the structural shocks are important over short horizons, whereas others are more important over long horizons. *** provide some references on spectral analysis *** Before providing a general formula for the spectral density function for a DSGE model, we provide a brief discussion of the linear cyclical model, which will be useful for interpreting some of the formulas presented below.

Suppose that y_t is a scalar time series that follows the process

$$y_t = 2 \sum_{j=1}^{m} a_j \big(\cos\theta_j \cos(\omega_j t) - \sin\theta_j \sin(\omega_j t) \big), \tag{54}$$

where $\theta_j \sim iidU[-\pi, \pi]$ and $0 \leq \omega_j \leq \omega_{j+1} \leq \pi$. The random variables θ_j are determined in the infinite past and cause a phase shift of the cycle. In a nutshell, the model in (54) 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 are typically comprised of cycles that have a duration of 8 to 32 quarters, which would correspond to $\omega_j \in [0.196, 0.785]$ for quarterly t.

Figure 2: Forecast Error Variance Decomposition



Notes: For each panel: plot "stacked bar plot" that captures the forecast error variance decomposition. Green is ϕ_t, pink is λ_t, black is z_t, blue is $\epsilon_{R,t}$.

Using Euler's formula, we rewrite the cyclical model in terms of an exponential function. This alternative representation is less intuitive but mathematically more convenient.

$$y_t = \sum_{j=-m}^{m} A(\omega_j) e^{i\omega_j t}, \tag{55}$$

where $\omega_{-j} = -\omega_j$ and $i = \sqrt{-1}$. Let $a_{-j} = a_j$ and

$$A(\omega_j) = \begin{cases} a_j(\cos\theta_{|j|} + i\sin\theta_{|j|}) & \text{if } j > 0 \\ a_j(\cos\theta_{|j|} - i\sin\theta_{|j|}) & \text{if } j < 0 \end{cases}. \tag{56}$$

It can be verified that expressions (54) and (55) are identical by showing that

$$A(\omega_j)e^{i\omega_j t} + A(\omega_{-j})e^{-i\omega_j t} = 2\left[a_j \cos\theta_j \cos(\omega_j t) - a_j \sin\theta_j \sin(\omega_j t)\right] \tag{57}$$

for $j = 1, \ldots, m$.

The spectral distribution function of y_t on the interval $\omega \in (-\pi, \pi]$ is defined as

$$F_{yy}(\omega) = \sum_{j=-m}^{m} \mathbb{E}[A(\omega_j)\overline{A(\omega_j)}]\mathbb{I}\{\omega_j \leq \omega\}, \tag{58}$$

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

$$f_{yy}(\omega) = dF_{yy}(\omega)d\omega. \tag{59}$$

If a process has a spectral density function $s_{yy}(\omega)$ then the covariances can be expressed as

$$\Gamma_{yy}(h) = \int_{(-\pi,\pi]} e^{ih\omega} f_{yy}(\omega)d\omega. \tag{60}$$

For the linear cyclical model in (54) the spectral It can also be shown that the autovariances are given by

$$\Gamma_{yy}(h) = \sum_{j=-m}^{m} \mathbb{E}[A(\omega_j)\overline{A(\omega_j)}]e^{i\omega_j h} = \sum_{j=-m}^{m} a_j^2 e^{i\omega_j h}. \tag{61}$$

The spectral density uniquely determines the entire sequence of autocovariances. Moreover, the converse is also true. The spectral density can be obtained by evaluating the autocovariance generating function of y_t at $z = e^{-i\omega}$:

$$f_{yy}(\omega) = \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Gamma_{yy}(h)e^{-i\omega h}.$$

The formulas presented for a scalar process y_t have a straightforward extension to the vector case. Recall that for our DSGE model defined by the state-space system (41) and (43) the autocovariance function for the state vector s_t was defined as $\Gamma_{ss}(h) = \Phi_1^h \Gamma_{ss}(0)$. Thus,

$$\begin{aligned} f_{ss}(\omega) &= \frac{1}{2\pi} \sum_{h=-\infty}^{\infty} \Phi_1^h \Gamma_{yy}(0)e^{-i\omega h} \\ &= \frac{1}{2\pi}\left(I - \Phi_1' e^{i\omega}\right)^{-1} \Phi_\epsilon \Phi_\epsilon' \left(I - \Phi_1 e^{-i\omega}\right)^{-1}. \end{aligned} \tag{62}$$

The contribution of shock j to the spectral density is given by

$$f_{ss}^{(j)}(\omega) = \frac{1}{2\pi}\big(I - \Phi_1' e^{i\omega}\big)^{-1}\Phi_\epsilon \mathcal{I}^{(j)}\Phi_\epsilon'\big(I - \Phi_1 e^{-i\omega}\big)^{-1}. \tag{63}$$

The spectral density for the observables y_t (and the contribution of shock j to the spectral density) can be easily obtained as

$$f_{yy}(\omega) = \Psi_1 f_{ss}(\omega)\Psi_1' \quad \text{and} \quad f_{yy}^{(j)}(\omega) = \Psi_1 f_{ss}^{(j)}(\omega)\Psi_1'. \tag{64}$$

Illustration: Figure 3 depicts the DSGE model-implied spectral density functions for output growth, the labor share, inflation, and interest rates. Each panel stacks the contributions of the four shocks to the spectral densities. *** elaborate ***

9.2.3 Impulse Response Functions

An important tool for studying the effects of innovations to exogenous shocks onto the endogenous DSGE model variables and the observables y_t are impulse response functions. Formally, impulse responses in a DSGE model can be defined as the difference between two conditional expectations:

$$\text{IRF}(i,j,h) = \mathbb{E}\big[y_{i,t+h} \,\big|\, s_{t-1}, \epsilon_{j,t+1} = 1\big] - \mathbb{E}\big[y_{i,t+h} \,\big|\, s_{t-1},\big]. \tag{65}$$

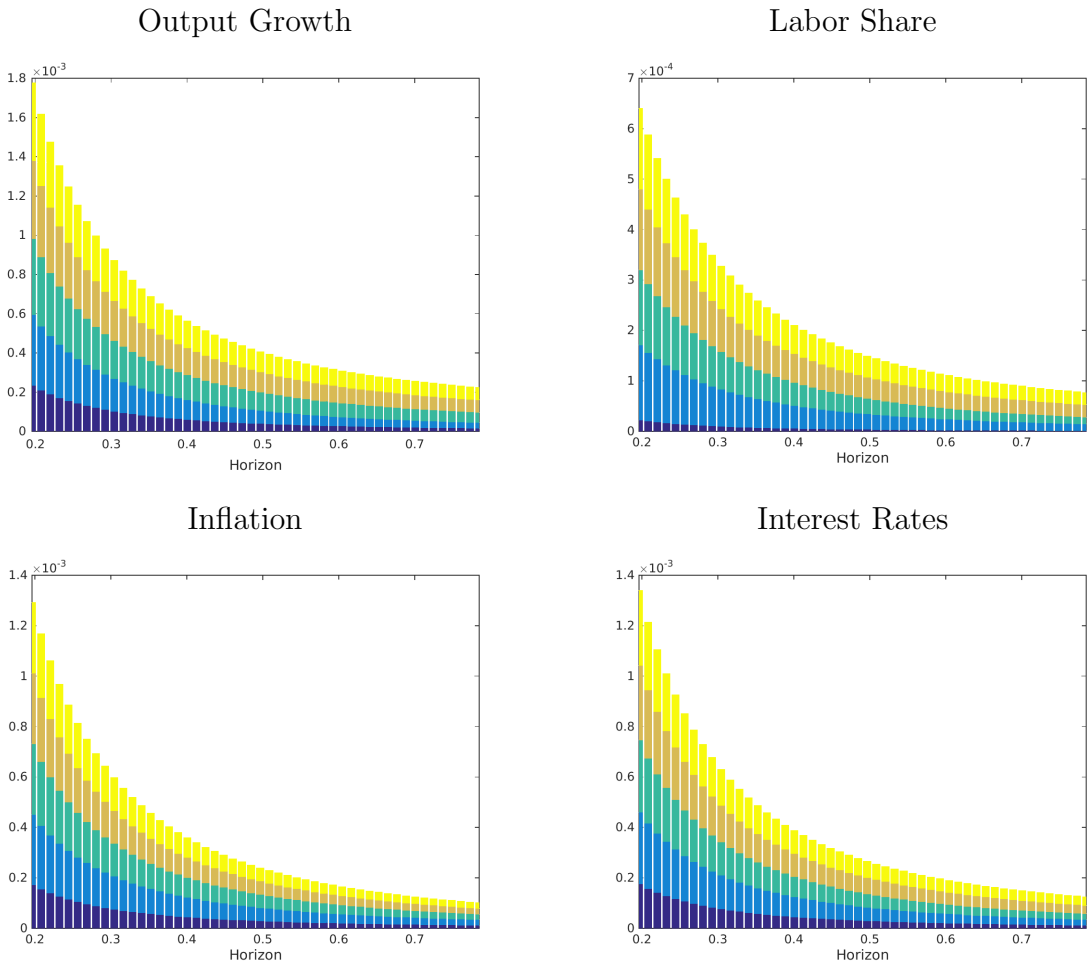
Both expectations are conditional on the initial state s_{t-1} and integrate over current and future realizations of the shocks ϵ_t. However, the first term also conditions on $\epsilon_{j,t} = 1$, whereas the second term averages of $\epsilon_{j,t}$. In a linearized DSGE model with a state space representation of the form (41) and (43), 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

$$\text{IRF}(.,j,h) = \Psi_1 \frac{\partial}{\partial \epsilon_{j,t}} s_{t+h} = \Psi_1 \Phi_1^h [\Phi_\epsilon]_{.j}, \tag{66}$$

where $[A]_{.j}$ is the j'th column of matrix A.

Illustration: Figure 4 depicts the impulse response function of log output to the four structural shocks. *** elaborate ***

Figure 3: Spectral Decomposition



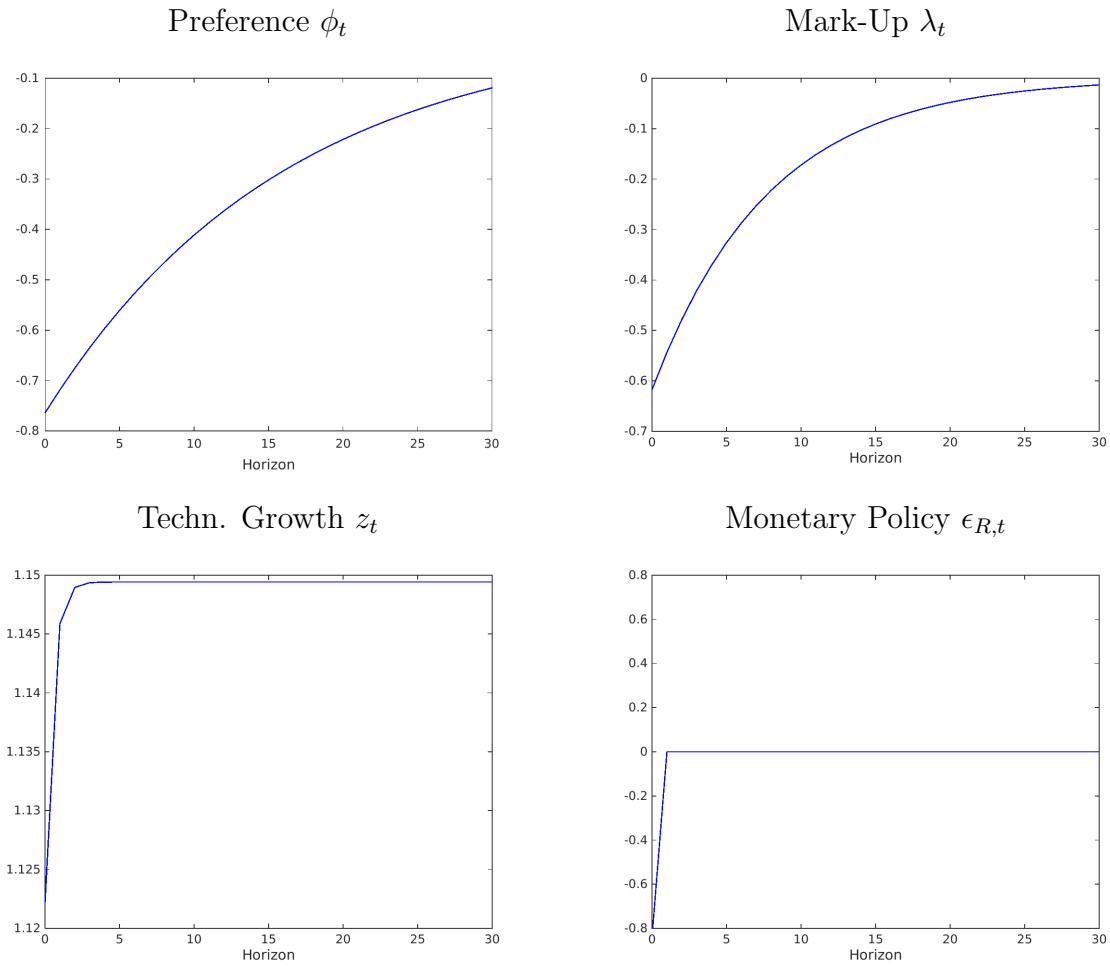
Notes: For each panel: plot "stacked lines plot" that capture the cumulative contribution of the shocks to the spectral density. Top line should correspond to total spectral density, bottom line to contribution of shock 1 to spectral density. Basically, a line version of the stacked bar plot.

9.2.4 Conditional Moment Restrictions

The intertemporal optimality conditions take the form of conditional moment restrictions. For instance, re-arranging the terms in the New Keynesian Phillips (26) curve a bit, we can write

$$\mathbb{E}_{t-1}\left[\hat{\pi}_{t-1} - \beta\hat{\pi}_t - \kappa_p(\widehat{lsh}_{t-1} + \lambda_{t-1})\right] = 0. \tag{67}$$

Figure 4: Impulse Responses of Output Growth $100\ln(X_t/X_{t-1})$



Notes:

The conditional moment condition can be converted into an unconditional moment condition as follows. Let $\mathcal{F}_t$ denote the sigma algebra generated by the infinite histories of $\{y_\tau, s_\tau, \epsilon_\tau\}_{\tau=-\infty}^{t}$ and let z_t *** this is bad notation because z_t is technology growth *** be a random variable that is measurable with respect to $\mathcal{F}_t$, meaning that its value is determined based on information on current and past (y_t, s_t, ϵ_t). Then for every such z_t,

$$\mathbb{E}\big[z_t\big(\hat{\pi}_{t-1} - \beta\hat{\pi}_t - \kappa_p(\widehat{lsh}_{t-1} + \lambda_{t-1})\big)\big] = \mathbb{E}\big[z_t\mathbb{E}_{t-1}\big[\hat{\pi}_{t-1} - \beta\hat{\pi}_t - \kappa_p(\widehat{lsh}_{t-1} + \lambda_{t-1})\big]\big] = 0. \quad (68)$$

The moment condition derived from the New Keynesian Phillips curve involves the latent price markup shock λ_t, which will cause difficulties if one tries to use the condition in a

estimation objective function. Now consider the Euler equation (22) instead. Recall from the definition of detrended output in deviations from its steady state

$$\hat{x}_t - \hat{x}_{t-1} + z_{t+1} = \ln X_t - \ln X_{t-1} - \ln \gamma.$$

Thus, we can write

$$\mathbb{E}_{t-1}\big[- \ln(X_t/X_{t-1}) + \ln R_{t-1} - \ln \pi_t - \ln(1/\beta)\big] = 0. \tag{69}$$

The terms γ and $\ln \pi^*$ 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

$$\mathbb{E}_{t-1}\big[\ln R_t - \ln(\gamma/\beta) - \psi \ln \pi_t - (1-\psi) \ln \pi^*] = 0. \tag{70}$$

9.2.5 Analytical Calculation of Moments versus Simulation Approximations

- In linearized DSGE models with Gaussian innovations we can compute the autocovariance function and impulse response functions directly from the state-space representation of the DSGE model.
- In DSGE models solved with perturbation methods the moments can also be computed analytically. See Andreasen, Fernández-Villaverde, and Rubio-Ramírez (2013)
- In general nonlinear DSGE models the moments have to be computed by simulation. If the moment as a function of the parameter is of interest, then fixing the random seed may be useful.

9.3 Empirical Analogues

Our goal is to confront the DSGE model with observed macroeconomic data to determine suitable numerical values for the parameter vector θ and to assess the fit of the model. We previously provided a characterization of the joint distribution of observable and latent model variables denoted by $p(Y_{1:T}, S_{1:T}|\theta)$. Based on this distribution we computed various

moments, including autocovariances, the spectrum, and impulse response functions. Confronting the DSGE model 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 far 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 strong believe (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 9.2 and compare these estimates to their model counterparts. Before delving into the intricacies of econometric inference, we discuss some of the sample analogues of the population objects introduced in Section 9.2.

All data was downloaded from FRED, and we record the series name in parentheses. For 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. For Labor Share, we use Compensation of Employees (COE) divided by nominal GDP (GDP), both of which are quarterly and seasonally adjusted at the annual rate. We use the log labor share as the observable. For Inflation, we use the implicit price deflator (GDPDEF), which is seasonally adjusted. We also use growth rates for the deflator, and again we turn it from levels into growth rates by taking log differences. Lastly, for the interest rate, we use the Effective Federal Funds Rate (FEDFUNDS), which is monthly, and not seasonally adjusted. To turn the series quarterly, we take the average.

9.3.1 Autocovariances

The sample analogue of the population autocovariance $\Gamma_{yy}(h)$ is defined as

$$\hat{\Gamma}_{yy}(h) = \frac{1}{T}\sum_{t=h}^{T}(y_t - \hat{\mu}_y)(y_{t-h} - \hat{\mu}_y)', \quad \text{where} \quad \hat{\mu}_y = \frac{1}{T}\sum_{t=1}^{T} y_t. \tag{71}$$

Under suitable regularity conditions, which include the ergodicity and stationarity of the vector process y_t, the sufficiently fast decay of the serial correlation in y_t, and some bounds

on higher-order moments of y_t, the sample autocovariance $\hat{\Gamma}_{yy}(h)$ converges to the population autocovariance $\Gamma_{yy}(h)$ and satisfies a central limit theorem (CLT).

If the object of interest is the autocovariance function for $h = 0, \ldots, H$, then it might be more efficient to obtain these estimates indirectly by first estimating an auxiliary model and then converting the parameter estimates of the auxiliary model into estimates of the autocovariance. A natural class of auxiliary models is provided by linear vector autoregressions (VARs). For illustrative purposes consider the following VAR(1):

$$y_t = \Phi_1 y_{t-1} + \Phi_0 + u_t, \quad u_t \sim iid(0, \Sigma). \tag{72}$$

Using the Frisch-Waugh-Lovell Theorem, we can approximate the OLS estimator of Φ_1 by

$$\hat{\Phi}_1 = \hat{\Gamma}_{yy}(1)\big(\hat{\Gamma}_{yy}(0)\big)^{-1} + O_p(T^{-1}), \quad \hat{\Sigma} = \hat{\Gamma}_{yy}(0) - \hat{\Gamma}_{yy}(1)\big(\hat{\Gamma}_{yy}(0)\big)^{-1}\hat{\Gamma}'_{yy}(1) + O_p(T^{-1}) \tag{73}$$

The $O_p(T^{-1})$ terms arise because the range of the summations in the definition of the sample autocovariances in (71) and the definition of the OLS estimator are not exactly the same. Suppose now, we plug the OLS estimator into the autocovariance formulas associated with the VAR(1), see (45) and (46), then, up to an $O_p(T^{-1})$ term, we recover the sample autocovariances for $h = 0$ and use $\hat{\Phi}_1^h\hat{\Gamma}_{yy}(0)$ as an estimate of the autocovariance of order h:

$$\hat{\Gamma}^V_{yy}(0) = \hat{\Gamma}_{yy}(0) + O_p(T^{-1}), \quad \hat{\Gamma}^V_{yy}(h) = \Big(\hat{\Gamma}_{yy}(1)\big(\hat{\Gamma}_{yy}(0)\big)^{-1}\Big)^h \hat{\Gamma}_{yy}(0) + O_p(T^{-1}). \tag{74}$$

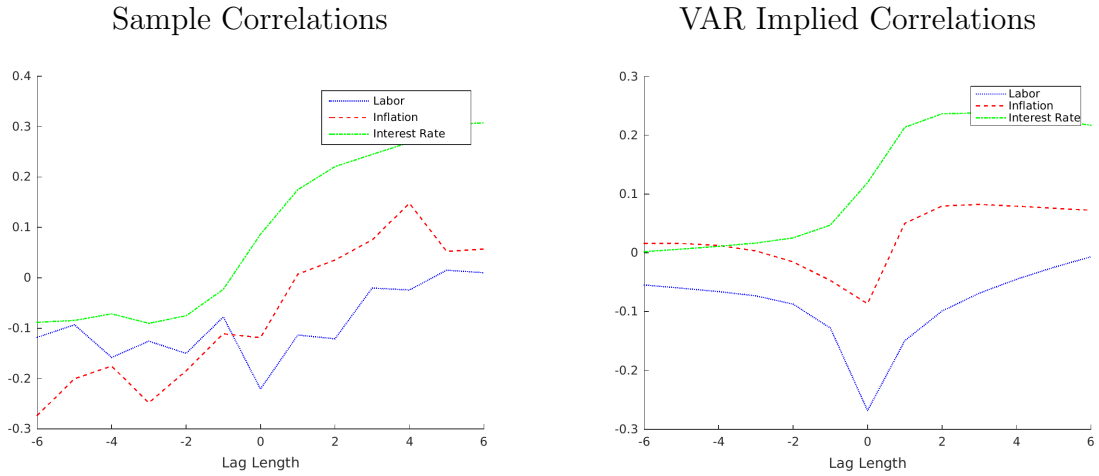
Note that for $h = 1$ we obtain $\hat{\Gamma}^V_{yy}(1) = \hat{\Gamma}_{yy}(1) + O_p(T^{-1})$. For $h > 1$ the VAR(1) plug-in estimate of the autocovariance matrix differs from the sample autocovariance matrix. If the actual time series are well approximated by a VAR(1), then the plug-in autocovariance estimate tends to be more efficient than the direct sample autocovariance estimate. A formal analysis in the context of multi-step forecasting is provided by 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)

$$y_t = \Phi_1 y_{t-1} + \ldots + \Phi_p y_{t-p} + \Phi_0 + u_t, \quad u_t \sim iid(0, \Sigma). \tag{75}$$

The appropriate lag length p could be determined with a model selection criterion, e.g., the Schwarz (1978) criterion. The notationally easiest way (but not the computationally fastest

Figure 5: Empirical Crosscorrelations



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):

$$\tilde{y}_t = \tilde{\Phi}_1 \tilde{y}_{t-1} + \tilde{\Phi}_0 + \tilde{u}_t, \quad \tilde{u}_t \sim iid(0, \tilde{\Sigma}), \tag{76}$$

where

$$\tilde{\Phi}_1 = \begin{bmatrix} \Phi_1 & \ldots & \Phi_{p-1} & \Phi_p \\ I_{n\times n} & \ldots & 0_{n\times n} & 0_{n\times n} \\ \vdots & \ddots & \vdots & \vdots \\ 0_{n\times n} & \ldots & I_{n\times n} & 0_{n\times n} \end{bmatrix}, \quad \tilde{\Phi}_0 = \begin{bmatrix} \Phi_0 \\ 0_{n(p-1)\times 1} \end{bmatrix},$$

$$\tilde{\epsilon}_t = \begin{bmatrix} \epsilon_t \\ 0_{n(p-1)\times 1} \end{bmatrix}, \quad \tilde{\Sigma} = \begin{bmatrix} \Sigma & 0_{n\times n(p-1)} \\ 0_{n(p-1)\times n} & 0_{n(p-1)\times n(p-1)} \end{bmatrix}.$$

The autocovariances for $\tilde{y}_t$ are then obtained by adjusting the VAR(1) formulas (74) 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 $y_t = M'\tilde{y}_t$.

Illustration: We estimate a VAR for output growth, labor share, inflation, and interest rates, based on a sample from 1984:Q1 to 2007:Q4. The lag length $p = 1$ is determined by BIC. Figure 5 shows sample cross correlations between output growth and leads and lags of the other three variables as well as correlations derived from the estimated VAR.

9.3.2 Spectrum

An intuitively plausible estimate of the spectrum is the sample periodogram, defined as

$$\hat{f}_{yy}(\omega) = \frac{1}{2\pi} \sum_{h=-T+1}^{T-1} \hat{\Gamma}_{yy}(h) e^{-i\omega h} = \frac{1}{2\pi} \left(\hat{\Gamma}_{yy}(0) + \sum_{h=1}^{T-1} (\hat{\Gamma}_{yy}(h) + \hat{\Gamma}_{yy}(h)') \cos \omega h \right). \quad (77)$$

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 $T \longrightarrow \infty$. 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

$$\bar{f}_{yy}(\omega) = \frac{\pi}{\lambda(T-1)/2} \sum_{j=1}^{(T-1)/2} K\left(\frac{\omega_j - \omega}{\lambda} \right) \hat{f}_{yy}(\omega_j). \quad (78)$$

An example of a kernel function is

$$K\left(\frac{\omega_j - \omega}{\lambda} \right) \hat{f}_{yy}(\omega_j) = \mathbb{I}\left\{ -\frac{1}{2} < \frac{\omega_j - \omega}{\lambda} < \frac{1}{2} \right\} = \mathbb{I}\{\omega_j \in B(\omega|\lambda)\},$$

where $B(\omega|\lambda)$ is a frequency band. The smoothed periodogram estimator $\bar{f}_{yy}(\omega)$ is consistent, provided that the bandwidth shrinks to zero, that is, $\lambda \longrightarrow 0$ as $T \longrightarrow \infty$, and the number of ω_j's in the band, given by $\lambda T(2\pi)$, tends to infinity.

An estimate of the spectral density can also be obtained indirectly through the estimation of the VAR(p) in (75). Define

$$\Phi = [\Phi_1, \ldots, \Phi_p, \Phi_0]' \quad \text{and} \quad M(z) = [Iz, \ldots, Iz^p],$$

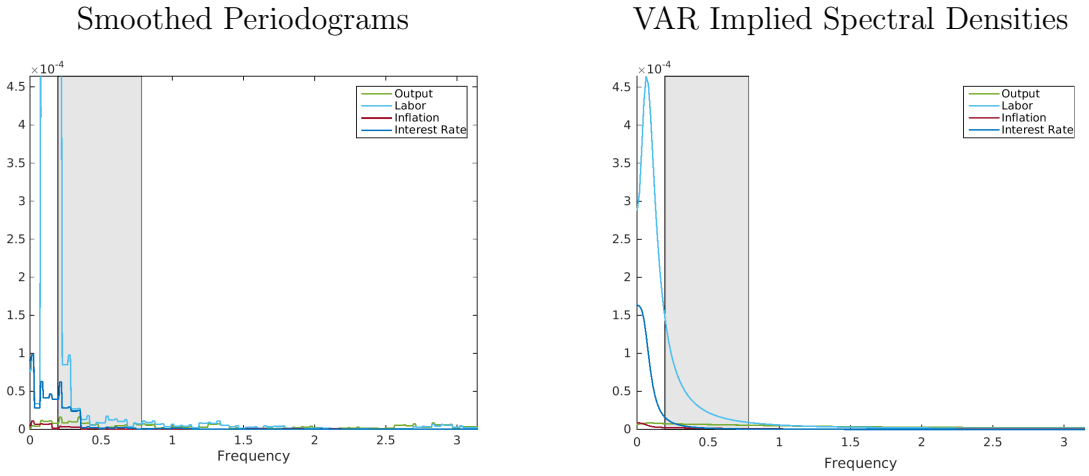
and let $\hat{\Phi}$ be an estimator of Φ. Then a VAR(p) plug-in estimator of the spectral density is given by

$$\hat{f}^V_{yy}(\omega) = \frac{1}{2\pi} [I - \hat{\Phi}' M'(e^{-i\omega})]^{-1} \hat{\Sigma} [I - M(e^{-i\omega})\hat{\Phi}]^{-1}. \quad (79)$$

This formula generalizes the VAR(1) spectral density in (62) to a spectral density for a VAR(p).

Illustration: Figure 6 depicts the VAR-based estimate of the spectrum and the smoothed sample periodogram. The sample period is 1984:Q1 to 2007:Q4. The VAR has one lag. The shaded area indicates business cycle frequencies.

Figure 6: Empirical Spectrum



9.3.3 Impulse Response Functions

Empirical analogues for the autocovariance function and the spectral density can be obtained either by computing sample autocovariances and the smoothed periodogram or by converting the parameters of an estimated VAR(p) into the objects of interest. No additional assumptions were required. Obtaining the empirical counterparts of the DSGE model impulse response functions requires an auxiliary model upon which one has to impose some additional structure that is not directly identifiable from the data. The VAR(p) in (75) 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 innovations in the structural shock innovations that contribute to the forecast error for several observables simultaneously. Thus, we express the one-step-ahead forecast errors as a linear combination of the structural innovations ϵ_t:

$$u_t = \Phi_\epsilon \epsilon_t = \Sigma_{tr} \Omega \epsilon_t, \tag{80}$$

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. The second equality ensures that the covariance matrix of u_t is preserved in the sense that

$$\Phi_\epsilon \Phi_\epsilon' = \Sigma_{tr} \Omega \Omega' \Sigma_{tr}' = \Sigma. \tag{81}$$

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, Eichenbaum, and Evans (1999), and Stock and Watson (2001). Conditional on an estimate of the reduced form coefficient matrices Φ and Σ and an identification scheme Ω, it is straightforward to express the impulse response as

$$\widehat{IRF}^V(.,j,h) = C_h(\hat{\Phi})\hat{\Sigma}_{tr}[\Omega]_{.j}, \tag{82}$$

where the moving average coefficient matrix $C_h(\hat{\Phi})$ can be obtained from the companion form representation of the VAR in (76): $C_h(\Phi) = M'\tilde{\Phi}_1^h M$ with $M' = [I_n, 0_{n\times n(p-1)}]$.

Illustration: we estimate a VAR in output growth, labor share, inflation, and interest rates. Monetary policy shocks are identified according to sign restrictions which leads identified sets for the IRFs. The upper and lower bounds for the identified sets of the IRFs are depicted in Figure 7.

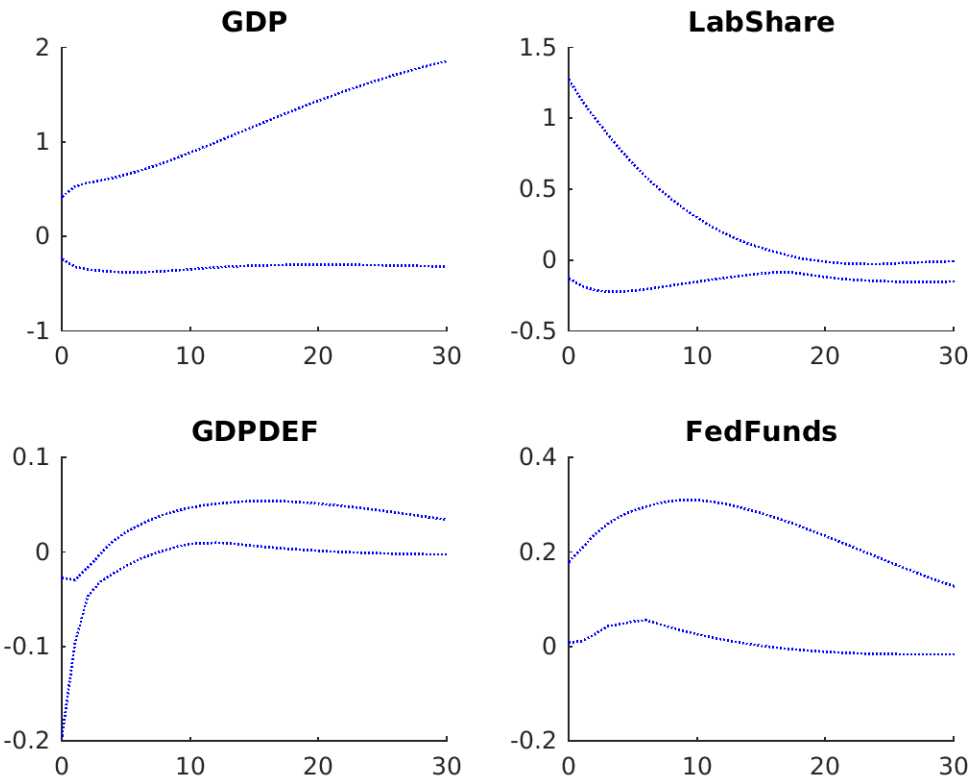
9.3.4 Conditional Moment Restrictions

The unconditional moment restrictions derived from the equilibrium conditions of the DSGE model discussed in Section 9.2.4 have sample analogues in which the population expectations are replaced by sample averages. These moment conditions are typically used to form generalized method of moments (GMM) objective functions. A complication arises if the moment conditions contain latent variables, e.g., the shock process λ_t in the moment condition (68) derived from the New Keynesian Phillips curve.

9.4 Dealing with Trends

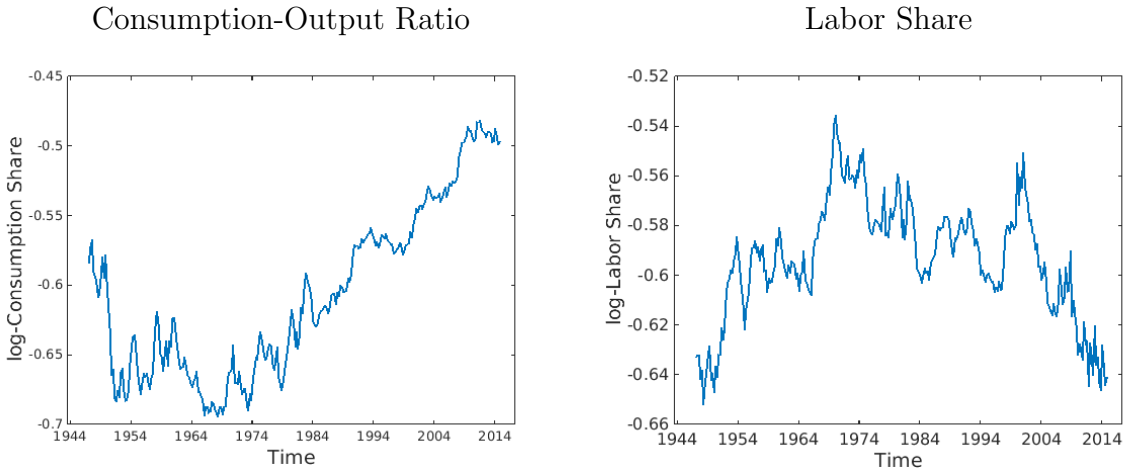
- Macroeconomic time series exhibit trends, which have deterministic and stochastic components.

Figure 7: Impulse Responses to a Monetary Policy Shock



- The simple DSGE model generates a common trend in output and real wages. So far, we used a stationarity inducing transformation by considering output growth and the labor share.
- Most DSGE models have strong co-trending implications, e.g., stationary labor share, which may be contradicted in the data.
- Various remedies:
  - ignore the mismatch;
  - incorporate more elaborate trends into DSGE model
  - detrend each data series separately and fit DSGE model to detrended data
  - detrend both model-implied as well as actual data

Figure 8: Consumption-Output Ratio and Labor Share (in Logs)



– create a hybrid model with a more flexible trend specification: Canova (2014)

Illustration: Figure 8 depicts the log consumption-output ratio and the labor share in logs. Consumption is defined as Personal Consumption Expenditure on Services (PCESV) plus Personal Consumption Expenditure on nondurable goods (PCND) divided by nominal GDP.

9.5 What Next?

Our DSGE model generates a joint probability distribution for the vector of states x_t and the vector of observables y_t. From this probability distribution, we derived very objects of interests: the autocovariance function and the spectral density; impulse response functions; conditional moment restrictions and the marginal distribution of the observables $Y_{1:T}$. We saw in Section 9.3 that these objects have empirical analogues.

Broadly speaking, DSGE model estimation is concerned with determining a range of appropriate values for the parameters θ based on the available data. This is done by defining an objective function $Q(\theta|Y_{1:T})$ that measures the discrepancy between a set of model implication and their empirical counterparts. An example would be the discrepancy between model-implied autocovariances and sample autocovariances. An estimator of θ can be obtained as the extremum (set estimator) or a level set (set estimator) of the objective function $Q(\theta|Y_{1:T})$.

Formal econometric analysis tends to use the probabilistic structure of the DSGE model to determine the objective function $Q(\theta|Y_{1:T})$, to derive measures of uncertainty for θ, and to obtain absolute measures of fit for a specific model as well as relative measures of fit for collections of models. Complications arise if the probabilistic structure of the DSGE model is misspecified. In this case, formal econometric analysis requires a reference model with a well-specified probabilistic structure, from which statistical measures of uncertainty and fit can be derived. Before delving into the details of DSGE model estimation, we discuss an object that is key for statistical inference, namely the likelihood function

10 The Likelihood Function

- An important object for statistical inference is the likelihood function: $p(Y_{1:T}|\theta)$. (sometimes we write $p(Y|\theta)$).
- Thus far, we have characterized the joint distribution of $p(Y_{1:T}, S_{1:T}|\theta)$. In order to obtain the likelihood function we need to integrate out the (hidden) states $S_{1:T}$. We use a filter to do so.
- Factorization:
$$p(Y_{1:T}|\theta) = \prod_{t=1}^{T} p(y_t|Y_{1:t-1}, \theta). \tag{83}$$
- General state-space representation:
$$\begin{aligned} y_t &= \Psi(s_t, t; \theta) + u_t, \quad u_t \sim F_u(\cdot; \theta) \\ s_t &= \Phi(s_{t-1}, \epsilon_t; \theta), \quad \epsilon_t \sim F_\epsilon(\cdot; \theta). \end{aligned} \tag{84}$$
- Due to the first-order Markov structure of the state transition equation neither the states $s_{t-2}, s_{t-3}, \ldots$ nor the observations $y_{t-1}, y_{t-2}, \ldots$ provide any additional information about s_t conditional on s_{t-1}. Thus,
$$p(s_t|s_{t-1}) = p(s_t|s_{t-1}, S_{1:t-2}) = p(s_t|s_{t-1}, S_{1:t-2}, Y_{1:t-1}). \tag{85}$$
- In our linearized DSGE model: The state-transition equation (41) describes a conditional distribution of $s_t|s_{t-1}$, which we generically denote by its density $p(s_t|s_{t-1})$. If

the innovations ϵ_t are normally distributed, then this conditional distribution is given by

$$s_t|s_{t-1} \sim N\big(\Phi_1 s_{t-1}, \Phi_\epsilon \Phi_\epsilon'\big). \tag{86}$$

The conditional distribution of $y_t|s_t$, which we denote by $p(y_t|s_t)$ is simply a point-mass at $y_t = \Psi_0 + \Psi_1 s_t$ in the absence of measurement errors. If the measurement equation (43) is augmented by normally distributed measurement errors, say, $u_t \sim iidN(0, \Sigma_u)$, then $y_t|s_t \sim N\big(\Psi_0 + \Psi_1 s_t, \Sigma_u\big)$.

10.1 A Generic Filter and Smoother

We now describe a generic filter that can be used to recursively compute the conditional distributions $p(y_t|Y_{1:t-1})$ and $p(s_t|Y_{1:t})$. The former is the predictive distribution of y_t given past observations $Y_{1:t-1}$ and appears in the factorization of the likelihood function in (83). The latter summarizes the information about the state s_t conditional on the current and past observations $Y_{1:t}$. The filter is summarized in Algorithm 1.

Algorithm 1 (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:

   $$p(s_t|Y_{1:t}, \theta) = p(s_t|y_t, Y_{1:t-1}) = \frac{p(y_t|s_t, Y_{1:t-1}) p(s_t|Y_{1:t-1})}{p(y_t|Y_{1:t-1})}$$

While Algorithm 1 suffices for real time inference (i.e., based on information $Y_{1:t}$) about the current state s_t, it is not sufficient to conduct *ex post* inference conditional on the full sample $Y_{1:T}$. For *ex post* inference about the latent states, the joint posterior distribution $p(S_{1:T}|Y_{1:T})$ is the relevant object. The joint posterior distribution can be factorized into conditional distributions, starting from $p(s_T|Y_{1:T})$ which is obtained in the last step of Algorithm 1:

$$p(S_{1:T}|Y_{1:T}) = p(s_T|Y_{1:T}) \prod_{t=1}^{T-1} p(s_t|S_{t+1:T}, Y_{1:T}). \tag{87}$$

According to Bayes Theorem

$$p(s_t|S_{t+1}, Y_{1:T}) = \frac{p(s_t, S_{t+1:T}, Y_{1:T})}{\int p(s_t, S_{t+1:T}, Y_{1:T}) ds_t}. \tag{88}$$

The right-hand-side can be factorized as follows:

$$\begin{aligned} p(s_t, S_{t+1:T}, Y_{1:T}) &= \int p(S_{1:T}, Y_{1:T}) dS_{1:t-1} \qquad (89) \\ &= \int p(S_{1:t}, Y_{1:t}) \left(\prod_{j=1}^{T-t} p(s_{t+j}|s_{t+j-1}) p(y_{t+j}|s_{t+j}) \right) dS_{1:t-1} \\ &= p(s_t, Y_{1:t}) p(s_{t+1}|s_t) C(S_{t+1:T}, Y_{t+1:T}) \\ &= p(s_t|Y_{1:t}) p(Y_{1:t}) p(s_{t+1}|s_t) C(S_{t+1:T}, Y_{t+1:T}) \end{aligned}$$

By combining (88) and (89) we can deduce

$$p(s_t|S_{t+1}, Y_{1:T}) = \frac{p(s_t|Y_{1:t}) p(s_{t+1}|s_t)}{\int p(s_t|Y_{1:t}) p(s_{t+1}|s_t) ds_t} = p(s_t|s_{t+1}, Y_{1:t}) \tag{90}$$

because the terms $p(Y_{1:t})$ and $C(S_{t+1:T}, Y_{t+1:T})$ cancel from the numerator and denominator. Using (90) we can simplify the factorization of the joint posterior density of the states considerably:

$$p(S_{1:T}|Y_{1:T}) = p(s_T|Y_{1:T}) \prod_{t=1}^{T-1} \frac{p(s_t|Y_{1:t}) p(s_{t+1}|s_t)}{\int p(s_t|Y_{1:t})}, \tag{91}$$

where $p(s_t|Y_{1:t})$ is generated by Algorithm 1 and $p(s_{t+1}|s_t)$ is obtained from the state-transition equation. The algorithm that generates the sequence of distributions $p(s_t|S_{t+1}, Y_{1:T})$ is called a smoother and summarized in Algorithm 2.

Algorithm 2 (Generic Smoother)**.**

Run Algorithm 1 to obtain $p(s_t|Y_{1:t})$, $t = 1, \ldots, T$. For $t = T-1$ to t, let

$$p(s_t|S_{t+1}, Y_{1:T}) = \frac{p(s_t|Y_{1:t}) p(s_{t+1}|s_t)}{\int p(s_t|Y_{1:t}) p(s_{t+1}|s_t) ds_t}. \tag{92}$$

Table 3: Conditional Distributions for Kalman Filter

	Distribution	Mean and Variance
$s_{t-1}\|Y_{1:t-1}$	$N\big(\bar{s}_{t-1\|t-1}, P_{t-1\|t-1}\big)$	Given from Iteration $t-1$
$s_t\|Y_{1:t-1}$	$N\big(\bar{s}_{t\|t-1}, P_{t\|t-1}\big)$	$\bar{s}_{t\|t-1} = \Phi_1 \bar{s}_{t-1\|t-1}$
		$P_{t\|t-1} = \Phi_1 P_{t-1\|t-1}\Phi_1' + \Phi_\epsilon \Sigma_\epsilon \Phi_\epsilon'$
$y_t\|Y_{1:t-1}$	$N\big(\bar{y}_{t\|t-1}, F_{t\|t-1}\big)$	$\bar{y}_{t\|t-1} = \Psi_0 + \Psi_1 \bar{s}_{t\|t-1}$
		$F_{t\|t-1} = \Psi_1 P_{t\|t-1}\Psi_1' + \Sigma_u$
$s_t\|Y_{1:t}$	$N\big(\bar{s}_{t\|t}, P_{t\|t}\big)$	$\bar{s}_{t\|t} = \bar{s}_{t\|t-1} + P_{t\|t-1}\Psi_1' F_{t\|t-1}^{-1}(y_t - \bar{y}_{t\|t-1})$
		$P_{t\|t} = P_{t\|t-1} - P_{t\|t-1}\Psi_1' F_{t\|t-1}^{-1}\Psi_1 P_{t\|t-1}$
$s_t\|(S_{t+1:T}, Y_{1:T})$	$N\big(\bar{s}_{t\|t+1}, P_{t\|t+1}\big)$	$\bar{s}_{t\|t+1} = \bar{s}_{t\|t} + P_{t\|t}\Phi_1' P_{t+1\|t}^{-1}(s_{t+1} - \Phi_1 \bar{s}_{t\|t})$
		$P_{t\|t+1} = P_{t\|t} - P_{t\|t}\Phi_1' P_{t+1\|t}^{-1}\Phi_1 P_{t\|t}$

10.2 Likelihood Function for a Linearized DSGE Model

For illustrative purposes consider our prototypical DSGE model. Due to the simple structure of the model, we can use (34), (35), (38), and (39) to solve for the latent shocks ϕ_t, λ_t, z_t, and $\epsilon_{R,t}$ as a function of $\hat{x}_t$, $\widehat{lsh}_t$, $\hat{\pi}_t$, and $\hat{R}_t$. Thus, we can deduce from (43) 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.

Now suppose that we drop the labor share and the interest rates from the definition of y_t. In this case it is no longer possible to uniquely determine s_t as a function of y_t and $\hat{x}_0$, because we only have two equations, (34) and (38), and four unknowns. The filter in Algorithm 1 now essentially solves a system of underdetermined 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 1 are Gaussian. In this case the Kalman filter can be used recursively compute the means and covariance matrices of these distributions. To complete the model specification we make the following distributional assumptions about the initial state s_0:

$$s_0 \sim N\big(\bar{s}_{0|0}, P_{0|0}\big).$$

In stationary models it is common to assume that $\bar{s}_{0|0}$ and $P_{0|0}$ corresponds to the invariant distribution associated with the law of motion of s_t in (41). The four conditional distributions in the description of Algorithm 1 for a linear Gaussian state space model are summarized in Table 3. Detailed derivations can be found in textbook treatments of the Kalman filter and smoother, e.g., Hamilton (1994) or Durbin and Koopman (2001). *** Discuss efficient smoothing; singularity of $P_{t|t}$ matrix. ***

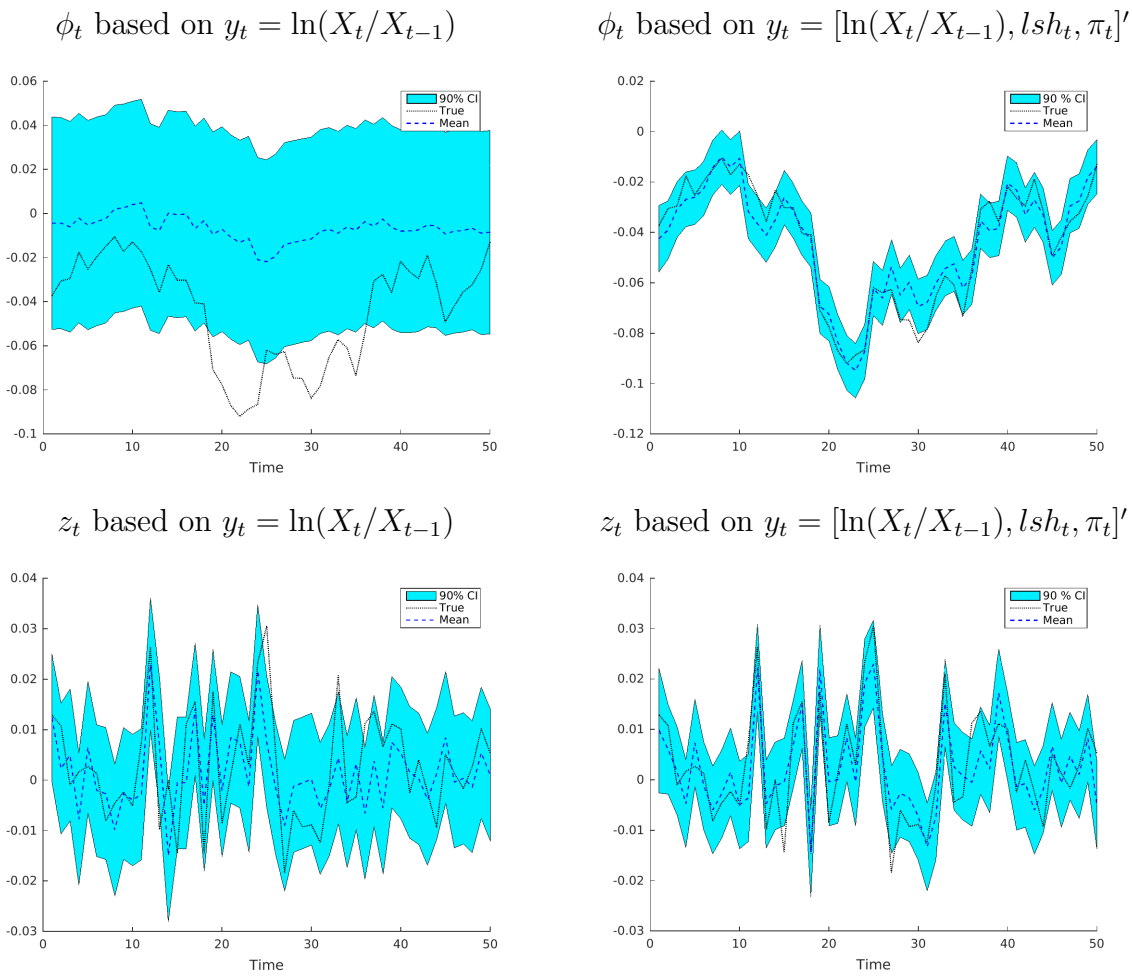
Illustration: We simulate $T = 50$ observations from the DSGE model. Figure 9 depicts the filtered shock processes ϕ_t and z_t based on observations of only output growth and based on observations of output growth, the labor share, and inflation. The figure depicts $\mathbb{E}[s_t|Y_{1:t}]$ as well as 90% credible intervals around the mean based on $\mathbb{V}[s_t|Y_{1:t}]$. Note that the estimates of the latent states become more accurate as the number of observables included in the definition of y_t increases.

10.3 Likelihood Function for Nonlinear DSGE Models

There exists a large literature on particle filters. Surveys and tutorials can be found, for instance, in Arulampalam, Maskell, Gordon, and Clapp (2002), Cappé, Godsill, and Moulines (2007), Doucet and Johansen (2011), Creal (2012). Kantas, Doucet, Singh, Maciejowski, and Chopin (2014) discuss using particle filters in the context of estimating the parameters of a state space models. These papers provide detailed references to the literature. The basic bootstrap particle filtering 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, de Freitas, and Gordon (2001). Textbook treatments of the statistical theory underlying particle filters can be found in Cappé, Moulines, and Ryden (2005), Liu (2001), and Del Moral (2013). First application in DSGE model literature: Fernández-Villaverde and Rubio-Ramírez (2007).

In the basic version of the particle filter the time t particles were generated by simulating the state transition equation forward. 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 $\tilde{s}_t^j$ the actual observation y_t lies far in the tails of the model-implied distribution of $y_t|(\tilde{s}_t^j, \theta)$. The

Figure 9: Filtered States



Notes: The filtered states are based on a simulated sample of $T = 50$ observations.

particle filter can be generalized by allowing $\tilde{s}_t^j$ in the forecasting step to be drawn from a generic importance sampling density $g_t(\cdot|s_{t-1}^j, \theta)$, which leads to the following algorithm:

Algorithm 3 (Generic Particle Filter)**.**

1. **Initialization.** Draw the initial particles from the distribution $s_0^j \overset{iid}{\sim} p(s_0)$ and set $W_0^j = 1$, $j = 1, \ldots, M$.

2. **Recursion.** For $t = 1, \ldots, T$:

   (a) **Forecasting** s_t**.** Draw $\tilde{s}_t^j$ from density $g_t(\tilde{s}_t|s_{t-1}^j, \theta)$ and define the importance weights

   $$\omega_t^j = \frac{p(\tilde{s}_t^j|s_{t-1}^j, \theta)}{g_t(\tilde{s}_t^j|s_{t-1}^j, \theta)}. \tag{93}$$

   An approximation of $\mathbb{E}[h(s_t)|Y_{1:t-1}, \theta]$ is given by

   $$\hat{h}_{t,M} = \frac{1}{M}\sum_{j=1}^{M} h(\tilde{s}_t^j)\omega_t^j W_{t-1}^j. \tag{94}$$

   (b) **Forecasting** y_t**.** Define the incremental weights

   $$\tilde{w}_t^j = p(y_t|\tilde{s}_t^j, \theta)\omega_t^j. \tag{95}$$

   The predictive density $p(y_t|Y_{1:t-1}, \theta)$ can be approximated by

   $$\hat{p}(y_t|Y_{1:t-1}, \theta) = \frac{1}{M}\sum_{j=1}^{M} \tilde{w}_t^j W_{t-1}^j. \tag{96}$$

   (c) **Updating.** Define the normalized weights

   $$\tilde{W}_t^j = \frac{\tilde{w}_t^j W_{t-1}^j}{\frac{1}{M}\sum_{j=1}^{M} \tilde{w}_t^j W_{t-1}^j}. \tag{97}$$

   An approximation of $\mathbb{E}[h(s_t)|Y_{1:t}, \theta]$ is given by

   $$\tilde{h}_{t,M} = \frac{1}{M}\sum_{j=1}^{M} h(\tilde{s}_t^j)\tilde{W}_t^j. \tag{98}$$

(d) **Selection.** *Case (i):* If $\rho_t = 1$ 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 $j =, 1 \ldots, M$.
Case (ii): If $\rho_t = 0$, let $s_t^j = \tilde{s}_t^j$ and $W_t^j = \tilde{W}_t^j$ for $j =, 1 \ldots, M$.
An approximation of $\mathbb{E}[h(s_t)|Y_{1:t}, \theta]$ is given by

$$\bar{h}_{t,M} = \frac{1}{M} \sum_{j=1}^{M} h(s_t^j) W_t^j. \tag{99}$$

3. **Likelihood Approximation.** The approximation of the log likelihood function is given by

$$\ln \hat{p}(Y_{1:T}|\theta) = \sum_{t=1}^{T} \ln \left(\frac{1}{M} \sum_{j=1}^{M} \tilde{w}_t^j W_{t-1}^j \right). \tag{100}$$

- Bootstrap particle filter sets $g_t(\tilde{s}_t^j|s_{t-1}^j, \theta) = p(\tilde{s}_t^j|s_{t-1}^j, \theta)$.
- Evaluation of $p(\tilde{s}_t^j|s_{t-1}^j, \theta)$.
- Conditionally-optimal particle filter sets

$$g_t(\tilde{s}_t|s_{t-1}^j) = p(\tilde{s}_t|y_t, s_{t-1}^j), \tag{101}$$

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

$$\tilde{w}_t^j = \int p(y_t|s_t) p(s_t|s_{t-1}^j) ds_t. \tag{102}$$

In a typical (nonlinear) DSGE model applications it is not possible to sample directly from $p(\tilde{s}_t|y_t, s_{t-1}^j)$.

- Approximately conditionally-optimal particle filter: In a typical DSGE model application, sampling from the conditionally-optimal importance distribution is infeasible or computationally too costly. Alternatively, one could try to sample from an approximately conditionally-optimal importance distribution. For instance, if the DSGE model 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 3 to the first-order approximation of the DSGE model. More generally, as suggested in Guo, Wang, and Chen (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 (2014).

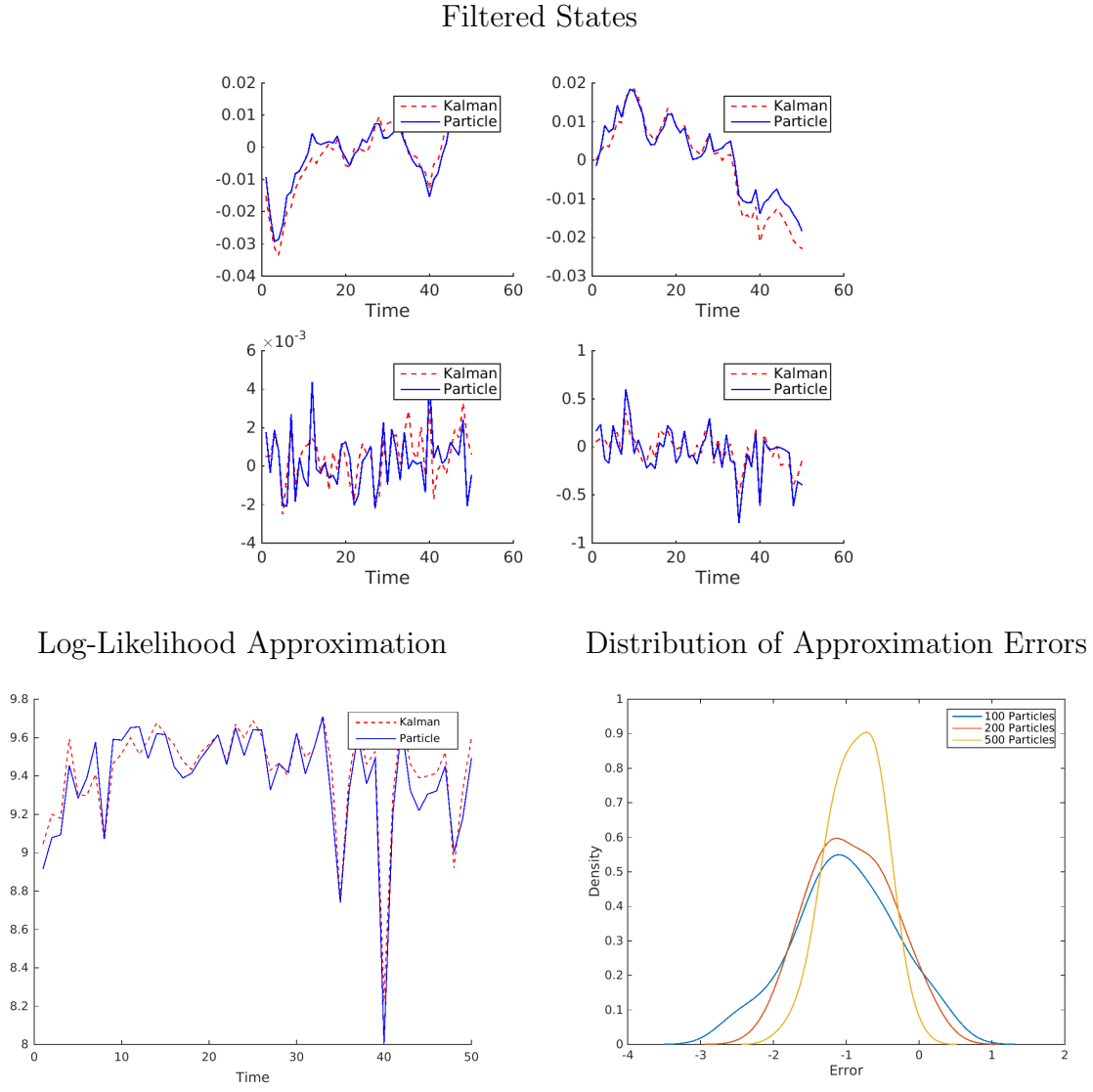
- Conditionally linear models.
- Role of measurement errors u_t.
- Various shortcuts.

Illustration: Show output from particle filter using the same design as for the Kalman filter above. Illustrate numerical accuracy of likelihood approximation as a function of the number of particles.

11 Statistical Inference

- Basic problem: infer DSGE model parameter vector θ from observations Y; provide measure of uncertainty. Once parameter values have been determined: explain the past: what shocks contributed to economic fluctuations; how do these shocks propagate? predict the future: generate DSGE model-based forecasts; policy counterfactuals and welfare analysis.
- We have seen that DSGE model generates family of distributions $p(Y|\theta)$, $\theta \in \Theta$ - though we are free to ignore certain features of this distribution.
- DSGE models have a high degree of theoretical coherence. This means that the functional forms and parameters of equations that describe the behavior of economic agents are tightly restricted by optimality and equilibrium conditions.
- Family of probability distribution (likelihood function) for empirical models with strong degree of theoretical coherence tend to be more restrictive than likelihood functions associated with atheoretical models such as vector autoregressions. A challenge for statistical inference arises if 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.

Figure 10: Particle Filtered States and Log-Likelihood



Notes: The filtered states are based on a simulated sample of $T = 50$ observations. We use the bootstrap particle filter with $M = 100$ particles. (All four observables; some measurement error; $N_{run} = 100$ repetitions to compute the density of the log-likelihood approximation error.)

- Two modes of statistical inference – frequentist and Bayesian:
  - Frequentist inference takes a pre-experimental perspective and focuses on the behavior of estimators and test statistics, which are functions of the observations Y, in repeated sampling under the distribution $\mathbb{P}_\theta^Y$, conditioning on a "true" parameter value θ. Procedures have to be well behaved for all values of $\theta \in \Theta$. Notion of a *data generating process* (DGP).
  - Bayesian inference takes a post-experimental 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 $p(Y|\theta)$.
- In turn, we will provide more details on both modes of inference.

11.1 Identification

- A point θ_0 is globally identified if $p(Y|\tilde{\theta}) = p(Y|\theta_0)$ with probability one implies that $\tilde{\theta} = \theta_0$ for any $\tilde{\theta} \in \Theta$. If the statement is true only for values of $\tilde{\theta}$ in an open neighborhood of θ_0, then θ_0 is locally identified.
- Finite T versus infinite T.
- Focus on entire distribution of the data vs. the first two moments / spectral density.
- The idea of Komunjer and Ng (2011) is to examine the relationship between the coefficients of the state-space representation and the structural DSGE model parameters θ.
- Roughly speaking the idea is to examine the mapping from structural parameters θ into the reduced-form parameters, say, ϕ. Suppose

$$\phi = f(\theta)$$

and

$$\frac{\partial}{\partial \theta'} f(\theta)$$

has full column rank at $\tilde{\theta}$, then the structural parameters are locally identifiable at $\tilde{\theta}$.

- The difficulty with this approach is to define the vector of reduced-form parameters.

- DSGE models are summarized by reduced-form parameters, i.e., state-space representation. Recall that in the linearized DSGE model we have a restricted version:

$$\begin{aligned} y_t &= \Psi_0(\theta) + \Psi_1(\theta) \\ s_t &= \Phi_1(\theta)s_{t-1} + \Phi_\epsilon(\theta)\epsilon_t \end{aligned} \tag{103}$$

  An unrestricted state-space model could be written as

$$\begin{aligned} y_t &= \Psi_0 + \Psi_1 s_t \\ s_t &= \Phi_1 s_{t-1} + \Phi_\epsilon \epsilon_t, \end{aligned} \tag{104}$$

  where

$$\phi' = [vec(\Psi_0)', vec(\Psi_1)', vec(\Phi_1)', vec(\Phi_\epsilon)']$$

  The problem is that the matrices Ψ_1, Φ_1, and Φ_ϵ themselves are not identifiable. For instance, 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

$$\tilde{s}_t = As_t, \quad \tilde{\epsilon}_t = \Omega\epsilon_t, \quad \tilde{\Psi}_1 = \Psi_1 A^{-1}, \quad \tilde{\Phi}_1 = \Phi_1 A^{-1}, \quad \tilde{\Phi}_\epsilon = A\Phi_\epsilon\Omega'$$

  and write

$$\begin{aligned} y_t &= \Psi_0 + \tilde{\Psi}_1 \tilde{s}_t \\ s_t &= \tilde{\Phi}_1 \tilde{s}_{t-1} + \tilde{\Phi}_\epsilon \epsilon_t \end{aligned} \tag{105}$$

  Thus, the number of identifiable reduced-form parameters is smaller than the number of elements in the Ψ and Φ matrices.

- Example from Schorfheide (2013):

$$y_t = [1\ 1]s_t, \quad s_t = \begin{bmatrix} \theta_1^2 & 0 \\ 1-\theta_1^2-\theta_1\theta_2 & (1-\theta_1^2) \end{bmatrix} s_{t-1} + \begin{bmatrix} 1 \\ 0 \end{bmatrix} \epsilon_t, \quad \epsilon_t \sim iidN(0,1). \tag{106}$$

  How many identifiable reduced-form parameters are there? Letting L denote the lag operator with the property that $Ly_t = y_{t-1}$, we can write the law of motion of y_t as an restricted ARMA(2,1) process:

$$\big(1-\theta_1^2 L\big)\big(1-(1-\theta_1^2)L\big)y_t = \big(1-\theta_1\theta_2 L\big)\epsilon_t. \tag{107}$$

This suggests that there are three. However, if one of the roots of the AR polynomial is equal to the root of the MA polynomial, the process reduces to an AR(1) process, which has only one identifiable reduced-form parameter. In this example, θ_2 is not identifiable if θ_1 is 0, because it enters the model only multiplicatively. Moreover, given θ_1 and θ_2, we obtain an observationally equivalent process by choosing $\tilde{\theta}_1$ and $\tilde{\theta}_2$ such that

$$\tilde{\theta}_1 = \sqrt{1-\theta_1^2}, \quad \tilde{\theta}_2 = \theta_1\theta_2/\tilde{\theta}_1.$$

Here we switched the values of the two roots of the autoregressive lag polynomial.

- Alternatively, examine the Jacobian associated with the mapping between structural parameters and the mean and covariance matrix of $Y_{1:T}$ (see Iskrev (2010));

- or the mapping between θ and the spectral density (see Qu and Tkachenko (2012)), which requires

$$G(\theta) = \int_{-\pi}^{\pi} \left(\frac{\partial}{\partial\theta'} vec(f_{yy}(\omega)')\right)' \left(\frac{\partial}{\partial\theta'} f_{yy}(\omega)\right) d\omega \tag{108}$$

  to be of full rank.

- Some numerical issues: accuracy of derivatives and of matrix rank.

- As we remove information, identification gets more problematic. This can be seen from the example model: suppose we only observe the labor share. We can write the law of motion of the labor share as ARMA(3,3) which has 8 reduced form parameters including the mean. This is fewer than the 13 DSGE model parameters. Also, removing information from the shocks eliminates information. Suppose we only focus on the response to the monetary policy shock. Then we can at most identify β, ζ, ν, and σ_R. See discussion by Canova and Sala (2009)

- Some recent work on global identification: Qu and Tkachenko (2014), Kociecki and Kolasa (2015)

11.2 Frequentist Inference

- Some notation: we have to distinguish between the DSGE models, say, $M_1, \ldots, M_J$, and a reference model M_0. This reference M_0 model could by a VAR, a general linear

process, or some nonlinear statistical model. For now, we focus on a single DSGE model M_1 versus a reference model M_0.

- We also use the notation $\|a\|_W = a'Wa$.

11.2.1 "Correct" Specification of DSGE Model

- "Correct" specification acknowledges that the DSGE model is an abstraction but regards the family of distributions $p(Y|\theta)$ to be rich enough to capture the salient features of the observables Y.
- Identification is required as regularity condition for "standard" inference procedures.
- DSGE model itself is the data generating process and $p(Y|\theta, M_1)$ describes the sampling distribution of Y under which the behavior of estimators and test statistics is being analyzed.
- Example maximum likelihood estimator:

$$\hat{\theta}_{mle} = \text{argmax}_{\theta\in\Theta} \ \ln p(Y|\theta, M_1) \tag{109}$$

  Under suitable regularity conditions (identification and stationarity of $\{y_t\}$ process), the MLE is consistent and asymptotically normal. Inference can be based on Wald, LM, or LR tests. Confidence sets can be obtained by inverting test statistics.

- The likelihood function can be replaced by the objective function of an estimator that matches sample statistics to model-implied population analogues:
  - sample statistic $\hat{m}_T(Y)$;
  - model-implied population statistic $\mathbb{E}[\hat{m}_T(Y)|\theta, M_1]$
  - simulation-based approximation of of model-implied population statistic: let $Y^*(\theta, M_1)$ be a sequence of λT simulated observations from DSGE model M_1 given parameter θ. Then we use $\hat{m}_{\lambda T}(Y^*(\theta, M_1))$ to denote the sample statistic computed from the simulated observations.

– Example (i): $\hat{m}_T(Y)$ could be composed of sample autocovariance matrix. E.g,

$$\hat{m}_T(Y) = \left[vech(\hat{\Gamma}_{yy}(0))', vec(\hat{\Gamma}_{yy}(1))'\right] = \frac{1}{T}\sum_{t=1}^{T} m(y_{t-1:t}).$$

Example (ii): $\hat{m}_T(Y)$ could be the OLS estimator of the coefficients of a VAR(1). For simplicity, assume that the data are demeaned and the VAR(1) does not contain an intercept, then

$$\hat{m}_T(Y) = vec\left(\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_{t-1}'\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_t'\right).$$

- If the moments can be evaluated analytically, we can define the objective function as

$$Q_T(\theta|Y) = \left\|\hat{m}_T(Y) - \mathbb{E}[\hat{m}_T(Y)|\theta, M_1]\right\|_{W_T}, \tag{110}$$

where W_T is a symmetric positive-definite weight matrix. If the moments have to be evaluated by simulation, we can use the objective function

$$Q_T(\theta|Y) = \left\|\hat{m}_T(Y) - \hat{m}_T\big(Y^*(\theta, M_1)\big)\right\|_{W_T}, \tag{111}$$

- Finally, let

$$\hat{\theta}_{mm} = \text{argmax}_{\theta\in\Theta}\, Q_T(\theta|Y). \tag{112}$$

- Under correct specification it makes sense to let the model-implied probability distribution of the data determine the choice of the objective function for estimators and test statistics to increase efficiency.

11.2.2 Incompleteness and Misspecification of DSGE Model

- Incompleteness: Suppose we remove all but the monetary policy shock from the DSGE model. In this case, the DSGE model is incomplete because it does not contain sufficiently many shocks to explain the variability in the observed data y_t. The DSGE model generates a counterfactual singular distribution. Another example of incompleteness would be if we only consider a subset of the moment conditions implied by the model, e.g., the moment condition implied by the Euler equation. This does not completely specify the distribution of Y and the data generating process.

- Misspecification: the DSGE model takes the form of a state-space model. The coefficient matrices Φ_1, Φ_ϵ, Ψ_0, and Ψ_1. are restricted functions of the DSGE model parameters θ. These functions could be misspecified. Note that an omitted state variable can also be interpreted as a misspecification of these functions.

- In either case we need a data generating process from which we can derive the sampling distribution of Y:

  1. the data generating process could be a fully-specified reference model such as a VAR, $p(Y|\phi, M_0)$, where ϕ is a finite-dimensional parameter vector;
  2. the data generating process could be a general process for $\{y_t\}$ that restricts the moments, degree of serial correlation, to ensure that certain sample moments are convergent;
  3. under incompleteness, the data generating process could be obtained by completing the model, e.g., by adding additional shocks to a version of the DSGE model with only a monetary policy shock, or by adding additional equilibrium conditions to an Euler equation.
  4. We write $p(Y|M_0)$.

- The concept of "true" values is elusive, but we can distinguish the following cases:

  1. Model is incompletely specified: it makes sense to refer to "true" values, in the sense that we can imagine the DGP to be the incomplete model plus a set of equations (with additional parameters) that complete the distribution of Y.
  2. Model is misspecified: define pseudo-true (or pseudo-optimal) values, e.g., based on likelihood / Kullback-Leibler distance

  $$\theta_0(KL) = \text{argmax}_{\theta\in\Theta} \int p(Y|M_0) \ln p(Y|\theta, M_1) dY \tag{113}$$

  or based on a population analogue of the moment discrepancy function

  $$\theta_0(Q, W) = \text{argmin}_{\theta\in\Theta}\, Q(\theta|M_0), \tag{114}$$

  where

  $$Q(\theta|M_0) = \left\| \mathbb{E}[\hat{m}_T(Y)|M_0] - \mathbb{E}[\hat{m}(Y)|\theta, M_1] \right\|_W$$

  Note that $\theta_0(\cdot)$ is loss function (or discrepancy) dependent.

- Under misspecification, the probability limit $\theta_0(\cdot)$ depends on the loss function / estimation objective function. In particular, in case of the moment-based objective function it depends on the selection of moments in the vector $\hat{m}_T(Y)$ and it depends on the weight matrix W. Thus, under misspecification, the choice of W should reflect the relevant loss function rather than be determined by the covariance of the sample moments $\hat{m}_T(Y)$.

- The sampling properties of estimators and test statistics have to be derived from the reference model M_0.

11.3 Bayesian Inference

11.3.1 "Correct" Specification of DSGE Model

- Under the Bayesian paradigm, the calculus of probability is not only used 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, i.e., 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:

$$p(\theta|Y, M_1) = \frac{p(Y|\theta, M_1)p(\theta|M_1)}{p(Y|M_1)}, \quad p(Y|M_1) = \int p(Y|\theta, M_1)p(\theta|M_1)d\theta. \tag{115}$$

- The posterior distribution contains all the information about θ conditional on sample information Y.

- Some subtle terminology: under the Bayesian paradigm a model is a joint distribution for data and parameters, i.e. likelihood function $p(Y|\theta, M_1)$ and prior $p(\theta|M_1)$.

- The posterior distribution of transformations of θ, say $h(\theta)$, e.g., autocovariances, IRFs, etc., can be derived from $p(\theta)$.

- A word on identification: for non-identifiable parameter the conditional prior distribution equals the conditional posterior distribution; lack of identification may generate difficulties for posterior simulators.

- Use decision-theoretic setup to derive point estimators, interval estimators, predictions, etc. for $h(\theta)$. The key idea is to specify a loss function $L\big(h(\theta),\delta\big)$ and to find the optimal decision δ_* that minimizes the posterior expected loss:

$$\delta_* = \text{argmin}_{\delta\in\mathcal{D}} \int L\big(h(\theta),\delta\big)p(\theta|Y)d\theta. \tag{116}$$

- Practical difficulty: characterization of moments of $p(h(\theta)|Y,M_1)$. In practice, we typically use Monte Carlo integration:

$$\mathbb{E}[h(\theta)|Y] = \int h(\theta)p(\theta|Y)d\theta \approx \frac{1}{N}\sum_{i=1}^{N} h(\theta^i), \tag{117}$$

  where the θ^i's are distributed according to $p(\theta|Y)$. We will subsequently discuss different strategies of generating the θ^i draws.

- Modern Bayesian computational techniques: MCMC and sequential Monte Carlo sampling.

- The relative fit of two DSGE models M_1 and M_2 can be summarized through posterior model probabilities. Let $\pi_{j,0}$ be prior probability of model $j = 1, 2$. Call ratios of probabilities odds. Then the posterior odds of M_1 versus M_2 are given by

$$\frac{\pi_{1,T}}{\pi_{2,T}} = \frac{\pi_{1,0}}{\pi_{2,0}}\frac{p(Y|M_1)}{p(Y|M_2)}, \tag{118}$$

  where the first factor on the r.h.s 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 (115). Posterior odds and probabilities can be used for model selection and for model averaging. Applications: Rabanal and Rubio-Ramírez (2005), Smets and Wouters (2007).

11.3.2 Misspecification of DSGE Model

- Researcher believes that there is some other more densely parameterized model, e.g., a VAR, that could have (or did) generate the data.

- Schorfheide (2000): loss function based evaluation of DSGE models. Researcher is interested in the relative ability of two (or more) DSGE models to explain certain population characteristics φ, e.g., autocovariances or impulse responses. However, the DSGE models may be potentially misspecified and the researcher considers a reference model M_0. Overall posterior:

$$p(\varphi|Y) = \sum_{j=0,1,2} \pi_{j,T} p(\varphi|Y, M_j). \tag{119}$$

  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. Compute DSGE model specific predictions:

$$\hat{\varphi}_{(j)} = \text{argmin}_{\tilde{\varphi}} \int L(\tilde{\varphi}, \varphi) p(\varphi|Y, M_j) d\varphi, \quad j = 1, 2 \tag{120}$$

  Compare models based on

$$\int L(\hat{\varphi}_{(j)}, \varphi) p(\varphi|Y) d\varphi. \tag{121}$$

- Geweke (2010): incomplete econometric models; models of moments (also DeJong, Ingram, and Whiteman (1996)). The idea is that the DSGE model is not meant to be a model for the sample of observations of Y, but only a model for certain population moments φ. He shows that under this assumption one can define model odds as

$$\frac{\pi_{1,T}}{\pi_{2,T}} = \frac{\pi_{1,0}}{\pi_{2,0}} \frac{\int p(\varphi|M_1) p(\varphi|Y, M_0) d\varphi}{\int p(\varphi|M_2) p(\varphi|Y, M_0) d\varphi} \tag{122}$$

  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.

- Ingram and Whiteman (1994) and Del Negro and Schorfheide (2004): using DSGE models to construct prior distributions for reference models such as VARs. Consider companion form VAR in (76). Use the DSGE model to generate a prior distribution for $(\tilde{\Phi}_1, \tilde{\Phi}_0, \tilde{\Sigma})$ and combine this prior with VAR likelihood function

$$p(Y, \tilde{\Phi}_0, \tilde{\Phi}_1, \tilde{\Sigma}, \theta|\lambda) = p(Y|\tilde{\Phi}_0, \tilde{\Phi}_1, \tilde{\Sigma}) p(\tilde{\Phi}_0, \tilde{\Phi}_1, \tilde{\Sigma}|\theta, \lambda) p(\theta) \tag{123}$$

Define restriction functions, e.g., by population regression,

$$\tilde{\Phi}_0^*(\theta), \quad \tilde{\Phi}_1^*(\theta), \quad \tilde{\Sigma}^*(\theta)$$

and center the prior $p(\tilde{\Phi}_0, \tilde{\Phi}_1, \tilde{\Sigma}|\theta, \lambda)$ on this restriction function, allowing for deviations (misspecification). The parameter λ is a hyperparameter that controls the magnitude of the deviations (prior variance) from the restriction function. Application to forecasting performance of Smets-Wouters model: Del Negro, Schorfheide, Smets, and Wouters (2007). Application to policy analysis: Del Negro and Schorfheide (2009).

- Interpretation of posterior odds if all models are misspecified: find the model that is closest to M_0 in a Kullback-Leibler sense. Fernandez-Villaverde and Rubio-Ramirez (2004). 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:

$$\lambda p(y_t|Y_{1:t-1}, M_1) + (1-\lambda)p(y_t|Y_{1:t-1}, M_2).$$

The weight λ can be determined based on

$$\prod_{t=1}^{T} \lambda p(y_t|Y_{1:t-1}, M_1) + (1-\lambda)p(y_t|Y_{1:t-1}, M_2),$$

see Geweke and Amisano (2011) and Geweke and Amisano (2012). Dynamic version with λ depending on time t are provided by Waggoner and Zha (2012) and Del Negro, Hasegawa, and Schorfheide (2014).

12 Frequentist Estimation Techniques

Discuss both estimation and basic model evaluation. Each estimation procedure comes with an evaluation procedure. How well did we do achieving the estimation objective?

12.1 Likelihood-Based Estimation

- Makes sense if model is well specified

- Early papers that estimate versions of the neoclassical stochastic growth model using MLE are Altug (1989), McGrattan (1994), and Leeper and Sims (1995). Many papers followed.

- Obstacle 1: stochastic singularity. Imagine we remove all but the technology shock from our simple DSGE model. In this case, we have one shock and four observables. It is easy to see from the DSGE model solution that the DSGE model places probability one on the event that

$$\beta \ln R_t - \ln \pi_t = \beta \ln(\pi^* \gamma/\beta) - \ln \pi^*.$$

  Because for actual data $\beta \ln R_t - \ln \pi_t$ the likelihood function is equal to zero and not usable for inference. To overcome this obstacle, researchers either have to add "measurement" errors or restrict the number of observables in the vector of y_t to the number of exogenous shocks ϵ_t. Choosing variables to include in the likelihood function: Canova, Ferroni, and Matthes (2014)

- Obstacle 2: identification issues led researcher to fix a subset of the parameters during the maximization of the likelihood function.

- Evaluation of the likelihood function: see Section 10

- Maximization of the likelihood function: various numerical techniques are available *** provide a good reference ***. Key distinction: some methods are gradient-based, e.g., Newton-Raphson and other methods do not rely on gradient, e.g., simulated annealing. If the likelihood function is generated by particle filter, then it is non-differentiable. In the resampling step particles are resampled based on a discontinuous cumulative distribution function. A small change in the parameter θ will cause a small change in the importance weights, which will potentially lead to a different set of resampled particles. As a result the particle approximation of the likelihood function is discontinuous even if the true likelihood function is not. See survey of Kantas, Doucet, Singh, Maciejowski, and Chopin (2014) and the solution proposed by Malik and Pitt (2011).

- Classical machinery: for a formal analysis of maximum likelihood estimators in state-space models, see, for instance, Cappé, Moulines, and Ryden (2005).

– Log-likelihood function $\ell_T(\theta|Y) = \ln p(Y|\theta)$. Recall that

$$\ell_T(\theta|Y) = \sum_{t=1}^{T} \ln p(y_t|Y_{1:t-1}) = \sum_{t=1}^{T} \ln \int p(y_t|s_t,\theta)p(s_t|Y_{1:t-1})ds_t. \tag{124}$$

Note that the summands are not stationary. Under the assumption that the sequence $\{s_t, y_t\}$ is stationary if initialized in the infinite past, we can approximate the log likelihood function by

$$\ell_T^s(\theta|Y) = \sum_{t=1}^{T} \ln \int p(y_t|s_t,\theta)p(s_t|Y_{-\infty:t-1})ds_t, \tag{125}$$

such that

$$\left|\ell_T(\theta|Y) - \ell_T^s(\theta|Y)\right| = \text{small}.$$

– Consistency requires: $T^{-1}\ell_T^s(\theta|Y) \longrightarrow \ell^s(\theta)$ uniformly almost surely, where $l^s(\theta)$ is deterministic and minimized at the "true" θ_0.

– Frequentist asymptotics rely on second-order approximation of log-likelihood function. Define score (vector of first derivatives) $\nabla_\theta \ell_T^s(\theta|Y)$ and observed information matrix $-\nabla_\theta^2 \ell_T^s(\theta|Y)$ and let

$$\begin{aligned}\ell_T^s(\theta|Y) &= \ell_T^s(\theta_0|Y) + T^{-1/2}\nabla_\theta \ell_T^s(\theta_0|Y)\sqrt{T}(\theta-\theta_0) \\ &\quad + \frac{1}{2}\sqrt{T}(\theta-\theta_0)'\left[\nabla_\theta^2 \ell_T^s(\theta_0|Y)\right]\sqrt{T}(\theta-\theta_0) + \text{small}\end{aligned}$$

– If the maximum is attained in the interior, the first-order conditions can be approximated by

$$\sqrt{T}(\theta-\theta_0) = \left[-\nabla_\theta^2 \ell_T^s(\theta_0|Y)\right]^{-1} T^{-1/2}\nabla_\theta \ell_T^s(\theta_0|Y). \tag{126}$$

– The score process satisfies a CLT: $T^{-1/2}\nabla_\theta \ell_T(\theta|Y) \Longrightarrow N(0, \mathcal{I}(\theta_0))$, where $\mathcal{I}(\theta_0)$ is the Fisher information matrix. The formal definition of Fisher information is delicate for this model and therefore omitted.

– $\| -\nabla_\theta^2 \ell_T(\theta|Y) - \mathcal{I}(\theta_0)\|$ converges to zero uniformly in a neighborhood around θ_0.

– Overall, this leads to

$$\sqrt{T}(\hat{\theta}_{mle} - \theta_0) \Longrightarrow N\big(0, \mathcal{I}^{-1}(\theta_0)\big). \tag{127}$$

- In frequency domain we can use Whittle likelihood. Recall DSGE model implied spectral density $f_{yy}(\omega|\theta)$ and sample periodogram $\hat{f}_{yy}(\omega)$. In our linearized DSGE model with Gaussian innovations we can directly characterize the joint distribution of the observables (rather than solving a filtering problem). Let Y by a $T \times n$ matrix composed of rows y_t'. Then the joint distribution of Y is given by

$$Y|\theta \sim N\left(I \otimes \Phi_0(\theta), \begin{bmatrix} \Gamma_{yy}(0|\theta) & \Gamma_{yy}(1|\theta) & \dots & \Gamma_{yy}(T-1|\theta) \\ \Gamma_{yy}'(1|\theta) & \Gamma_{yy}(0|\theta) & \dots & \Gamma_{yy}(T-2|\theta) \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{yy}'(T-1|\theta) & \Gamma_{yy}'(T-2|\theta) & \dots & \Gamma_{yy}(0|\theta) \end{bmatrix}\right) \tag{128}$$

  For large T the joint density can be approximated by the so-called Whittle likelihood function

$$p_W(Y|\theta) \propto \left(\prod_{j=0}^{T-1} \left|2\pi f_{yy}^{-1}(\omega|\theta)\right|\right)^{1/2} \exp\left\{-\frac{1}{2}\sum_{j=0}^{T-1} tr\left[f_{yy}^{-1}(\omega_j|\theta)\hat{f}_y y(\omega_j)\right]\right\} \tag{129}$$

  where 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. Applications: Christiano and Vigfusson (2003) and Sala (2015).

- Identification-robust inference:

  – Guerron-Quintana, Inoue, and Kilian (2013): the key idea is to obtain direct 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 of the system matrices discussed above, they might have to be re-parameterized in terms of an identifiable reduced form parameter $\phi = f(\theta)$. See Table 1 for possible definition of ϕ. Note that many elements of $\Phi_1(\theta)$ and $\Phi_\epsilon(\theta)$ do not depend on θ. One can then form a likelihood ratio test of $H_0: \theta = \theta_0$ based on

$$LR = 2\left[\ln p(Y|\hat{\phi}, M_1^\phi) - \ln p(Y|f(\theta_0), M_1^\phi)\right] \Longrightarrow \chi^2_{\dim(\phi)}$$

    Here M_1^ϕ refers to the state-space representation of DSGE model, parameterized in terms of the identifiable reduced-form parameter vector ϕ. Degrees of freedom depend on dimension of ϕ, which means that it is important to use a minimal

state-variable representation of the DSGE model solution. Paper discusses implementation based on the output from a Bayesian estimation. Invert LR statistic to generate confidence intervals. Subsector inference is possible by projection of the joint confidence set. Strongly identified parameters could be concentrated out.

– Andrews and Mikusheva (2015): Lagrange multiplier test which is also identification robust. Notation: we have the score $\nabla_\theta \ell_T(\theta|Y)$ and the quadratic variation of the score process. Define

$$s_{T,t}(\theta) = \nabla_\theta \ell(\theta|Y_{1:t}) - \nabla_\theta \ell(\theta|Y_{1:t-1}), \quad J_T(\theta) = \sum_{t=1}^{T} s_{T,t}(\theta) s'_{T,t}(\theta).$$

The Lagrange multiplier statistic can be defined as

$$LM = \nabla'_\theta \ell_T(\theta_0|Y)[J_T(\theta_0)]^{-1} \nabla_\theta \ell_T(\theta_0|Y) \Longrightarrow \chi^2_{\dim(\theta_0)}$$

– Frequency domain version of LM test is provided by Qu (2014).

12.2 (Simulated) Method of Moments Estimation and Indirect Inference

- Recall method of moments estimator, define by the objective function $Q(\theta|Y)$ in (110) and (111). Examples for $\hat{m}_T(Y)$: a subset of the autocovariances $\Gamma_{yy}(h)$, estimates of the parameters of an approximating model, e.g., the VAR(p) in (75) as in Smith (1993). If $\hat{m}_T(Y)$ consists of parameter estimates of a reference model (or approximating model), then the procedure is also called indirect inference.

- For some DSGE models and moment functions the model-implied moments can be evaluated directly, e.g., the autocovariance function of linearized DSGE models, but, in particular for nonlinear DSGE models, $\mathbb{E}[\hat{m}_T(Y)|\theta, M_1]$ is often replaced by the simulation approximation $\hat{m}_{\lambda T}(Y^*(\theta, M_1))$, where Y^* is a sample of size λT simulated from DSGE model M_1 given the parameter value θ.

- Some sample statistics, e.g., autocovariances, can be expressed as sample averages, e.g.,

$$\hat{m}_T(Y) = \frac{1}{T} \sum_{t=1}^{T} m(y_{t-p:1}) \tag{130}$$

in which case (assuming stationarity)

$$\mathbb{E}[\hat{m}_T(Y)|\theta, M_1] = \mathbb{E}[m(y_{t-p:t})|\theta, M_1], \tag{131}$$

which can be calculated from the model. On the other hand, if $\hat{m}_T(Y)$ corresponds to the OLS estimator of a VAR(1) without intercept then,

$$\mathbb{E}[\hat{m}_T(Y)] = \mathbb{E}\left[\left(\frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_{t-1}'\right)^{-1}\frac{1}{T}\sum_{t=1}^{T} y_{t-1}y_t'\middle|\theta, M_1\right], \tag{132}$$

which is difficult to evaluate even for a linearized DSGE model. In this case, it could be replaced by the probability limit of $\hat{m}_{\lambda T}(Y^*(\theta, M_1))$, which is given by the population regression

$$\left(\mathbb{E}[y_{t-1}y_{t-1}'|\theta, M_1]\right)^{-1}\mathbb{E}[y_{t-1}y_t'|\theta, M_1]. \tag{133}$$

- Conceptual issue: models with stochastic singularities. If the DSGE model is incomplete, in the sense that it does not contain a full set of shocks, then comparing the autocovariance from the DSGE model to the sample autocovariance does not make sense. Also, in this case the sampling distribution of $\hat{m}_T(Y)$ should not be derived from DSGE model.

- Notation:

$$\begin{aligned} G_T(\theta|Y) &= \hat{m}_T(Y) - \mathbb{E}[\hat{m}_T(Y)|\theta, M_1] \\ Q_T(\theta|Y) &= \left\|G_T(\theta|Y)\right\|_W \\ D(\theta) &= \nabla_\theta \mathbb{E}[m(y_{t-p:t})'|\theta, M_1] \\ \Omega(\theta) &= \lim_{t=1}^{T} \mathbb{V}\left[\sqrt{T}G_T(\theta|Y)\right] \end{aligned}$$

  We also have the "score" $\nabla_\theta Q_T(\theta|Y)$ and the "hessian" $\nabla^2_\theta Q_T(\theta|Y)$.

- Note that under suitable regularity conditions $Q_T(\theta|Y) \xrightarrow{a.s.} Q(\theta)$, where $Q(\theta) \geq 0$ is uniquely minimized at $\theta = \theta_0$.

- Extremum estimator asymptotics rely on second-order approximation of objective function:

$$\begin{aligned} TQ_T(\theta|Y) &= \sqrt{T}\nabla_\theta Q_T(\theta_0|Y)\sqrt{T}(\theta-\theta_0)' \\ &\quad + \frac{1}{2}\sqrt{T}(\theta-\theta_0)'\left[\frac{1}{T}\nabla^2_\theta Q_T(\theta_0|Y)\right]\sqrt{T}(\theta-\theta_0) + \text{small}. \end{aligned} \tag{134}$$

- If the minimum is obtained in the interior, then

$$\sqrt{T}(\hat{\theta}_{mm} - \theta_0) = \left[-\frac{1}{T}\nabla_\theta^2 Q_T(\theta_0|Y)\right]^{-1} \sqrt{T}\nabla_\theta Q_T(\theta_0|Y) + \text{small}. \tag{135}$$

- The "score" for the quadratic objective function is given by

$$\sqrt{T}\nabla_\theta Q_T(\theta_0|Y) = \nabla_\theta G_T(\theta_0|Y) W \sqrt{T} G_T(\theta_0|Y) \tag{136}$$

  and its distribution depends on the distribution of

$$\begin{aligned}\sqrt{T}G(\theta_0|Y) &= \sqrt{T}\big(\hat{m}_T(Y) - \mathbb{E}[\hat{m}_T(Y)|\theta_0, M_1]\big) \\ &\quad +\sqrt{T}\big(\hat{\mathbb{E}}[\hat{m}_T(Y)|\theta_0, M_1] - \mathbb{E}[\hat{m}_T(Y)|\theta_0, M_1]\big) \\ &= I + II,\end{aligned}$$

  say. Term II drops out if the moments for the estimation objective function are computed analytically.

- Under suitable regularity conditions

$$\sqrt{T}G_T(\theta_0|Y) \Longrightarrow N\big(0, \Omega\big) \tag{137}$$

  and we now consider the asymptotic covariance matrix

$$\Omega = \mathbb{V}_\infty[I] + \mathbb{V}_\infty[II]. \tag{138}$$

- The distribution of term II always depends on the model-implied distribution of the moments, regardless of whether the model is correctly specified or not. Its asymptotic variance is given by

$$\mathbb{V}_\infty[II] = \lim_{T\longrightarrow\infty} T\mathbb{V}\left[\hat{m}_{\lambda T}(Y^*(\theta_0, M_1))\right] \tag{139}$$

- We can write the limit variance of term I as

$$\mathbb{V}_\infty[I] = \lim_{T\longrightarrow\infty} T\mathbb{V}[\hat{m}_T(Y)] \tag{140}$$

  We can derive a formula for $\mathbb{V}_\infty[I]$ under the assumption that the model is correct as well as under the assumption that the model is incorrect, ie., derive the limit distribution of $\hat{m}_T(Y)$ under M_0 or M_1.

- Under the assumption that the model is "correctly" specified

$$\Omega = (1 + 1/\lambda) \lim_{T \longrightarrow \infty} T\mathbb{V}\big[\hat{m}_T(Y)|\theta_0, M_1\big].$$

- Ω is a long-run covariance matrix if $\hat{m}_T(Y)$ is a sample average.

- Limit distribution of methods of moments estimator:

$$\sqrt{T}\big(\hat{\theta}_{mm} - \theta_0\big) \Longrightarrow N\big(0, (DWD')^{-1}DW\Omega WD'(DWD')^{-1}\big) \tag{141}$$

- Under "correct" specification, optimal weight matrix is given by

$$W^{-1} = \lim_{T \longrightarrow \infty} T\mathbb{V}\big[\hat{m}_T(Y)\big].$$

- Early work on the asymptotics of simulation-based extremum estimators: Pakes and Pollard (1989). DSGE model applications in Lee and Ingram (1991) and Smith (1993).

- Some performance comparisons in Ruge-Murcia (2007) and Ruge-Murcia (2012).

- Testing hypotheses about coefficients and construction of confidence intervals based on quasi-likelihood (ratio, LM, Wald) statistics.

- Testing overidentifying restrictions based on $TQ_T(\hat{\theta}_{mm}|Y) \longrightarrow \chi^2$ under the optimal weight matrix. Degrees of freedom depend on the number of over-identifying restrictions.

- Issues arising from stationarity inducing transformations: Gorodnichenko and Ng (2010)

- Indirect inference for misspecified DSGE models: Dridi, Guay, and Renault (2007)

- Explict formulae for moments of pruned models solved with perturbation methods: Andreasen, Fernández-Villaverde, and Rubio-Ramírez (2013).

Illustration: use different choices of $m(\cdot)$ function, some of which are based on simulation approximations. Apply repeatedly and show how sampling variation of estimator changes.

12.3 Impulse Response Function Matching

- Sometimes, DSGE models are specified only with a few exogenous shocks that are unable to explain all the variability in the data, e.g., only a technology shock or only a monetary policy shock. To compare models and data we have to purge the effects of the unspecified shocks from the data. This can be done by "filtering" the data through a VAR and estimating the impulse response function to a particular structural shock. One then can estimate the parameters based on discrepancy between model-implied and empirical impulse response function. Any mismatch between the impulse responses provides some indication about the misspecification of the propagation mechanism to the considered structural shock.

- Influential empirical papers utilizing this approach: Rotemberg and Woodford (1997), Christiano, Eichenbaum, and Evans (2005), Altig, Christiano, Eichenbaum, and Linde (2011).

- This sounds like the moment-matching estimator described above, where $\hat{m}_T(Y)$ is now a VAR impulse response function and $\mathbb{E}[\hat{m}_T(Y)|\theta, M_1]$ is the impulse response function obtained from the DSGE model. Unfortunately, a few complications arise:

  - empirical impulse responses are based on finite-order VAR(p)'s.
  - DSGE models, if linearized, have a state-space representation and (i) can be expressed as a low-order VAR(p) in terms of the observables and the structural shocks of the DSGE model; (ii) can be expressed as an infinite-order VAR, where the innovations correspond to the structural shocks of the DSGE model; (iii) can be expressed as an infinite-order VAR, but the innovations of this VAR do not correspond to the innovations of the structural shocks.
  - only in case (i) there is a direct match between the empirical IRF and its model-based counterpart.

- For formal check whether a DSGE model falls within one of the three categories described above, see Fernández-Villaverde, Rubio-Ramírez, Sargent, and Watson (2007). Here is an example: consider the two observationally-equivalent DSGE models

$$\begin{aligned} M_1 &: \quad y_t = \epsilon_t + \theta\epsilon_{t-1} \\ M_2 &: \quad y_t = \theta\epsilon_t + \epsilon_{t-1}, \end{aligned} \tag{142}$$

where $0 < \theta < 1$. The two models are observationally equivalent. Let L denote the lag operator and note that the root of the MA polynomial of model M_1 is outside of the unit circle, whereas the root for model M_2 is inside the unit circle. Thus, for M_1 the MA polynomial is invertible and we can express y_t as an AR(∞) process:

$$M_1 : \; y_t = -\sum_{j=1}^{\infty}(-\theta)^j y_{t-j} + \epsilon_t. \tag{143}$$

Thus, the estimation of an autoregressive model with many lags can reproduce the monotone impulse response function of model M_1, but not the hump-shaped response of model M_2.

- If the DSGE model generates non-invertible moving average terms its impulse responses cannot be approximated by a VAR(∞) and a direct comparison of VAR and DSGE IRFs will be misleading.

- Consistency requirement for IRF matching estimator: suppose the only source of misspecification of the DSGE model is the omission of the "other" shocks. In this case, it should be possible to consistently estimate the DSGE model parameters θ that are related to the transmission of the shock under consideration.

  - in general, this would require that the DSGE model allows for a VAR(∞) representation;
  - letting the number of lags in the empirical VAR go to infinity as $T \longrightarrow \infty$;
  - a VAR identification scheme that correctly identifies the shock of interest if data were generated from a "completion" of the DSGE model. This either works with long-run restrictions or if certain adjustment lags are built into the DSGE model that justify the use of zero-restrictions.

- Some debate in the literature about what can be learned from IRF matching: Chari, Kehoe, and McGrattan (2008) and Christiano, Eichenbaum, and Vigfusson (2007).

- In practice: $\mathbb{E}[\hat{m}_T(Y)|\theta, M_1]$ is typically replaced by the population IRF computed from the DSGE model, denoted by $IRF(\cdot|\theta, M_1)$.

- Asymptotics: the 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.

Common procedures: use first-order asymptotics and delta method to derive limit distribution of $\hat{m}_T(Y)$, see Lutkepohl (1990) and Mittnik and Zadrozny (1993). Phillips (1998) Alternatively, use bootstrap approximation as in Kilian (1999, 1998). For persistent VARs see Phillips (1998), Rossi and Pesavento (2006), and Pesavento and Rossi (2007).

- Issue: choice of weight matrix - preferences versus statistical considerations. Moreover, if the number of responses is large, then the distribution of the IRFs becomes singular, meaning that there is redundant information: Guerron-Quintana, Inoue, and Kilian (2014).
- Nonlinear IRFs: Ruge-Murcia (2014)

Illustration: Compare different versions of the IRF matching estimator: vary VAR specification and weight matrix for IRF discrepancies. Use long sample so that parameter uncertainty is irrelevant. Show discrepancy between IRFs and parameter estimates.

12.4 GMM Estimation

- Equilibrium conditions of DSGE model can be rewritten as

$$\mathbb{E}[g(y_{t-p:t}|\theta, M_1)] = 0 \tag{144}$$

if and only if $\theta = \theta_0$. We discussed in Section 9.2.4 how to use instrumental variables convert conditional moment restrictions into unconditional moment restrictions. To form an estimator, replace population moments by sample moments and let

$$G_T(\theta|Y) = \frac{1}{T}\sum_{t=1}^{T} g(y_{t-p:t}|\theta, M_1). \tag{145}$$

The GMM objective function is then given by

$$Q_T(\theta|Y) = G_T(\theta|Y)'W_T G_T(\theta|Y) \tag{146}$$

and looks identical to the objective function studied in Section 12.2.

- GMM estimator was originally proposed by Hansen (1982), who provided the first-order asymptotics for the estimator. Since then, an explosion of work on GMM estimators. GMM machinery can be applied to estimation of DSGE models.

- Early applications of GMM estimators to DSGE models are Christiano and Eichenbaum (1992), Burnside, Eichenbaum, and Rebelo (1993). These papers consider sufficiently many moment conditions to be able to estimate all the parameters of the DSGE model.

- In general the GMM estimation can be applied to a subset of the equilibrium conditions, e.g., 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 model solution is computationally costly, this can speed up the estimation process. However, GMM has difficulties dealing with latent variables that appear in equilibrium conditions. Would be difficult to estimate, say the coefficients of the λ_t process in our DSGE model because it is not observed.

- Robust to misspecification of other model aspects - but model needs to be solved to be used for actual analysis

- The recent literature has focused on identification-robust inference in view of difficulties with the identification of Phillips curve parameters and policy rule parameters in New Keynesian DSGE models. Survey on weak instrument literature: Stock, Wright, and Yogo (2002). Identification-robust inference for New Keynesian Phillips curve:Mavroeidis (2005), Kleibergen and Mavroeidis (2009), and Mavroeidis, Plagborg-Mller, and Stock (2014). Identification robust inference for monetary policy rule is provided in Mavroeidis (2010) and general identification problems are discussed by Cochrane (2011).

- Laplace type estimators: Chernozukov and Hong (2003) and Kormilitsina and Nekipolov (2013).

- GMM with latent variables: Gallant, Giacomini, and Ragusa (2013).

13 Bayesian Estimation Techniques

- This section relies heavily on Herbst and Schorfheide (2015).
- To implement Bayesian inference we need to specify a prior distribution $p(\theta|M_1)$ and develop numerical techniques to generate draws from the posterior distribution and approximate the posterior mean and quantiles of transformations $h(\theta)$.
- There are many computational techniques available to approximate posterior expectations. The most widely-used technique in the estimation of DSGE models is the Metropolis-Hastings algorithm, which we discuss in Section 13.2.
- Alternatively, one can use sequential Monte Carlo techniques, that are similar in spirit to the particle filter discussed in Section 10.3, except that the particles represent the static parameter θ instead of a sequence of latent states s_t.

13.1 Prior Distributions

- Prior distributions are an essential part of the parameter inference in DSGE models.
- Various branches of econometrics use prior distributions differently: (i) keep the prior "flat" so that the posterior reflects the shape of the likelihood function; (ii) use the prior to deal with a proliferation of parameters; (iii) use the prior to regularize the likelihood function (make it more ellipitical); (iv) use the prior to add substantive information to the estimation problem. The Bayesian analysis of DSGE models uses (iii) and (iv).
- Prior elicitation: Del Negro and Schorfheide (2008). Distinguish steady-state related parameters, exogenous shock parameters, and endogenous propagation parameters (give examples in the context of DSGE model).
- Use non-sample information, e.g., pre-sample or micro-level studies.
- Provide some more specifics in the context of stylized DSGE model.

Table 4: Prior Distribution

Name	Domain	Prior		
		Density	Para (1)	Para (2)
Steady-State Related Parameters $\theta_{(ss)}$				
$100(1-1/\beta)$	$\mathbb{R}^+$	Gamma	0.50	0.50
$100\ln\pi^*$	$\mathbb{R}^+$	Gamma	1.00	0.50
$100\ln\gamma$	$\mathbb{R}$	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.

13.2 MCMC Methods

- Denote posterior distribution by $\pi(\theta)$ and expectations by $\mathbb{E}_\pi[\theta]$.

- Direct sampling and computation of posterior moments is not possible for DSGE models.

- We begin with a Markov chain Monte Carlo algorithm, which constructs a Markov chain such that the stationary distribution associated with this Markov chain is unique and equals the posterior distribution of interest. MCMC algorithms generate a sequence of draws θ^i $i = 1, \ldots, N$ and the distribution of draw i converges to the target posterior distribution as $N \longrightarrow \infty$. Moreover, sample averages of draws converge to posterior expectations:

$$\frac{1}{N - N_0} \sum_{i=N_0+1}^{N} h(\theta^i) \xrightarrow{a.s.} \mathbb{E}_\pi[h(\theta)] \tag{147}$$

- An MCMC algorithm generates a Markov transition kernel $K(\theta^i|\theta^{i-1})$ with the property that

$$\int K(\theta^i|\theta^{i-1})\pi(\theta^{i-1})d\theta^{i-1} = \pi(\theta^i) \tag{148}$$

13.2.1 Metropolis-Hastings Algorithm

- We focus on a particular class of MCMC algorithms, namely, Metropolis-Hastings (MH) algorithms A first version of such an algorithm had been constructed by Metropolis, Rosenbluth, Rosenbluth, Teller, and Teller (1953) to solve a minimization problem and was later generalized by Hastings (1970). Tierney (1994) proved important convergence results for MCMC algorithms and Monte Carlo averages computed based on their output. Chib and Greenberg (1995) provide an excellent introduction to MH algorithms. Detailed textbook treatments can be found, for instance, in Robert and Casella (2004) or Geweke (2005).

- 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. The proposed draw is always accepted if it raises the posterior density (relative to θ^{i-1}) and it is sometimes accepted even if it lowers the posterior density. 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 4 (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 $\alpha(\vartheta|\theta^{i-1})$ by the product of likelihood and prior, which does not require the evaluation of the marginal data density $p(Y)$. Algorithm 4 describes how to generate a parameter draw θ^i conditional on a parameter draw θ^{i-1}. Thus, implicitly it characterizes a Markov transition kernel $K(\theta|\tilde{\theta})$, where the conditioning value $\tilde{\theta}$ corresponds to the parameter draw from iteration $i-1$.

13.2.2 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:

$$\vartheta|\theta^i \sim N\big(\theta^i, c^2\hat{\Sigma}\big) \tag{149}$$

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. Suppose θ, comprises two parameters, say β and δ, that are highly correlated in the posterior distribution. If the

variance of the proposal distribution does not capture this correlation, e.g., the matrix $\hat{\Sigma}$ is diagonal, then the draw ϑ is unlikely to reflect the fact that if β is large then δ should also be large, and vice versa. Therefore, $p(\vartheta|Y)$ is likely to be smaller than $p(\theta^{i-1}|Y)$, and so the proposed draw will be rejected with high probability. As a consequence, the chain will have a high rejection rate, exhibit a high autocorrelation, and the Monte Carlo estimates derived from it will have a high variance.

A good choice for $\hat{\Sigma}$ seeks to incorporate information from the posterior, to potentially capture correlations discussed above. Obtaining this information can 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 MCMC. 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 as the sample size $T \longrightarrow \infty$. There exists a large literature on the asymptotic normality of posterior distributions. Fundamental conditions can be found, for instance, in Johnson (1970).

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 non-elliptical and possibly multi-modal, or if the likelihood function is replaced by a non-differentiable 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 $\hat{\Sigma}$, e.g. 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 which constantly (or automatically) adjust $\hat{\Sigma}$ are known as adaptive samplers and require substantially more elaborate theoretical justifications.

- discuss accuracy assessment through long-run covariance matrix and multiple runs of the chain.

Illustration: Generate data from DSGE model. Use MCMC to generate posterior draws. Show some output from MCMC: sequences of draws, effect of initialization, variance across chains.

13.2.3 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 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 subsectors 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 bth block of ith draw of θ and let $\theta_{<b}^i$ refer to the ith draw of all of the blocks before b and similarly for $\theta_{>b}^i$. Algorithm 5 describes a generic Block MH algorithm.

Algorithm 5 (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 $\vartheta_b \sim q(\cdot| \left[\theta_{<b}^i, \theta_b^{i-1}, \theta_{\geq b}^{i-1}\right])$.

   (b) With probability,

$$\alpha = \max \left\{ \frac{p(\left[\theta_{<b}^i, \vartheta_b, \theta_{>b}^{i-1}\right]|Y) q(\theta_b^{i-1}, |\theta_{<b}^i, \vartheta_b, \theta_{>b}^{i-1})}{p(\theta_{<b}^i, \theta_b^{i-1}, \theta_{>b}^{i-1}|Y) q(\vartheta_b|\theta_{<b}^i, \theta_b^{i-1}, \theta_{>b}^{i-1})}, 1 \right\},$$

   set $\theta_b^i = \vartheta_b$, otherwise set $\theta_b^i = \theta_b^{i-1}$.

In order to make the Block MH algorithm operational the researcher has to decided how to allocate parameters to blocks in each iteration and how to choose the proposal distribution $q(\cdot| \left[\theta_{<b}^i, \theta_b^{i-1}, \theta_{>b}^{i-1}\right])$ 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}, should 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. The first three papers in the

DSGE model literature to consider blocking were Curdia and Reis (2009), Chib and Ramamurthy (2010), and Herbst (2011). Curdia and Reis (2009) group the parameters by type: economic – those related to agents' preferences and production technologies – and statistical – those governing the exogenous processes driving the model. The rationale for this grouping is that it is relatively straightforward to design proposal distributions for the statistical parameters. However, the grouping is unlikely to be optimal, because, for instance, economic parameters related to the persistence generated by the internal propagation mechanism of a DSGE model may be highly correlated with the parameters of the exogenous processes. 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. While there will be correlated blocks sometimes, the randomization ensures that this feature does not persist. Key to the algorithm is that the block configuration is independent of the Markov chain. This is crucial for ensuring the convergence of the chain. Otherwise, the chain is said to be adaptive and the asymptotic theory is substantially more complicated. Herbst (2011) constructs a Block MH algorithm in which the blocking is explicitly based on the posterior correlation structure which is approximated based on draws from a burn-in period. He provides evidence that the distributional blocking procedure outperforms the random blocking.

In order to tailor the block-specific proposal distributions, Chib and Ramamurthy (2010) advocates 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. This algorithm is called Tailorized Random Block MH (TaRBMH) algorithm. While the TaRBMH algorithm is very successful in reducing the persistence of the Markov chain relative to the benchmark RWMH-V algorithm, the downside is that the algorithm is very slow due to the likelihood evaluations required to execute the simulated annealing step and the computation of the Hessian.

13.2.4 Marginal Likelihood Approximations

Discuss Geweke (1999)'s harmonic mean estimator and refer to An and Schorfheide (2007) and Herbst and Schorfheide (2015) for a comparison with the method proposed by Chib and Jeliazkov (2001) and Sims, Waggoner, and Zha (2008).

13.2.5 Particle MCMC

For nonlinear DSGE models the exact likelihood function has to be replaced by a particle filter approximation, which leads to an algorithm that belongs to the class of particle MCMC methods. We refer to the combination of a particle-filter approximated likelihood and the MH algorithm as PFMH algorithm. This idea was first proposed for the estimation of nonlinear DSGE models by Fernández-Villaverde and Rubio-Ramírez (2007). The statistical theory underlying the PFMH algorithm is very complex and beyond the scope of this chapter. We refer the interested reader to Andrieu, Doucet, and Holenstein (2010) for a careful exposition and to Flury and Shephard (2011) for other applications in econometrics. By replacing the exact likelihood function $p(\theta|Y)$ with the particle filter approximation $\hat{p}(Y|\theta)$ in Algorithm 4 one might expect to obtain draws from the approximate posterior $\hat{p}(\theta|Y)$ instead of the exact posterior $p(\theta|Y)$. The surprising implication of the theory developed in Andrieu, Doucet, and Holenstein (2010) is that the distribution of draws from the PFMH algorithm that replaces $p(Y|\theta)$ by $\hat{p}(Y|\theta)$ in fact does converge to the exact posterior. The key requirement is that the likelihood approximation generated by the particle filter is unbiased. The algorithm takes the following form:

Algorithm 6 (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 3.

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.

13.3 SMC Methods

- References: Chopin (2002) showed how to adapt the particle filtering techniques to conduct posterior inference for a static parameter vector. Textbook treatments of SMC algorithms can be found, for instance, in Liu (2001) and Cappé, Moulines, and Ryden (2005). The volume by Doucet, de Freitas, and Gordon (2001) discusses many applications and practical aspects of SMC. Creal (2012) provides a recent survey focusing on SMC applications in econometrics. The first paper that applied SMC techniques to posterior inference in a small-scale DSGE models is Creal (2007). Herbst and Schorfheide (2014) developed the algorithm further, provided some convergence results for an adaptive version of the algorithm building on the theoretical analysis of Chopin (2004), and showed that a properly tailored SMC algorithm delivers more reliable posterior inference for large-scale DSGE models with multi-modal posterior than the widely-used RMWH-V algorithm. Much of exposition in this section borrows from Herbst and Schorfheide (2014) and Herbst and Schorfheide (2015). An additional advantage of the SMC algorithms over MCMC algorithms, on the computational front, highlighted by Durham and Geweke (2012), is that SMC is much more amenable to parallelization. Durham and Geweke (2012) show how to implement a SMC algorithm on graphical processing unit (GPU), facilitating massive speed gains in estimations.

- Construct a sequence of posterior distributions. Suppose ϕ_n, $n = 1, \ldots, N_\phi$, is a sequence that slowly increases from zero to one. We can define a sequence of tempered posteriors as

$$\pi_n(\theta) = \frac{[p(Y|\theta)]^{\phi_n} p(\theta)}{\int [p(Y|\theta)]^{\phi_n} p(\theta) d\theta} \quad n = 0, \ldots, N_\phi, \quad \phi_n \uparrow 1. \tag{150}$$

  Alternatively, one could construct the sequence of posteriors by sequentially adding observations to the likelihood function, that is, $\pi_n(\theta)$ is based on $p(Y_{1:\lfloor \phi_n T \rfloor}|\theta)$:

$$\pi_n^{(D)}(\theta) = \frac{p(Y_{1:\lfloor \phi_n T \rfloor}) p(\theta)}{\int p(Y_{1:\lfloor \phi_n T \rfloor}) p(\theta) d\theta}. \tag{151}$$

  This data tempering is particularly attractive in sequential applications. Due to the fact that individual observations are not divisible, the data tempering approach is slightly less flexible.

- 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

$$\bar{h}_{n,N} = \frac{1}{N}\sum_{i=1}^{N} W_n^i h(\theta^i) \xrightarrow{a.s.} \mathbb{E}_\pi[h(\theta_n)]. \tag{152}$$

Illustration: Generate data from DSGE model. Implement SMC algorithm. Show waterfall plot of sequence of tempered posteriors. Maybe illustrate propagation of particles. Can simplify DSGE model by fixing a subset of the parameters.

13.3.1 The SMC Algorithm

Starting from stage $n-1$ particles $\{\theta_{n-1}^i, W_{n-1}^i\}_{i=1}^N$ the algorithm proceeds in three steps, using Chopin (2004)'s terminology: *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 7 (Generic SMC Algorithm with Likelihood Tempering)**.**

1. **Initialization.** ($\phi_0 = 0$). Draw the initial particles from the prior: $\theta_1^i \overset{iid}{\sim} p(\theta)$ and $W_1^i = 1$, $i = 1, \ldots, N$.

2. **Recursion.** For $n = 1, \ldots, N_\phi$,

   (a) **Correction.** Reweight the particles from stage $n-1$ by defining the incremental weights

   $$\tilde{w}_n^i = [p(Y|\theta_{n-1}^i)]^{\phi_n - \phi_{n-1}} \tag{153}$$

   and the normalized weights

   $$\tilde{W}_n^i = \frac{\tilde{w}_n^i W_{n-1}^i}{\frac{1}{N}\sum_{i=1}^N \tilde{w}_n^i W_{n-1}^i}, \quad i = 1, \ldots, N. \tag{154}$$

   An approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$ is given by

   $$\tilde{h}_{n,N} = \frac{1}{N}\sum_{i=1}^{N} \tilde{W}_n^i h(\theta_{n-1}^i). \tag{155}$$

(b) **Selection.**

Case (i): If $\rho_n = 1$, resample the particles via multinomial resampling. Let $\{\hat{\theta}\}_{i=1}^N$ 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 $W_n^i = 1$.
Case (ii): If $\rho_n = 0$, let $\hat{\theta}_n^i = \theta_{n-1}^i$ and $W_n^i = \tilde{W}_n^i$, $i = 1, \ldots, N$. An approximation of $\mathbb{E}_{\pi_n}[h(\theta)]$ is given by

$$\hat{h}_{n,N} = \frac{1}{N} \sum_{i=1}^N W_n^i h(\hat{\theta}_n^i). \tag{156}$$

(c) **Mutation.** Propagate the particles $\{\hat{\theta}_i, W_n^i\}$ via N_{MH} steps of a 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

$$\bar{h}_{n,N} = \frac{1}{N} \sum_{i=1}^N h(\theta_n^i) W_n^i. \tag{157}$$

3. For $n = N_\phi$ ($\phi_{N_\phi} = 1$) the final importance sampling approximation of $\mathbb{E}_\pi[h(\theta)]$ is given by:

$$\bar{h}_{N_\phi,N} = \sum_{i=1}^N h(\theta_{N_\phi}^i) W_{N_\phi}^i. \tag{158}$$

13.3.2 Tailoring the SMC Algorithm

- The transition kernel in the mutation step can be generated by N_{MH} steps of Algorithm 5. A single step may be sufficient. The tuning constants c^2 and $\hat{\Sigma}$ for a RWMH algorithm can be constructed adaptively; $\hat{\Sigma}$ is obtained as the approximation of the posterior covariance matrix from the output of the correction step and c^2 can be adjusted adaptively based on rejection rates at earlier stages to reach a desired rejection rate.

- Remaining choices: tempering schedule, number of particles, number of MH steps.

- See Herbst and Schorfheide (2015) for illustrations and trade-offs in the choice of tuning parameters.

13.3.3 Marginal Likelihood Approximation and SMC^2

- Marginal likelihood can be computed based on unnormalized weights.
- If the exact likelihood function is replaced by a particle filter approximation then we obtain an SMC^2 algorithm, see Chopin, Jacob, and Papaspiliopoulos (2012) and the DSGE model application in Herbst and Schorfheide (2015).

13.4 Model Evaluation

- Assessing the relative fit of DSGE models using posterior odds; Bayesian model averaging.
- References: Lancaster (2004) and Geweke (2005). The absolute fit of a Bayes model can be characterized using prior or posterior predictive checks. Let $Y^*_{1:T}$ be an artificial sample of length T. The predictive distribution for $Y^*_{1:T}$ based on time t information set $\mathcal{F}_t$ is
$$p(Y^*_{1:T}|\mathcal{F}_t) = \int p(Y^*_{1:T}|\theta)p(\theta|\mathcal{F}_t)d\theta. \tag{159}$$
Here $\mathcal{F}_0$ corresponds to prior information and $\mathcal{F}_T$ corresponds to posterior distribution. Consider transformation $\mathcal{S}(Y_{1:T})$, e.g., sample autocovariances $vech(\hat{\Gamma}_{yy}(0))$. If the value of this statistic computed from actual data falls far into the tails of the prior or posterior predictive distribution, then it indicates that the model is inconsistent with this particular feature of the data. Application: Chang, Doh, and Schorfheide (2007).

13.5 Bayesian Inference At Work

There are a large number of papers that apply Bayesian techniques to estimate DSGE model. Below we are referencing a few of them.

- Early applications of Bayesian estimation of DSGE models: DeJong, Ingram, and Whiteman (2000), Schorfheide (2000), Otrok (2001), Fernandez-Villaverde and Rubio-Ramirez (2004), and Rabanal and Rubio-Ramírez (2005).
- Large scale models usable for policy analysis: Smets and Wouters (2003) and Smets and Wouters (2007).

- Policy analysis under uncertainty: Levin, Onatski, Williams, and Williams (2006)
- Allowing for indeterminacy: Lubik and Schorfheide (2004)
- Open economy models: Lubik and Schorfheide (2006)
- Adding stochastic volatility: Justiniano and Primiceri (2008)
- Fat-tailed error distributions: Curdia, Del Negro, and Greenwald (2014)
- Fully nonlinear estimation: Fernández-Villaverde and Rubio-Ramírez (2007), Fernández-Villaverde and Rubio-Ramírez (2008)
- Regime switches in intercepts Schorfheide (2005a) and slope coefficients Bianchi (2013)
- Forecasting: handbook chapter of Del Negro and Schorfheide (2013)
- Identification and robustness to prior: Koop, Pesaran, and Smith (2013) and Müller (2012).
- Choice of observables: Guerron-Quintana (2010)

13.6 Limited Information Bayesian Inference

Bayesian inference requires a likelihood function $p(Y|\theta)$. However, as we have seen, many of the classical approaches to DSGE model estimation, e.g., (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 of the likelihood function. These approaches are often called limited information approaches (as opposed to full-information inference based on the likelihood function). In this subsection we explore several Bayesian approaches to limited-information inference.

- Lubik and Schorfheide (2005) estimated 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, e.g. output gap and inflation in the case of our stylized model. This generates a VAR for output, inflation, and interest rates, with cross-coefficient restrictions that depend on the monetary policy rule parameters.

Kleibergen and Mavroeidis (2014) uses a similar approach, focusing on the specification of prior distributions, and shows how it can be applied to the estimation of a New Keynesian Phillips curve.

- Suppose we know the sampling distribution $p(\hat{\theta}|\theta)$ of an estimator $\hat{\theta}$. Then, instead of updating our beliefs conditional on the observed sample Y, we could update the beliefs about θ based on the realization of $\hat{\theta}$:

$$p(\theta|\hat{\theta}) = \frac{p(\hat{\theta}|\theta)p(\theta)}{\int p(\hat{\theta}|\theta)p(\theta)}. \tag{160}$$

This idea dates back at least to Pratt, Raiffa, and Schlaifer (1965). The problem with this approach is that we typically do not have a closed-form representation of the density $p(\hat{\theta}|\theta)$. We could replace $p(\hat{\theta}|\theta)$ by a simulation-based approximation, an idea set forth by Boos and Monahan (1986). Alternatively, we could replace the finite-sample distribution by a limit distribution, e.g.,

$$\sqrt{T}(\hat{\theta} - \theta)|\theta \Longrightarrow N\big(0, V(\theta)\big). \tag{161}$$

However, it is important that this convergence is uniform in θ, because we want to approximate a posterior density for θ. Thus, we would need something like: for every converging sequence $\theta_T \longrightarrow \theta$,

$$\sqrt{T}(\hat{\theta} - \theta_T)|\theta_T \Longrightarrow N\big(0, V(\theta)\big). \tag{162}$$

Note that this uniform convergence is typically not attainable as θ_T approaches the boundary of the stationary region of the parameter space. This idea is explored in Kwan (1999). Rather than making statements about the approximation of posterior distribution, Müller (2013) shows that decisions based on the quasi-posterior are asymptotically optimal under fairly general conditions.

- Kim (2002) constructs a limited-information likelihood function as follows. Suppose the data is generated under the probability measure $\mathbb{P}$ and at θ_0 we have the moment condition $\mathbb{E}_{\mathbb{P}}[g(y_{t-p:t}|\theta_0)] = 0$, thus we are considering the GMM setup. Using a weight matrix that corresponds to the inverse of the long-run covariance matrix of the sample analogue of the moment conditions and imposing uniform integrability, then, $\lim_{T\longrightarrow\infty} \mathbb{E}_{\mathbb{P}}[Q_T(\theta_0|Y)] = k$, where k is the dimension of θ.

To obtain a limited-information likelihood function, consider the class of distributions $\mathcal{P}(\theta)$ defined as follows:

$$\mathcal{P}(\theta) = \left\{ P \mid \lim_{T \longrightarrow \infty} \mathbb{E}_P[TQ_T(\theta|Y)] = k \right\}. \tag{163}$$

Because $\mathcal{P}(\theta)$ is not a singleton, we do not have a likelihood function yet. To obtain a unique distribution for each θ, let's project the "true" distribution $\mathbb{P}$ onto the set $\mathcal{P}(\theta)$ using the Kullback-Leibler discrepancy as the metric:

$$P^*(\theta|Y) = \text{argmin}_{P \in \mathcal{P}(\theta)} \int \ln(dP/d\mathbb{P}) dP, \tag{164}$$

where $p^*(\theta|Y) = dP/d\mathbb{P}$ is the Radon-Nikodym derivative of P with respect to $\mathbb{P}$. The solution takes the form

$$p^*(\theta|Y) \propto \exp\left\{ -\frac{1}{2} Q_T(\theta|Y) \right\}. \tag{165}$$

Practical issues: dependence of weight matrix in $Q_T(\theta)$ objective function on θ; use of quadratic expansion of the log likelihood to obtain Gaussian posterior.

This idea has been used by Christiano, Trabandt, and Walentin (2010) to construct a Bayesian impulse response function matching estimator. Inoue and Shintani (2014) construct a limited-information marginal likelihood that can be used for model selection.

- Nonparametric likelihood function: empirical likelihood as in Lazar (2003) and Schennach (2005). Roughly, assign probability p_t to observation y_t and let the likelihood function be $\prod_{t=1}^T p_t$. Impose side constraint $\sum_{t=1}^T p_t g(y_{t-p:t}|\theta)$ and concentrate 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 use 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.

14 Conclusion

(TBW)

References

Altig, D., L. Christiano, M. Eichenbaum, and J. Linde (2011): "Firm-Specific Capital, Nominal Rigidities and the Business Cycle," *Review of Economic Dynamics*, 14(2), 225–247.

Altug, S. (1989): "Time-to-Build and Aggregate Fluctuations: Some New Evidence," *International Economic Review*, 30(4), 889–920.

An, S., and F. Schorfheide (2007): "Bayesian Analysis of DSGE Models," *Econometric Reviews*, 26(2-4), 113–172.

Andreasen, M. M. (2013): "Non-Linear DSGE Models and the Central Difference Kalman Filter," *Journal of Applied Econometrics*, 28(6), 929–955.

Andreasen, M. M., J. Fernández-Villaverde, and J. F. Rubio-Ramírez (2013): "The Pruned State-Space System for Nonlinear DSGE Models: Theory and Empircal Applications," *Manuscript, University of Pennsylvania*, (4).

Andrews, I., and A. Mikusheva (2015): "Maximum Likelihood Inference in Weakly Identified DSGE Models," *Quantitative Economics*, forthcoming.

Andrieu, C., A. Doucet, and R. Holenstein (2010): "Particle Markov Chain Monte Carlo Methods," *Journal of the Royal Statistical Society Series B*, 72(3), 269–342.

Arulampalam, S., S. Maskell, N. Gordon, and T. Clapp (2002): "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Transactions on Signal Processing*, 50(2), 174–188.

Bianchi, F. (2013): "Regime Switches, Agents' Beliefs, and Post-World War II U.S. Macroeconomic Dynamics," *Review of Economic Studies*, 80(2), 463–490.

Boos, D. D., and J. F. Monahan (1986): "Bootstrap Methods Using Prior Information," *Biometrika*, 73(1), 77–83.

Burnside, C., M. Eichenbaum, and S. Rebelo (1993): "Labor Hoarding and the Business Cycle," *Journal of Political Economy*, 101(2), 245–73.

Canova, F. (2014): "Bridging Cyclical DSGE Models and the Raw Data," *Journal of Monetary Economics*, 67, 1–15.

Canova, F., F. Ferroni, and C. Matthes (2014): "Choosing the Variables to Estimate Singular DSGE Models," *Journal of Applied Econometrics*, 29(7), 1099–1117.

Canova, F., and L. Sala (2009): "Back to Square One: Identification Issues in DSGE Models," *Journal of Monetary Economics*, 56, 431 – 449.

Cappé, O., S. J. Godsill, and E. Moulines (2007): "An Overview of Existing Methods and Recent Advances in Sequential Monte Carlo," *Proceedings of the IEEE*, 95(5), 899–924.

Cappé, O., E. Moulines, and T. Ryden (2005): *Inference in Hidden Markov Models*. Springer Verlag.

Chang, Y., T. Doh, and F. Schorfheide (2007): "Non-stationary Hours in a DSGE Model," *Journal of Money, Credit, and Banking*, 39(6), 1357–1373.

Chari, V. V., P. J. Kehoe, and E. R. McGrattan (2008): "Are Structural VARs with Long-Run Restrictions Useful in Developing Business Cycle Theory?," *Journal of Monetary Economics*, 55(8), 1337–1352.

Chib, S., and E. Greenberg (1995): "Understanding the Metropolis-Hastings Algorithm," *The American Statistician*, 49, 327–335.

Chib, S., and I. Jeliazkov (2001): "Marginal Likelihoods from the Metropolis Hastings Output," *Journal of the American Statistical Association*, 96(453), 270–281.

Chib, S., and S. Ramamurthy (2010): "Tailored Randomized Block MCMC Methods with Application to DSGE Models," *Journal of Econometrics*, 155(1), 19–38.

Chopin, N. (2002): "A Sequential Particle Filter for Static Models," *Biometrika*, 89(3), 539–551.

——— (2004): "Central Limit Theorem for Sequential Monte Carlo Methods and its Application to Bayesian Inference," *Annals of Statistics*, 32(6), 2385–2411.

Chopin, N., P. E. Jacob, and O. Papaspiliopoulos (2012): "SMC^2: An Efficient Algorithm for Sequential Analysis of State-Space Models," *arXiv:1101.1528*.

Christiano, L., and M. Eichenbaum (1992): "Current Real-Business Cycle Theories and Aggregate Labor-Market Fluc," *American Economic Review*, 82(3), 430–450.

Christiano, L. J., M. Eichenbaum, and C. L. Evans (1999): "Monetary Policy Shocks: What Have We Learned and to What End," in *Handbook of Macroeconomics*, ed. by J. B. Taylor, and M. Woodford, vol. 1a, chap. 2, pp. 65–148. North Holland, Amsterdam.

——— (2005): "Nominal Rigidities and the Dynamic Effects of a Shock to Monetary Policy," *Journal of Political Economy*, 113(1), 1–45.

Christiano, L. J., M. Eichenbaum, and R. Vigfusson (2007): "Assessing Structural VARs," in *NBER Macroeconomics Annual 2006*, ed. by D. Acemoglu, K. Rogoff, and M. Woodford, vol. 21, pp. 1–72. MIT Press, Cambridge.

Christiano, L. J., M. Trabandt, and K. Walentin (2010): "DSGE Models for Monetary Policy Analysis," in *Handbook of Monetary Economics*, ed. by B. M. Friedman, and M. Woodford, vol. 3 of *Handbook of Monetary Economics*, chap. 7, pp. 285–367. Elsevier.

Christiano, L. J., and R. J. Vigfusson (2003): "Maximum Likelihood in the Frequency Domain: The Importance of Time-to-plan," *Journal of Monetary Economics*, 50(4), 789–815.

Cochrane, J. H. (1994): "Shocks," *Carnegie Rochester Conference Series on Public Policy*, 41(4), 295–364.

——— (2011): "Determinacy and Identification with Taylor Rules," *Journal of Political Economy*, 119(3), 565 – 615.

Creal, D. (2007): "Sequential Monte Carlo Samplers for Bayesian DSGE Models," *Unpublished Manuscript, Vrije Universitiet*.

——— (2012): "A Survey of Sequential Monte Carlo Methods for Economics and Finance," *Econometric Reviews*, 31(3), 245–296.

Curdia, V., M. Del Negro, and D. L. Greenwald (2014): "Rare Shocks, Great Recessions," *Journal of Applied Econometrics*, 29(7), 1031–1052.

Curdia, V., and R. Reis (2009): "Correlated Disturbances and U.S. Business Cycles," *Working Paper*.

DeJong, D. N., B. F. Ingram, and C. H. Whiteman (1996): "A Bayesian Approach to Calibration," *Journal of Business Economics and Statistics*, 14(4), 1–9.

——— (2000): "A Bayesian Approach to Dynamic Macroeconomics," *Journal of Econometrics*, 98(2), 203 – 223.

Del Moral, P. (2013): *Mean Field Simulation for Monte Carlo Integration*. Chapman & Hall/CRC.

Del Negro, M., R. Hasegawa, and F. Schorfheide (2014): "Dynamic Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance," *NBER Working Paper*, 20575.

Del Negro, M., and F. Schorfheide (2004): "Priors from General Equilibrium Models for VARs," *International Economic Review*, 45(2), 643 – 673.

——— (2008): "Forming Priors for DSGE Models (and How it Affects the Assessment of Nominal Rigidities)," *Journal of Monetary Economics*, 55(7), 1191–1208.

——— (2009): "Monetary Policy with Potentially Misspecified Models," *American Economic Review*, 99(4), 1415–1450.

——— (2013): "DSGE Model-Based Forecasting," in *Handbook of Economic Forecasting*, ed. by G. Elliott, and A. Timmermann, vol. 2, forthcoming. North Holland, Amsterdam.

Del Negro, M., F. Schorfheide, F. Smets, and R. Wouters (2007): "On the Fit of New Keynesian Models," *Journal of Business and Economic Statistics*, 25(2), 123–162.

Doucet, A., N. de Freitas, and N. Gordon (eds.) (2001): *Sequential Monte Carlo Methods in Practice*. Springer Verlag.

Doucet, A., and A. M. Johansen (2011): "A Tutorial on Particle Filtering and Smoothing: Fifteen Years Later," in *Handook of Nonlinear Filtering*, ed. by D. Crisan, and B. Rozovsky. Oxford University Press.

Dridi, R., A. Guay, and E. Renault (2007): "Indirect Inference and Calibration of Dynamic Stochastic General Equilibrium Models," *Journal of Econometrics*, 136(2), 397–430.

Durbin, J., and S. J. Koopman (2001): *Time Series Analysis by State Space Methods*. Oxford University Press.

Durham, G., and J. Geweke (2012): "Adaptive Sequential Posterior Simulators for Massively Parallel Computing Environments," *Unpublished Manuscript*.

Fernandez-Villaverde, J., and J. F. Rubio-Ramirez (2004): "Comparing dynamic equilibrium models to data: a Bayesian approach," *Journal of Econometrics*, 123(1), 153 – 187.

Fernández-Villaverde, J., and J. F. Rubio-Ramírez (2007): "Estimating Macroeconomic Models: A Likelihood Approach," *Review of Economic Studies*, 74(4), 1059–1087.

——— (2008): "How Structural are Structural Parameters?," in *NBER Macroeconomics Annual 2007*, ed. by D. Acemoglu, K. Rogoff, and M. Woodford, vol. 22. University of Chicago Press, Chicago, University of Chicago Press.

Fernández-Villaverde, J., J. F. Rubio-Ramírez, T. J. Sargent, and M. W. Watson (2007): "ABCs (and Ds) of Understanding VARs," *American Economic Review*, 97(3), 1021–1026.

Flury, T., and N. Shephard (2011): "Bayesian Inference Based Only On Simulated Likelihood: Particle Filter Analysis of Dynamic Economic Models," *Econometric Theory*, 27, 933–956.

Gallant, A. R., R. Giacomini, and G. Ragusa (2013): "Generalized Method of Moments with Latent Variables," *CEPR Discussion Papers*, (DP9692).

Geweke, J. (1999): "Using Simulation Methods for Bayesian Econometric Models: Inference, Development, and Communication," *Econometric Reviews*, 18(1), 1–126.

——— (2005): *Contemporary Bayesian Econometrics and Statistics*. John Wiley & Sons, Inc.

——— (2010): *Complete and Incomplete Econometric Models*. Princeton University Press, Princeton, New Jersey.

Geweke, J., and G. Amisano (2011): "Optimal Prediction Pools," *Journal of Econometrics*, 164, 130–141.

——— (2012): "Prediction with Misspecified Models," *American Economic Review: Papers & Proceedings*, 103(3), 482–486.

Gorodnichenko, Y., and S. Ng (2010): "Estimation of DSGE Models When the Data are Persistent," *Journal of Monetary Economics*, 57(3), 325–340.

Guerron-Quintana, P., A. Inoue, and L. Kilian (2013): "Frequentist Inference in Weakly Identified Dynamic Stochastic General Equilibrium Models," *Quantitative Economics*, 4, 197–229.

——— (2014): "Impulse response matching estimators for DSGE models," .

Guerron-Quintana, P. A. (2010): "What You Match Does Matter: The Effects of Observable Variables on DSGE Estimation," *Journal of Applied Econometrics*, 25, 774–804.

Guo, D., X. Wang, and R. Chen (2005): "New Sequential Monte Carlo Methods for Nonlinear Dynamic Systems," *Statistics and Computing*, 15, 135.147.

Hamilton, J. D. (1994): *Time Series Analysis*. Pri.

Hansen, L. P. (1982): "Large Sample Properties of Generalized Method of Moments Estimators," *Econometrica*, 50(4), 1029–54.

Hastings, W. (1970): "Monte Carlo Sampling Methods Using Markov Chains and Their Applications," *Biometrika*, 57, 97–109.

Herbst, E. (2011): "Gradient and Hessian-based MCMC for DSGE Models," *Unpublished Manuscript, University of Pennsylvania*.

Herbst, E., and F. Schorfheide (2014): "Sequential Monte Carlo Sampling for DSGE Models," *Journal of Applied Econometrics*, forthcoming.

——— (2015): *Bayesian Estimation of DSGE Models*. Princeton University Press, forthcoming.

Ingram, B., and C. Whiteman (1994): "Supplanting the Minnesota Prior- Forecasting Macroeconomic Time Series Using Real Business Cycle Model Priors," *Journal of Monetary Economics*, 49(4), 1131–1159.

Inoue, A., and M. Shintani (2014): "Quasi-Bayesian Model Selection," *Manuscript, Vanderbilt University*.

Iskrev, N. (2010): "Local Identification of DSGE Models," *Journal of Monetary Economics*, 2, 189–202.

Johnson, R. (1970): "Asymptotic Expansions Associated with Posterior Distributions," *Annals of Mathematical Statistics*, 41, 851–864.

Justiniano, A., and G. E. Primiceri (2008): "The Time-Varying Volatility of Macroeconomic Fluctuations," *American Economic Review*, 98(3), 604–641.

Kantas, N., A. Doucet, S. Singh, J. Maciejowski, and N. Chopin (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," *Review of Economics and Statistics*, 80(2), 218–230.

——— (1999): "Finite-Sample Properties of Percentile and Percentile-t Bootstrap Confidence Intervals for Impulse Responses," *Review of Economics and Statistics*, 81(4), 652–660.

Kim, J.-Y. (2002): "Limited Information Likelihood and Bayesian Analysis," *Journal of Econometrics*, 107(1-2), 175 – 193.

King, R. G., C. I. Plosser, and S. Rebelo (1988): "Production, Growth, and Business Cycles: I The Basic Neoclassical Model," *Journal of Monetary Economics*, 21(2-3), 195–232.

Kitamura, Y., and T. Otsu (2011): "Bayesian Analysis of Moment Condition Models Using Nonparametric Priors," *Manuscript, Yale University and LSE*.

Kleibergen, F., and S. Mavroeidis (2009): "Weak Instrument Robust Tests in GMM and the New Keynesian Phillips Curve," *Journal of Business & Economic Statistics*, 27(3), 293–311.

——— (2014): "Identification Issues in Limited-Information Bayesian Analysis of Structural Macroeconomic Models," *Journal of Applied Econometrics*, 29, 1183–1209.

Kociecki, A., and M. Kolasa (2015): "Global Identification of Linearized DSGE Models," *Manuscript, Bank of Poland*.

Kollmann, R. (2014): "Tractable Latent State Filtering for Non-Linear DSGE Models Using a Second-Order Approximation and Pruning," *Computational Economics*, forthcoming.

Komunjer, I., and S. Ng (2011): "Dynamic Identification of DSGE Models," *Econometrica*, 79(6), 1995–2032.

Koop, G., H. M. Pesaran, and R. P. Smith (2013): "On Identification of Bayesian DSGE Models," *Journa of Business & Economic Statistics*, 31(3), 300–314.

Kwan, Y. K. (1999): "Asymptotic Bayesian Analysis Based on a Limited Information Estimator," *Journal of Econometrics*, 88, 99–121.

Lancaster, T. (2004): *An Introduction to Modern Bayesian Econometrics*. Blackwell Publishing.

Lazar, N. A. (2003): "Bayesian Empirical Likelihood," *Biometrika*, 90(2), 319–326.

Lee, B.-S., and B. F. Ingram (1991): "Simulation estimation of time-series models," *Journal of Econometrics*, 47(2-3), 197–205.

Leeper, E. M., and C. A. Sims (1995): "Toward a Modern Macroeconomic Model Usable for Policy Analysis," in *NBER Macroeconomics Annual 1994*, ed. by S. Fischer, and J. J. Rotemberg, pp. 81–118. MIT Press, Cambridge.

Levin, A., A. Onatski, J. C. Williams, and N. Williams (2006): "Monetary Policy Under Uncertainty in Micro-founded Macroeconometric Models," in *NBER Macroeconomics Annual 2005*, ed. by M. Gertler, and K. Rogoff, vol. 20, pp. 229–287. MIT Press, Cambridge.

Liu, J. S. (2001): *Monte Carlo Strategies in Scientific Computing*. Springer Verlag.

Lubik, T., and F. Schorfheide (2005): "Do Central Banks Respond to Exchange Rate Movements? A Structural Investigation," *Journal of Monetary Economics*, 54(4), 1069 – 1087.

——— (2006): "A Bayesian Look at the New Open Macroeconomics," *NBER Macroeconomics Annual 2005*.

Lubik, T. A., and F. Schorfheide (2004): "Testing for Indeterminancy: An Application to U.S. Monetary Policy," *American Economic Review*, 94(1), 190–217.

Lutkepohl, H. (1990): "Asymptotic Distributions of Impulse Response Functions and Forecast Error Variance Decompositions of Vector Autoregressive Models," *The Review of Economics and Statistics*, 72(1), 116–25.

Malik, S., and M. K. Pitt (2011): "Particle Filters for Continuous Likelihood Evaluation and Maximization," *Journal of Econometrics*, 165, 190–209.

Mavroeidis, S. (2005): "Identification Issues in Forward-Looking Models Estimated by GMM, with an Application to the Phillips Curve," *Journal of Money, Credit and Banking*, 37(3), 421–48.

——— (2010): "Monetary Policy Rules and Macroeconomic Stability: Some New Evidence," *The American Economic Review*, 100(1), 491–503.

Mavroeidis, S., M. Plagborg-Mller, and J. H. Stock (2014): "Empirical Evidence on Inflation Expectations in the New Keynesian Phillips Curve," *Journal of Economic Literature*, 52(1), 124–88.

McGrattan, E. R. (1994): "The Macroeconomic Effects of Distortionary Taxation," *Journal of Mon*, 33(3), 573–601.

Metropolis, N., A. Rosenbluth, M. Rosenbluth, A. Teller, and E. Teller (1953): "Equations of State Calculations by Fast Computing Machines," *Journal of Chemical Physics*, 21, 1087–1091.

Mittnik, S., and P. A. Zadrozny (1993): "Asymptotic Distributions of Impulse Responses, Step Responses, and Variance Decompositions of Estimated Linear Dynamic Models," *Econometrica*, 61(4), 857–70.

Müller, U. (2012): "Measuring Prior Sensitivity and Prior Informativeness in Large Bayesian Models," *Journal of Monetary Economics*, 59, 581–597.

——— (2013): "Risk of Bayesian Inference in Misspecified Models, and the Sandwich Covariance Matrix," *Econometrica*, 81(5), 1805–1849.

Otrok, C. (2001): "On Measuring the Welfare Costs of Business Cycles," *Journal of Monetary Economics*, 47(1), 61–92.

Pakes, A., and D. Pollard (1989): "Simulation and the Asymptotics of Optimization Estimators," *Econometrica*, 57(5), 1027–57.

Pesavento, E., and B. Rossi (2007): "Impulse response confidence intervals for persistent data: What have we learned?," *Journal of Economic Dynamics and Control*, 31(7), 2398 – 2412.

Phillips, P. C. B. (1998): "Impulse response and forecast error variance asymptotics in nonstationary VARs," *Journal of Econometrics*, 83(1-2), 21–56.

Pratt, J. W., H. Raiffa, and R. Schlaifer (1965): *Introduction to Statistical Decision Theory*. Wiley, New York.

Qu, Z. (2014): "Inference in Dynamic Stochastic General Equilibrium Models with Possible Weak Identification," *Quantitative Economics*, 5, 457–494.

Qu, Z., and D. Tkachenko (2012): "Identification and Frequency Domain Quasi-Maximum Likelihood Estimation of Linearized DSGE Models," *Quantitative Economics*, 3, 95–132.

Qu, Z., and D. Tkachenko (2014): "Local and Global Parameter Identification in DSGE Models: Allowing for Indeterminancy," *Manuscript, Boston University*.

Rabanal, P., and J. F. Rubio-Ramírez (2005): "Comparing New Keynesian Models of the Business Cycle: A Bayesian Approach," *Journal of Monetary Economics*, 52(6), 1151–1166.

Robert, C. P., and G. Casella (2004): *Monte Carlo Statistical Methods*. Springer.

Rossi, B., and E. Pesavento (2006): "Small-sample confidence intervals for multivariate impulse response functions at long horizons," *Journal of Applied Econometrics*, 21(8), 1135–1155.

Rotemberg, J. J., and M. Woodford (1997): "An Optimization-Based Econometric Framework for the Evaluation of Monetary Policy," in *NBER Macroeconomics Annual 1997*, ed. by B. S. Bernanke, and J. J. Rotemberg. MIT Press, Cambridge.

Ruge-Murcia, F. (2012): "Estimating nonlinear DSGE models by the simulated method of moments: With an application to business cycles," *Journal of Economic Dynamics and Control*, 36, 914938.

——— (2014): "Indirect Inference Estimation of Nonlinear Dynamic General Equilibrium Models: With an Application to Asset Pricing under Skewness Risk," Working paper, McGill University.

Ruge-Murcia, F. J. (2007): "Methods to estimate dynamic stochastic general equilibrium models," *Journal of Economic Dynamics and Control*, 31(8), 2599–2636.

Sala, L. (2015): "DSGE Models in the Frequency Domain," *Journal of Applied Econometrics*, 30, 219–240.

Schennach, S. M. (2005): "Bayesian Exponential Tilted Empirical Likelihood," *Biometrika*, 92, 31–46.

Schorfheide, F. (2000): "Loss Function-based Evaluation of DSGE Models," *Journal of Applied Econometrics*, 15, 645–670.

——— (2005a): "Learning and Monetary Policy Shifts," *Review of Economic Dynamics*, 8(2), 392–419.

——— (2005b): "VAR Forecasting Under Misspecification," *Journal of Econometrics*, 128(1), 99 – 136.

——— (2013): "Estimation and Evaluation of DSGE Models: Progress and Challenges," in *Advances in Economics and Econometrics: Tenth World Congress*, ed. by D. Acemoglu, M. Arellano, and E. Dekel, vol. III, chap. 5, pp. 184–230. Cambridge University Press.

Schwarz, G. (1978): "Estimating the Dimension of a Model," *Annals of Statistics*, 6(2), 461–464.

Shin, M. (2014): "Bayesian GMM," *Manuscript, University of Pennsylvania*.

Sims, C. A., D. Waggoner, and T. Zha (2008): "Methods for Inference in Large Multiple-Equation Markov-Switching Models," *Journal of Econometrics*, 146(2), 255–274.

Smets, F., and R. Wouters (2003): "An Estimated Dynamic Stochastic General Equilibrium Model of the Euro Area," *Journal of the European Economic Association*, 1(5), 1123–1175.

Smets, F., and R. Wouters (2007): "Shocks and Frictions in US Business Cycles: A Bayesian DSGE Approach," *American Economic Review*, 97, 586–608.

Smith, A A, J. (1993): "Estimating Nonlinear Time-Series Models Using Simulated Vector Autoregressions," *Journal of Applied Econometrics*, 8(S), S63–84.

Stock, J. H., and M. W. Watson (2001): "Vector Autoregressions," *Journal of Economic Perspectives*, 15(4), 101–115.

Stock, J. H., J. H. Wright, and M. Yogo (2002): "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics*, 20(4), 518–529.

Tierney, L. (1994): "Markov Chains for Exploring Posterior Distributions," *The Annals of Statistics*, 22(4), 1701–1728.

Waggoner, D., and T. Zha (2012): "Confronting model misspecification in macroeconomics," *Journal of Econometrics*, 171(2), 167184.

Mathematical Notation

- Sequences: $Y_{1:t} = \{y_1, \ldots, y_t\}$
- Steady states: x is steady state of x_t
- Deviations: $\hat{x}_t = \ln(x_t/x)$
- Autocovariances (y_t, s_{t-h}): $\Gamma_{ys}(h)$, $\hat{\Gamma}_{yy}(h)$.
- Identity matrix: I
- forecast horizon: h
- Index for MA representation; s. NOTE s is also state.
- Variable i, shock j
- Number of shocks n_ϵ
- Matrix elements $[A]_{ij}$, $[A]_{\cdot j}$
- Indicator function: $\mathbb{I}\{a \geq b\}$
- Frequencies and fundamental frequencies: ω and ω_j
- $i = \sqrt{-1}$
- complex conjugate $\bar{z}$
- Spectral distribution function $F_{yy}(\omega)$
- Spectral density function $f_{yy}(\omega)$
- Sample periodogram $\hat{f}_{yy}(\omega)$
- Smoothed sample $\bar{f}_{yy}(\omega)$
- VAR-based spectral estimate $\hat{f}^V_{yy}(\omega)$
- Forecast error covariance decomposition: FEVD(i, j, h)
- Impulse response function: IRF(i, j, h)

- Dimensions: n_y, n_s, n_ϵ
- Generic parameter vector: θ
- Measurement equation: Ψ
- Transition equation: Φ

Abbreviations

- DSGE
- *iid*
- VAR
- CLT