Approximation, Perturbation, and Projection Methods in Economic Analysis

KENNETH L. JUDD*

July 24, 1995

ABSTRACT. This article examines local and global approximation methods which have been used or have potential future value in economic and econometric analysis. While these methods are familiar, they are seldom developed within a general, formal analytical framework, a fact which has hindered understanding of these techniques and limited their application. We attempt to unify this literature, showing connections which have been ignored, and pointing out potential new directions. We review the foundations of basic asymptotic, or, perturbation, methods. We discuss their applications to economic modelling and econometrics. We next discuss global approximation methods, including orthogonal polynomials, interpolation theory, shape-preserving splines, and neural networks. We present the related projection method for solving operator equations, and illustrate its application to dynamic economic analysis, dynamic games, and equilibrium with asymmetric information. Finally, we discuss how the hybrid perturbation-Galerkin method combines the complementary strengths of local approximation procedures and the projection method to produce a promising new method.

CONTENTS

1. INTRODUCTION

The key technical problem in much of economic analysis is the determination of some unknown function. Important examples include the optimal policy functions of economic agents (such as the consumption function in macroeconomics), equilibrium price functions dynamic models, equilibrium strategies in games, and inference rules and price functions in asymmetric information problems. The usual approach is to make functional form assumptions on the structural elements of a model which lead to closed-form solutions for these functions; prominent examples of this approach are the linear-quadratic competitive structures discussed in Hansen and Sargent [60], the linear-quadratic dynamic game

*Hoover Institution and National Bureau of Economic Research. This research was supported by NSF Grant SBR-9309613. The author gratefully acknowledges comments from Bo Li, Ariel Pakes, John Rust, and Andrew Solnick.

structure exposited in Kydland [82, 83], the linear risk tolerance and Gaussian returns in Merton [96], and the exponential-Gaussian structure in Grossman [54]. Unfortunately, the desire for a closed-form solution often restricts the analysis. While these special cases may suffice for some purposes, they are often inadequate for a robust analysis. Such robustness is important for both theoretical analysis, where important elements may be ignored in cases with closed-form solutions, and in empirical work where misspecification of tastes and technology can ruin an otherwise valid approach.

The alternative is to assume more general and flexible functional forms and use approximation ideas to compute functions which are "close" to the true solution. In the first section we remind the reader of a variety of theoretical and empirical problems for which these methods are useful. In the rest of the paper, we will review the two basic approaches to the approximation of functions and the approximate solution of operator equations, representing two different kinds of data and objectives, and introduce a third which combines the strengths of the first two methods. *Local approximations* take as data the value of the unknown function f and its derivatives at a point x_0 and constructs a function which matches those properties at x_0. These constructions rely on Taylor's theorem, the implicit function theorem, and singularity theory, and lead to the construction of Taylor or Padé series, or other approximations of a simple form. These methods are called *perturbation*, or *asymptotic*, methods. The basic idea of asymptotic methods is to formulate a general problem, find a particular case which has a known solution, and then use that particular case and its solution as a starting point for computing approximate solutions to "nearby" problems. These methods are widely used in mathematical physics, particularly in quantum mechanics and general relativity theory, with much success. While economists have often used special versions of perturbation and asymptotic techniques, such as linearizing around a steady state, they often provide little formal justifications for their procedures, and sometimes proceed in an ad hoc and potentially invalid fashions. This has lead to some confusion as to the differences among various procedures. This is plausibly one reason why economists have generally not exploited the full range and power of these approximation techniques.

We will give simple examples of the perturbation methods and indicate the more substantive uses which have appeared in the economics literature. These applications include theoretical analyses of sunspot equilibria as well as quantitative analyses of economic policies and business cycles. We will interpret the phrase "computational economics" broadly in this chapter. The perturbation analysis work which theorists have done has been viewed as pure theory, and the authors made no apparent use of a computer. However, much of this work is really the outcome of algebraic manipulations which could be automated by symbolic mathematics software, such as Mathematica, Maple, or Macsyma. We take the view that in the future much of this type of theoretical analysis will be done by computer software, and is an interesting new avenue for computational economics. This literature is included here also because the linear approximations which these authors compute do have value as numerical approximations, and it is instructive to compare these methods with other "linear approximation" methods used in economics. Furthermore, these linear approximations are just the first step in higher–order Taylor series expansions which themselves may have substantial numerical value, even though this fact is generally not utilized in either the theoretical or applied literatures.

The other approaches to approximation are more global in nature. *L^p approximation* takes a given function f and finds a "nice" function g which is "close to" f in the sense of some L^p norm. To compute an L^p approximation of f, one ideally needs the entire

function, whereas we generally have information about f at only a finite number of values. *Interpolation* is any procedure which finds a "nice" function which exactly fits a finite set of prescribed conditions. *Regression* is similar to L^p approximation in that a some L^p norm is minimized, an L^2 norm in the case of least squares and L^∞ in the case of minimum absolute deviation. Regression also lies between L^p approximation and interpolation in that it uses n points of data to produce an approximation with $m < n$ free parameters which "nearly" satisfies the data. These approximation methods form the basis for *projection* methods, also known as *weighted residual* methods, for solving functional equations. Projection methods have been increasingly used in the physical sciences over the past twenty years. They have been used to solve various economic problems, ranging from dynamic growth models, dynamic games, and asset market equilibria with incomplete information.

Both perturbation and L^p approximation methods are important because of the increasing role of computation in economic analysis. Many computational economists eschew sophisticated approximation techniques, believing that simple methods of approximation combined with supercomputer technology will solve any problem they might have. This is not the attitude taken in other computationally intensive fields. In fact, an examination of the numerical analysis literature shows that over the past fifty years advances in numerical analysis have improved algorithm speed as much as hardware advances. Rice [108] presents a formal and substantive discussion of this issue for the problem of solving two- and three-dimensional elliptic partial differential equations, a class of numerical problems which arise naturally in continuous-time stochastic economic modelling. He argues that we were able to solve these problems 4 million to 50 billion times faster in 1978 than in 1945, of which a factor of 2,000 to 25 million can be attributed to software improvements, and a factor of 2,000 to hardware improvements. One reason for this improvement has been the application of the basic approximation ideas we present below. It is clear from examination of the mathematical and economic literature that even a modest application of modern approximation techniques can substantially improve the efficiency of most computational methods in economics. The objective of this review is to be retrospective and review actual applications, but also to be prospective and indicate where a more intensive use of well-known mathematical techniques can expand the range and quality of these applications in economics.

After discussing perturbation and projection methods, we move to a third approach to approximation which combines perturbation and projection methods. The perturbation and projection methods of solution differ substantially in their focus and procedures. However, we shall see that their strengths and weaknesses are complementary. This complementariness implies that a combined analysis using both methods will allow economists to analyze many economic problems in a robust and reliable fashion. This combined method is called the *hybrid perturbation-Galerkin* procedure. We will illustrate its advantages and potential in a simple example.

2. The Uses of Approximation Ideas: An Overview

Economic modelling problems have used a variety of approximation methods. In dynamic programming problems, one wants to solve out for the value function and the corresponding policy rule. The policy rule is then used for empirical analysis of the data since it determines the relationships among observable variables. The general approach is to find primitives – tastes and technology – which would generate observable processes whose statistical processes match the observed processes. A key step in this is computing the

behavioral rules which arise from the primitives. The closed-form approach[1] to this problem is exemplified in Sargent's[116] analysis of dynamic labor demand. However, the linear-quadratic approach has limitations. Rust[111] exemplifies the alternative approach where one assumes arbitrary tastes and technology and approximately solves the dynamic programming problem of the agents and for likelihood models for the data. However, Rust uses the very conservative discrete–state approximation method which is reliable but slow. The approximation ideas we discuss below have been successful in solving many dynamic programming problems which are more general than the linear–quadratic case but with substantially greater efficiency than the discrete–space approximation method. These solutions could also be used in maximum likelihood econometric procedures where such an increase in speed would be important.

The approximation ideas we discuss below have also been used in rational expectations equilibrium analysis. Closed-form solutions of rational expectations models are rare; agricultural economists realized the futility of this back in 1958 with Gustafson's[56] work on optimal grain stockpiling. A critical aspect of that problem is the nonnegativity constraint on grain stockpiles. This constraint leads to kinks in the storage rules and price functions. Gustafson used piecewise linear functions to approximate the relation between current price and the current total grain stock. Williams and Wright[123, 124, 125] extended the Gustafson analysis to include elastic supply. An important innovation in their solution was their observation that the conditional expectation of the *future* grain price is a smooth function of the *current* state of the market, and that this conditional expectation function characterizes equilibrium. This observation suggests that equilibrium can be approximated by low-order polynomial approximation of the conditional expectation function which characterizes equilibrium. This leads to a considerable improvement in efficiency over the alternative of using discrete-state or piecewise linear approximations of the current price law. Helmburger and Miranda[98] also use this approximation idea to solve equilibrium. More recently, Christiano and Fisher[32] use the same idea to model general equilibrium where a nonnegativity constraint on gross investment will occasionally bind. The history of computational rational expectations is just one example of where basic ideas in approximation ideas have been used to improve considerably the computational procedures. These approximation methods are also important in empirical work on structural models of commodity markets. Deaton and Laroque[43] used approximations of the rational expectations equilibrium to compute methods of moments estimates in a fully structural model of several commodity markets.

Dynamic games also have a similar dichotomy. Kydland exemplifies the closed-form approach to linear-quadratic games. In contrast, Kotlikoff, Shoven, and Spivak?? take a smooth approximation approach to solving a more general dynamic game. Miranda and Rui?? use modern approximation theory to solve nonlinear dynamic games.

The most common use of perturbation methods is the method of "linearizing around a steady state." Such linearizations tell us how a dynamical system evolves near a stable steady state. We can also use them to compute how a system reacts to shocks which move the steady state, such as tax policy or monetary policy changes. A particularly important

[1] Some may argue that the linear–quadratic model typically does not have a closed-form solution because it is generally necessary to solve a Ricatti equation, or, as in the case of dynamic games, a coupled system of Ricatti equations. While there are nontrivial problems associated with solving Ricatti equations, we currently have methods which are so reliable and accurate that the solutions are treated as if they were closed-form solutions with no computational error. Since the approximation problems are much worse when we leave the linear-quadratic paradigm, linear–quadratic modelling is, for the purposes of this review, more like closed–form modelling than the approximate solutions we will discuss.

case of this was Magill [93], who suggested that the linear approximations of stochastic growth models be used in macroeconometric analysis. Kydland and Prescott[85], and many later macroeconomists have successfully used a linear approximation computational approach to examine the empirical strength of the Real Business Cycle hypothesis. Similarly, many authors used linearization methods to analyze the impact of macroeconomic policy on dynamic equilibrium.

The key fact is that perturbation methods are just ways to take derivatives in complex problems. This implies that they have a variety of uses. For example, in maximum likelihood estimation, one must repeatedly compute derivatives of the likelihood function. Zadrozny[127] discusses how to compute such derivatives analytically in the case of linear quadratic models. For more general models, computing such derivatives is generally done numerically. However, perturbation methods could be used to solve for these derivatives analytically with considerable gains in accuracy and speed.

These are just a few examples of how approximation ideas are important in computational aspects of both theory and econometrics. We shall now discuss the formal mathematics behind these approximation ideas and illustrate their applications in simple examples.

3. The Mathematical Foundations of Regular Perturbation Methods

Some simple but powerful local approximation techniques are called *regular perturbation methods.* They are based on a few basic theorems including the well-known Taylor's theorem and the implicit function theorem for R^n as well as extensions to operators on infinite-dimensional spaces. We will first state the basic theorems which provide the foundation for regular perturbation methods in this section, and give examples of their use in the next section.

3.1. The Meaning of "Approximation". We often use the phrase "$f(x)$ approximates $g(x)$ for x near x_0", but the meaning of this phrase is seldom made clear. One trivial sense of the term is that $f(x_0) = g(x_0)$. While this is certainly a necessary condition, it is generally too weak to be a useful concept. Approximation usually means at least that $f'(x_0) = g'(x_0)$ as well. In this case, we say that "f is a first-order (or linear) approximation to g at $x = x_0$". In general, "f is an n'th order approximation of g at $x = x_0$" if and only if

$$\lim_{x \to x_0} \frac{\| f(x) - g(x) \|}{\| x - x_0 \|^n} = 0$$

3.2. Taylor Series Approximation. The most basic local approximation is described by Taylor's Theorem:

Theorem 1. *(Taylor's Theorem:) Suppose $f : R^n \to R^1$, and is C^{k+1}. Then for $x^0 \in R^n$ If $f \in C^{n+1}[a, b]$ and $x, x_0 \in [a, b]$, then*

$$\begin{aligned} f(x) \quad = \quad & f(x^0) + \sum_{i=1}^{n} \frac{\partial f}{\partial x_i}(x^0)\,(x_i - x_i^0) \\ & + \tfrac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n} \frac{\partial^2 f}{\partial x_i \partial x_j}(x^0)\,(x_i - x_i^0)\,(x_j - x_j^0) \\ & \vdots \\ & + \tfrac{1}{k!} \sum_{i_1=1}^{n} \cdots \sum_{i_k=1}^{n} \frac{\partial^k f}{\partial x_{i_1} \cdots \partial x_{i_k}}(x^0)\,(x_{i_1} - x_{i_1}^0) \cdots (x_{i_k} - x_{i_k}^0) \\ & + \mathcal{O}\left(\| x - x^0 \|^{k+1}\right) \end{aligned} \tag{1}$$

The Taylor series approximation of $f(x)$ based at x^0, (1), uses derivative information at x^0 to construct a polynomial approximation. f is *analytic on* $[a, b]$ exactly when this approximation converges to f on $[a, b]$ as k increases. Generally, this approximation is good only near x^0 and decays rapidly away from x^0.

3.3. Rational Approximation. *Padé approximation* uses the same derivative information as does a Taylor series approximation, but instead constructs a rational function to approximate f. The (m, n) Padé approximant of f at x_0 is a rational function

$$r(x) = \frac{p(x)}{q(x)} \tag{2}$$

where $p(x)$ and $q(x)$ are polynomials of degree m and n, and

$$\frac{d^k}{dx^k}\,(p - f\ q)\ (x_0), \quad k = 0, \cdots, m + n \tag{3}$$

The $m + n + 1$ derivative conditions in (3) suffice since $q(x_0)$ can be normalized to be 1. The problem of computing the coefficients of p and q is a (generally nonsingular) linear problem.

The experience is that Padé approximants are better global approximants than Taylor series approximations, that is, the error grows less rapidly as we move away from x_0. There are strong theorems confirming this for analytic functions; see Bender and Orszag [8] for an accessible treatment.

Rational approximation ideas have not been as widely used in economic analysis as Taylor series methods. Padé approximation has proved useful in econometric analysis. See Phillips [103] for a discussion of various generalizations of Padé expansions; in particular, he discusses the idea of using information at several points, not just one. Phillips also reviews applications to finite sample distribution theory. Below we will discuss another kind of application of Padé approximations..

3.4. Implicit Function Theorem. The next important tool is the Implicit Function Theorem in Euclidean spaces.

Theorem 2. *(Implicit Function Theorem:) If* $H(x, y) : R^n \times R^m \to R^m$ *is* C^1 *and* $H_y(x_0, y_0)$ *is not singular, then there is a unique function* C^0 *function* $h : R^n \to R^m$ *such that for* (x, y) *near* (x_0, y_0)

$$H(x, h(x)) = 0.$$

Furthermore, if H *is* C^k *then* h *is* C^{k-1} *and its derivatives can be computed by implicit differentiation of the identity* $H(x, h(x)) = 0$.

The Implicit Function Theorem states that h can be uniquely defined for x near zero by a relation of the form $H(x, h(x)) = 0$ whenever $H_y(0, h(0))$ is not singular. This allows us to implicitly compute the derivatives of h with respect to x as a functions of x. When we combine Taylor's theorem and the Implicit Function theorem, we have a way to compute a locally valid degree k polynomial approximation of the implicit function $h(x)$ whenever H is sufficiently differentiable. The derivative information could also be used to compute a Padé approximant.

The previous theorem applied to finite-dimensional problems. Frequently in economics we need to solve for unknown functions which are solutions to some operator equations. In these cases we need implicit function theorem for infinite dimensional spaces.

3.5. Generalizations to Function Spaces. To solve dynamic economic problems, we need generalizations of these theorems to functional spaces. It is necessary, therefore, to first introduce some terminology from functional analysis, and state a generalization of the implicit function theorem which has a straightforward computational implementation.

Suppose that X and Y are Banach spaces, i.e., normed complete vector spaces. A map $M : X^k \to Y$ is *k–linear* if it is linear in each of its k arguments. It is a *power* map if it is symmetric and k-linear, in which case it is denoted by $Mx^k \equiv M(x, x, \ldots, x)$. The norm of M is constructed from the norms on X and Y, and is defined by

$$||M|| = \sup_{||x_i||=1,\ i=1,2,\ldots,k} ||M(x_1, x_2, \ldots, x_k)||$$

For any fixed x_0 in X, consider the infinite sum in Y

$$Tx = \sum_{k=1}^{\infty} M_k (x - x_0)^k \tag{4}$$

where each of the M_k is a k-linear power map from X to Y. When the infinite series in (4) converges, T is a map from X to Y. The *majorant series* for T is

$$\sum_{k=0}^{\infty} ||M_k||\, ||x - x_0||^k$$

The important fact is that T will converge whenever its majorant series does.

Definition 3. *T is analytic at x_0 if and only if, for some neighborhood of x_0, it is defined and its majorant series converges.*

With these definitions, we can now state an analytic operator version of the Implicit Function Theorem, taken from Zeidler[128].

Theorem 4. *(Implicit Function Theorem for Analytic Operators:) Suppose that*

$$F(\epsilon, x) = \sum_{n,k=0}^{\infty} \epsilon^n M_{nk} x^k \tag{5}$$

defines an analytic operator, $F : U \subset R \times X \to Y$, where U is a neighborhood of $(0,0)$ in $R \times X$. Furthermore, assume that $F(0,0) = 0$ and that the operator $M_{01} : X \to Y$, representing the Frechet cross-partial derivative at $(0,0)$, is invertible. Consider the equation

$$F(\epsilon, x(\epsilon)) = 0 \tag{6}$$

implicitly defining a function $x(\epsilon) : R \to X$. The following are true:

1. *There is a neighborhood of $0 \in R$, V, and a positive number, $r > 0$, such that (6) has a unique solution $x(\epsilon)$ with $||x(\epsilon)|| < r$ for each $\epsilon \in V$.*
2. *The solution, $x(\epsilon)$, of (6) is analytic at $\epsilon = 0$, and, for some sequence of x_n in X, can be expressed as*

$$x(\epsilon) = \sum_{n=1}^{\infty} x_n \epsilon^n \tag{7}$$

where the coefficients x_n can be determined by substituting (7) into (6) and equating coefficients of like powers of ϵ.

3. *The radius of convergence of the power series representation in (7) is no less than that of the analytic map, $z(\epsilon) : R \to R$, defined implicitly for some neighborhood of 0 by*

$$0 = \sum_{n,k=0}^{\infty} \epsilon^n \ ||M_{nk}|| \ z(\epsilon)^k \tag{8}$$

Furthermore, for some sequence z_n of real numbers,

$$z(\epsilon) = \sum_{n=0}^{\infty} \epsilon^n \ z_n$$

represents the solution to (8) and $|z_n| > ||x_n||$.

See Zeidler[128] for a proof and discussion of this implicit function theorem. The mathematics of applying this method turns out to be elementary since the task is reduced to recursive computation of x_n terms, in term-by-term approach described above. The only requirement is to set up the problem so that it is expressed as an analytic operator with a nondegenerate radius of convergence. This theorem shows that the logic and intuition from the finite-dimensional implicit function theorem generalizes naturally and straightforwardly for analytic operators.

4. Applications of Regular Perturbation Methods to Economics

There have been many uses of local approximations in economics, implicit and explicit. The topic of comparative statics is nothing more than applications of the implicit function theorem. Comparative dynamics are technically more difficult problems, but fit into the same general framework. Recognizing these similarities will help us solve difficult problems. We will review some basic applications which have appeared and give examples of some possible future uses.

4.1. Comparative Statics: A Simple Rule of Thumb in Tax Theory. The topic of comparative statics is nothing more than applications of the implicit function theorem. One simple example of applying perturbation ideas is the impact of a tax on equilibrium. Suppose that $D(p)$ is demand at consumer price p, that $S(p)$ is supply at producer price p, and that a per unit tax of τ is applied. Then the equilibrium consumer price at tax rate τ can be expressed as the function $p(\tau)$ which is implicitly defined by $D(p(\tau)) = S(p(\tau)-\tau)$. We can expand this relation around $\tau = 0$, the tax–free equilibrium case, to study the impact of the tax on equilibrium. This analysis leads, for example, to the useful rule of thumb that the efficiency cost of a tax equals $\frac{1}{2}(\eta_D + \eta_S)\tau^2$ where η_D and η_S are the demand and supply elasticities at the $\tau = 0$ case. This quadratic approximation has been used extensively to intuitively discuss tax policies and as the formal basis for some quantitative tax analysis, as in the Barro[5] analysis of optimal tax policy.

This tax example is just one simple case where simple perturbation formulas, more commonly described as comparative statics, are useful approximations. We next examine dynamic applications of these perturbation ideas.

4.2. Comparative Dynamics: A Canonical Problem. Since it will be frequently used below, we will now describe a simple continuous-time[2] model of economic growth. Let

[2] It will be obvious that all of these methods can be applied in the same way to discrete-time models. Since there is no substantive distinction between the discrete-time and continuous-time literatures, I will discuss continuous-time and discrete-time papers together.

k be the capital stock, c the rate of consumption, and $f(k)$ the rate of output. Assume that the intertemporal utility function of the representative agent is $\int_0^\infty e^{-\rho t}u(c(t))dt$, and that the capital stock evolves according to $\dot{k} = f(k) - c$. The corresponding optimal growth problem is

$$\begin{aligned} V(k_0) &\equiv \max_{c(t)} \int_0^\infty e^{-\rho t}\, u(c)\, dt \\ \dot{k} &= f(k) - c \\ k(0) &= k_0 \end{aligned} \tag{9}$$

where $V(k)$ is the value function. Our examples will study the solution to this optimal growth problem. We will also examine the representative agent version of this problem. The competitive equilibrium will correspond to the social planning problem in the perfectly competitive, distortion free case, but not otherwise. We will also examine the equilibrium problem when taxes are present. While this model and its stochastic generalization appears to be special, it is in the same general family of dynamic optimization problems investigated by the papers of Sargent and Rust.

4.3. Perturbing Dynamic Equilibria. To illustrate the essential features of perturbation methods applied to dynamic equilibria, we apply them to study the effects of policy changes in a dynamic model of equilibrium with taxation. Brock and Turnovsky[21] shows that if we take the simple growth model behind (9) and add a tax on capital income, the resulting equilibrium solves the system of differential equations

$$\begin{aligned} \dot{c} &= \gamma(c)\; c\; (\rho - f'(k)(1-\tau)) \\ \dot{k} &= f(k) - c - g \end{aligned} \tag{10}$$

where $\gamma(c) \equiv u'(c)/(cu''(c))$ is the rate of intertemporal substitution in consumption, $\tau(t)$ is the tax on capital income at time t, $g(t)$ is government expenditure (on goods which do not affect utility) at t. The tax rates are exogenous, and c and k are the unknowns to be determined. Note that this includes the special case of $\tau = g = 0$, which is (9). The boundary conditions for (10) are the initial condition on the capital stock

$$k(0) = k_0 \tag{11}$$

and a stability condition on consumption

$$0 < \mid \lim_{t\to\infty}\; c(t) \mid < \infty \tag{12}$$

The conceptual experiment is as follows. We assume that the "old" tax policy was constant, $\tau(t) = \bar{\tau}$, and that it has been in place so long that, at $t = 0$, the economy is at the steady state corresponding to $\bar{\tau}$. Note that this also assumes that for $t < 0$, agents assumed that $\tau(t) = \bar{\tau}$ for all t, even $t > 0$. Hence, at $t = 0$, $k(0) = k^{ss}$. Suppose, however, that at $t = 0$, agents are told that future tax policy will be different. Say that they find out that the new tax rates are $\bar{\tau} + \tau(t)$, $t \geq 0$, that is $\tau(t)$ will be the change in the tax rate at time t. Similarly, they are told that the new expenditure policy is $\overline{g} + g(t)$. We also allow the possibility that the capital stock at $t = 0$ is changed by κ. The new system is

$$\begin{aligned} \dot{c} &= \gamma(c)\; c\; (\rho - f'(k)\; (1 - (\bar{\tau} + \tau(t)))) \\ \dot{k} &= f(k) - c - (\overline{g} + g(t)) \end{aligned} \tag{13}$$

together with $k(0) = k^{ss} + \kappa$, and (12). We will use perturbation methods to approximate the effects of the new policies τ and g on the dynamic paths for k and c.

We need to parameterize the new policy so that it fits the perturbation approach; that is, we need to imbed the shocked system (13) in a parameterized collection set of problems of the form $F(c,k,t,\epsilon) = 0$. We do this by defining

$$\tau(t,\epsilon) = \bar{\tau} + \epsilon\tau(t), \quad g(t,\epsilon) = \overline{g} + \epsilon g(t), \quad k(0,\epsilon) = k^{ss} + \epsilon\kappa$$

and the corresponding continuum of BVP's

$$\begin{aligned} c_t(t,\epsilon) &= \gamma(c(t,\epsilon))\ c(t,\epsilon)\ (\rho - f'(k(t,\epsilon))(1-\tau(t,\epsilon))) \\ k_t(t,\epsilon) &= f(k(t,\epsilon)) - c(t,\epsilon) - g(t,\epsilon) \\ k\,(0,\epsilon) &= k^{ss} + \epsilon\kappa \end{aligned} \tag{14}$$

plus (12).

The system (14) implicitly defines consumption and capital paths for any value of ϵ. In that way, it fits into our general implicit function framework in that we have an expression $F(c,k,t,\epsilon) = 0$ which implicitly defines the paths $c(t)$ and $k(t)$. As long as the functions involved in (14) are locally analytic, we can apply Theorem 4 above. With this apparatus in hand, we can now solve for the first-order perturbation of (14).

To solve for first–order approximations of the impact of ϵ on c and k, we differentiate (14) with respect to ϵ, evaluate the resulting differential equation at $\epsilon = 0$, and arrive at the following linear differential equation system for the unknown functions $c_\epsilon\,(t,0)$ and $k_\epsilon(t,0)$:

$$\begin{aligned} c_{\epsilon t}(t,0) &= \gamma(c^{ss})\ c^{ss}\ (-f''(k^{ss})(1-\overline{\tau})k_\epsilon(t,0) + (\rho - f'(k^{ss})(-\tau_\epsilon(t,0)))) \\ k_{\epsilon t}(t,0) &= f'(k^{ss})k_\epsilon(t,0) - c_\epsilon(t,0) - g(t) \\ k_\epsilon\,(0,0) &= \kappa \end{aligned} \tag{15}$$

plus the condition that c_ϵ and k_ϵ are both bounded. This is a linear boundary value problem with constant coefficients, which can be solved analytically. This is typical of perturbation methods: differentiate a nonlinear problem and one will arrive at a linear problem of the same type.

We then solve for $c_\epsilon(t,0)$ and $k_\epsilon(t,0)$ from(15). The result will allow us to compute a linear approximation for $c(t,1)$ and $k(t,1)$, the consumption and capital paths under the tax and spending changes; they are

$$\begin{aligned} c(t,1) &\cong f(k^{ss}) - \overline{g} + c_\epsilon(t,0) \\ k(t,1) &\cong k^{ss} + k_\epsilon(t,0) \end{aligned}$$

One can also compute the derivative of any dynamic quantity, such as lifetime utility and tax revenue, with respect to ϵ, thereby computing the marginal change in the consumption and capital path per dollar of extra revenue, per util of extra utility, or relative to any other quantity.

The resulting solutions can be very informative. For example, the initial shock to net investment (denoted by the derivative of $I \equiv f\,(k) - c - g$ with respect to ϵ at $t = 0$) is

$$I_\epsilon(0) = -\frac{\gamma\ c\ \rho}{1-\bar{\tau}}\ T(\mu) + (f'(k^{ss}) - \mu)\kappa + \mu G(\mu) - g(0) \tag{16}$$

where

$$\mu = \frac{\rho}{2(1-\bar{\tau})}\left(1+\sqrt{1+\frac{4\gamma(1-\bar{\tau})\theta_L\theta_c}{\sigma\theta_K}}\right) \tag{17}$$

is the positive eigenvalue of the linearized system (15), θ_K is capital's share of income, θ_L is labor's share, and θ_c is the steady state share of output which goes to consumption. $G(s)$ and $T(s)$ are the Laplace transforms[3] of the policy perturbations $g(t)$ and $\tau(t)$.

Perturbation methods yield algebraic formulas for quantities of interest. For example, the formula (16) tells us many things. First, future tax increases reduce investment. However, their effect is proportional to $T(\mu)$, which is essentially the average tax increase discounted at the positive eigenvalue, μ. From (17) it is clear that μ exceeds $f'(k)$, the marginal product of capital and ρ, the after-tax return. Hence, future tax increases are heavily discounted when determining their impact on current investment. Second, government spending has an ambiguous impact on investment – current government spending depresses investment and future spending increases investment, but again the future impact is discounted at rate μ. Third, since investment and output are related, we also know the initial impact of this policy shock on output. For example, if a future tax increase causes current investment to fall, then output in the future will also fall. Note that these shocks could be nonconstant, allowing us to consider partially anticipated shocks. These simple calculations address basic issues in macroeconomics.

Fourth, the presence of κ in (14) allows us to use the same approach to compute the effect of changes in the initial capital stock on consumption. The effect is intuitive: an increase in the capital stock of κ will increase output by $\kappa f'(k^{ss}) = \kappa\rho/(1-\overline{\tau})$ but will increase consumption by $\mu\kappa$, with (17) indicating that the increase in consumption is greater. Therefore, this procedure also tells us that the slope of the equilibrium policy function for consumption is $\rho/(1-\overline{\tau}) - \mu$.

We can also use this method to approximate solutions to the optimal growth model. We chose the tax example to make clear that the presence of a social planning equivalent plays no role in this procedure. However, if taxes and government spending are zero, then the problem reduces to the social planner's optimal growth problem. For example, the presence of the parameter κ in (14) means that the linear approximation to the consumption policy function near the steady state is $c^{ss} + (\rho - \mu)k$.

4.4. The Stable Manifold Theorem and Applications to Economic Theory. The analysis above is just a simple example of what perturbation analysis can do. Extending this type of analysis to several states is important in economics. These additional states will arise when we include heterogeneous capital or heterogeneous agents to our model. In this section we review the stable manifold theorem[4], which is the general statement of the linear approximation theory in dynamical systems, and its applications to economics. However, we will also note that we can compute approximations which go beyond those derived from the stable manifold theorem.

Multidimensional Dynamics. The methods used above can be extended to the case of several state variables by applying basic linear algebra and differential equation

[3] If $f(t) : R^1 \to R^n$, then the Laplace transform of $f(t)$ is $L\{f\} : R^1 \to R^n$, where $L\{f\}(s) \equiv \int_0^\infty e^{-st} f(t)dt$.

[4] We shall just discuss the procedure which is justified by the stable manifold theorem. An interested reader can find a formal statement of the stable manifold theorem in Coddington and Levinson[35].

theory. The most used mathematical theorem in this regard is the stable manifold theorem. Suppose we have a dynamic system

$$\dot{Z} = g(Z) \tag{18}$$

with a stationary point at Z^*; that is, $g(Z^*) = 0$. Then the local behavior of (18) for Z near Z^* is linearly approximated by the linear system

$$\dot{z} = A\ z \tag{19}$$

where $A = g_Z(Z^*)$ and $z \equiv Z - Z^*$. The solution to (19) is $z(t) = e^{At}\ z_0$[5]. The stable manifold theorem essentially says that the local behavior of (18) near Z^* is approximated with first-order accuracy by the local behavior of (19). In particular, if the linear system (19) has a k-dimensional stable space near Z^*, then (18) has a k-dimensional stable manifold[6] near Z^*.

This is a common situation in dynamic growth models, with and without distortions. Let $Z = \binom{X}{Y}$ where X is a list of predetermined variables and Y is a list of free variables; we use here the terminology of linear rational expectations models, as in Blanchard and Kahn[12], for example. The predetermined variables are the state variables, such as the distribution of the capital stock across sectors, or the distribution of wealth. The free variables are the decision variables, such as consumption and labor supply, and prices, all of which are endogenous at each moment. Suppose that there is a stationary point at $Z^* = \binom{X^*}{Y^*}$. Then the local behavior of the system is linearly approximated by (19) and the solution is $z(t) = e^{At}\ \binom{x_0}{y_0}$, where $x \equiv X - X^*$, $y \equiv Y - Y^*$, and y_0 is chosen to keep $z(t)$ bounded asymptotically. Let $\mathcal{Y}(x_0)$ be the set of all possible values for the free variables which together with the predetermined variables being equal to x_0 will imply a bounded path for $z(t)$. $\mathcal{Y}(x_0)$ may be a single value or a set of values.

In many economic models, $\mathcal{Y}(x_0)$ is a single-valued function which generates much valuable information, such as the dependence of prices, output, labor supply, and consumption on the state variables. As in the one-dimensional case, in general, they will allow one to compute linear approximations to the multidimensional equilibrium decision rules, even when the equilibrium cannot be reduced to a social planning problem. This procedure (which is equally valid for continuous-time and discrete-time systems) for computing a linear approximation is well-known; it is presented, for example, in detail in Chapter 6 of Stokey and Lucas[119]. Anderson[2] presents computer programs in Mathematica for solving such problems in discrete-time.

Comparative Dynamics. The general theory of such perturbations for optimal control problems has been worked out in a variety of papers. Oniki[101], and Araujo and Scheinkman[3] proved that optimal paths were differentiable with respect to parameters. Treadway[121] and Mortenson[100] used a heuristic approach to derive explicit formulas for local approximations near steady states. Lucas[92] and Otani[102] provided approximation formulas and formal justifications for them, the latter for the general optimal control problem. Caputo[25, 24] derived Slutsky–like expressions for comparative dynamics problems, and Lafrance and Barney[86] extended the analysis to the case of nondifferentiable constraints.

[5] For discrete time systems, $Z_{t+1} = g(Z_t)$, $Z^* = g(Z^*)$, $z_{t+1} = Az_t$, and $z_t = A^t z_0$.

[6] A stable manifold is a manifold, M, such that if $z(t_0)$ is in M than $Z(t)$ is in M for $t > t_0$ and converges to Z^*.

While the tax example in (15) above was quite simple, the robustness of the method to dynamic equilibrium analysis is obvious. This approach has been used to analyze many questions in dynamic economic policy. One can add labor supply, and other tax instruments. Judd [66, 68, 67] used this method to calculate the marginal efficiency cost of various tax innovations, and related impulse responses to tax changes for several macroeconomic variables. Laitner [87, 88, 89, 91] has written a series of papers on comparative dynamics, and applying them to difficult problems in dynamic tax incidence. His work includes overlapping generations applications of perturbation methods and large-dimension applications of the linearization procedure. Bovenberg [16, 18, 17] has used these methods to analyze international economic questions. He has computed the impact of taxation on capital flows, trade patterns, and terms of trade in dynamic models of international trade. All of these authors[7] use linearizations around the steady state to compute quantitative estimates of the impact of policy shocks. It is clear that these procedures can be used to analyze models with imperfect competition and externalities as well.

The linearization procedures appear to be much faster than alternative numerical methods, such as shooting. The disadvantage is that linearization procedures can produce only the first-order effects, and may miss higher–order effects. We next turn to that issue.

Higher-order Approximations. The stable manifold theorem calculation yields just linear approximations. However, proceeding as we did above, one could also compute second order approximations. This is typically not done, but there is no theoretical difficulty. In fact, when we compute the second differential of (14) one finds that the differential equations for $c_{\epsilon\epsilon}(t)$ and $k_{\epsilon\epsilon}(t)$ are the same as the differential equations for $c_\epsilon(t)$ and $k_\epsilon(t)$ in (15) except for different forcing terms. More specifically, if we write (15) in the form $x_t = Ax + \varphi(t)$, where $x = (c_\epsilon, k_\epsilon)$, then the corresponding equation for $c_{\epsilon\epsilon}(t)$ and $k_{\epsilon\epsilon}(t)$ has the same form except for the $\varphi(t)$ term. Since the difficult part of solving any linear differential equation lies in dealing with the linear operator A, we see that solving for $c_{\epsilon\epsilon}(t)$ and $k_{\epsilon\epsilon}(t)$ is essentially the same as solving for $c_\epsilon(t)$ and $k_\epsilon(t)$. More generally, the methods used in Bensoussan [10] presents the mathematical foundations for these methods in the finite-horizon case. In many models, these higher–order terms will be as easy to compute as the first–order effects. By adding a few higher–order terms to the linear term, one will end up with an accurate procedure far faster than standard differential equation solution methods. Below we will return to the problem of higher–order approximations in recursive equilibrium contexts.

Determinacy of Perfect Foresight Equilibria. The discussion above presumed that $\mathcal{Y}(x_0)$ is a single–valued function. There are many interesting cases where $\mathcal{Y}(x_0)$ is a correspondence, indicating that there are many choices for y_0 which satisfy the boundedness conditions. In fact, when there are too few unstable eigenvalues of the Jacobian $g_Z(Z^*)$, that is, the number of stable eigenvalues exceeds the number of predetermined variables, then $\mathcal{Y}(x_0)$ is a linear space. This implies that we have indeterminacy, that is, there is a linear continuum of prices and/or allocations which are consistent with equilibrium. They have proven useful for qualitatively analyzing many issues in dynamic general equilibrium. Kehoe and Levine[79] used this approach to study indeterminacy in infinite-horizon economic models. This, and many other papers, show that indeterminacy is possible in robust examples, and that the dimension of the indeterminacy can be large. Local determinacy of equilibrium is an important example where a key qualitative property of a model can be determined by straightforward computation.

[7] This is by no means a complete list of such analyses.

4.5. Perturbing Functional Equations from Recursive Equilibrium Analyses. A large variety of economic problems can be reduced to various kinds of functional equations, some more complex than the simple ordinary differential equations in time as in the example above. Stochastic models, in particular, do not generally reduce to such equations. In this section we shall take a functional approach to a simple growth model to illustrate the general applicability of perturbation methods to those functional equations arising from dynamic programming and recursive equilibrium.

Stationary, Deterministic Growth. We will first look at a single-sector, single good, continuous-time optimal growth problem, (9). The Bellman equation defining $V(k)$ is

$$\rho V(k) = \max_c \; u(c) + V'(k)(f(k) - c). \tag{20}$$

By the concavity of u and f, at each k there is a unique optimal choice of c, which satisfies the first order condition $u'(c) = V'(k)$. We will let $C(k)$, the policy function, denote that choice. (20) implies a differential equation for $C(k)$:

$$u''(C(k))\,C'(k)(f - C(k)) + u'(C(k))(f'(k) - \rho) = 0 \tag{21}$$

At the steady state, k^{ss}, $f(k^{ss}) = C(k^{ss})$, which, when substituted into (21) implies the condition $\rho = f'(k^{ss})$ which determines k^{ss}.

Our goal is to compute the Taylor series expansion of the policy function around the steady state. Specifically, we want to compute the coefficients of

$$C(k) \doteq C(k^{ss}) + C'(k^{ss})(k - k^{ss}) + C''(k^{ss})(k - k^{ss})^2/2 + ... \tag{22}$$

We have so far computed k^{ss}, $C(k^{ss})$, and $f'(k^{ss})$. We next move to $C'(k^{ss})$. At this point we must assume that $C(k)$ is C^∞. This assumption is clearly excessive, but not unrealistic if we also assume that $u(c)$ and $f(k)$ are also C^∞. In fact, Santos and Vila [115] shows that if u and f are C^k then the policy function is C^{k-2} near any stable steady state.

Differentiating (21) with respect to k yields[8]

$$0 = u'''C'C'(f - C) + u''C''(f - C) + u''C'(f' - C') + u''C'(f' - \rho) + u'f'' \tag{23}$$

which holds at each k and at the steady state, k^{ss}, reduces to

$$0 = -u''(C')^2 + u''C'f' + u'f'' \tag{24}$$

Hence $C'(k^{ss})$ must solve the quadratic equation (24), implying

$$C' = \frac{u''f' \pm \sqrt{(u''f')^2 + 4u''u'f''}}{2u''} \tag{25}$$

where all derivatives are evaluated at the steady-state levels for the capital stock and consumption. Since u and f are increasing and concave, (25) has two real solutions of opposite signs. The quadratic equation (24) is implied by the first-order conditions of the problem. We pick the positive root since only that solution is consistent with the second-order conditions.

[8] We drop arguments when they can be understood from context.

To demonstrate the ease with which higher-order terms can be calculated, we next $C''(k^{ss})$. Differentiating (23) with respect to k and imposing the steady state conditions yields an equation *linear* in the unknown $C''(k^{ss})$. Therefore, solving for $C''(k^{ss})$ is *easier* than solving for $C'(k^{ss})$. In fact, the solution for $C''(k^{ss})$ is

$$C''(k^{ss}) = \frac{2(\rho - C')u'''C'C' + 3u''C'f'' + u'f'''}{u''(3C' - 2\rho)}$$

where all functions are evaluated at k^{ss}. Note that the solution for $C''(k^{ss})$ involves $C'(k^{ss})$. The critical simplifying feature is that once we have solved the quadratic equation for $C'(k^{ss})$, we have a linear equation for $C''(k^{ss})$. Similarly, continued differentiation of (21) shows that every other derivative of C at k^{ss} can be defined linearly in terms of the steady–state derivatives of u, f, and lower order derivatives.

Judd and Guu [75] present Mathematica programs which compute arbitrary order Taylor and Padé expansions based on the derivatives of C at the steady state. Judd [74] shows that the 100 degree polynomial approximation to C is easily computed via a recursive formula. Table 1 displays the results for a variety of approximations. The assumptions are that $u(c) = c^{(1+\gamma)}/(1+\gamma)$ and $f(k) = \rho k^{\alpha}/\rho$ with $\rho = .04$, $\gamma = -2$, and $\alpha = .25$. To evaluate the quality of the approximations, we compute a normalized, unit–free version of the problem, (21), which is

$$E(k) \equiv \frac{u''(C(k))\, C'(k)(f - C(k)) + u'(C(k))(f'(k) - \rho)}{\rho\, u'(C(k^{ss}))} \tag{26}$$

We display the values of $E(k)$ for various values of k degree of approximation, and type of approximation. The notation $a(-n)$ denotes $a \times 10^{-n}$. The theoretical properties of the Taylor and Padé approximations are displayed in this example. As the degree of approximation increases, both approximations improve at all capital stocks in $[0, 2]$. Outside of $[0, 2]$, the Taylor approximation is poor and getting worse; however, the Padé approximation is doing very well even at $k = 3$ when $n = 15$.

Table 1: Euler Equation Errors

k	$n = 6$:		$n = 10$:		$n = 15$:	
	Taylor	Padé	Taylor	Padé	Taylor	Padé
0.1	9.7(−1)	2.7(−1)	5.2(−1)	3.0(−2)	2.6(−1)	1.5(−3)
0.3	6.3(−2)	5.0(−3)	1.2(−2)	5.3(−5)	1.6(−3)	1.3(−5)
0.6	6.2(−4)	1.5(−5)	1.2(−5)	5.5(−9)	1.0(−7)	6.3(−8)
0.8	3.6(−6)	4.7(−8)	4.4(−9)	1.5(−12)	1.2(−12)	7.8(−9)
1.0	0(0)	6.3(−16)	0(0)	6.3(−16)	0(0)	0(0)
1.3	3.6(−5)	1.5(−7)	2.3(−7)	3.8(−12)	4.6(−10)	7.9(−10)
1.6	3.7(−3)	8.7(−6)	3.7(−4)	2.2(−9)	2.4(−5)	1.4(−9)
2.0	1.0(−1)	1.3(−4)	7.9(−2)	1.5(−7)	6.8(−2)	3.1(−9)
2.5	9.6(−1)	8.7(−4)	7.9(−1)	3.0(−6)	1.7(2)	7.1(−9)
3.0	4.3(1)	3.0(−3)	1.3(3)	2.0(−5)	7.1(5)	3.7(−8)

Just because the Euler equation error is small does not imply that the approximation is close to the true solution. We make two points. First, in this case, we can check for

accuracy; in this case, we do find that the Euler equation error is a good indicator of accuracy. Second, if the Euler equation errors are small than the associated decision rule is one in which the agents are making decisions which are nearly optimal in the sense that the gain from doing the exactly optimal action improves the agent's welfare slightly. Since computation is costly for economic agents, we can only expect them to follow rules which are nearly optimal, and the appropriate sense of nearly optimal is not the distance from their decision and the optimal decision but the value to the agent of determining and taking the optimal action.

Non-Steady State Perturbations. The examples above computes a Taylor series for $C(k)$ around a particular capital stock, the steady state. There are other formulations which can also produce useful approximations. Recall that perturbation methods begin with a soluble case out of a continuum of cases, and uses differentiation to produce an approximation based on the soluble case. Instead of constructing an approximation based on knowing the value of C at some point, we can begin with a case where we know the entire solution and use that case to construct approximations. An example of this alternative is the continuum of problems

$$0 = C'(k,\epsilon)\left(f(k,\epsilon) - C(k,\epsilon)\right) + \gamma C(k,\epsilon)\ \left(\rho - f'(k,\epsilon)\right) \equiv F(k,\epsilon) \tag{27}$$

where γ is the constant relative risk aversion parameter, and

$$f(k,\epsilon) = (1-\epsilon)\rho k + \epsilon k^{\alpha}\rho/\alpha$$

At $\epsilon = 0$, we have a linear production function with a marginal product of capital equal to ρ, the pure rate of time preference; in this degenerate case, the solution is $C(k,\epsilon) = \rho k$, that is, consumption equals output. At all positive values for ϵ, the production function is concave and the unique steady state is $k = 1$. Suppose that we are really interested in the $\epsilon = 1$ case where f is the standard Cobb-Douglas production function.

The first perturbation of (27) implies that for all k and ϵ,

$$0 = C_{k\epsilon}(f - C) + C_k(f_\epsilon - C_\epsilon) + \gamma C_\epsilon\ (\rho - f_k) + \gamma C\ (-f_{k\epsilon})$$

which at $\epsilon = 0$ and $C = \rho k$ reduces to

$$0 = C_k(f_\epsilon - C_\epsilon) + \gamma C\ (-f_{k\epsilon})$$

and implies the solution

$$C_\epsilon(k,0) = k^{\alpha}\rho(\alpha^{-1} - \gamma) + (\gamma - \rho)k$$

Continued differentiation will yield more terms which can be use in a Taylor series approximation for the Cobb-Douglas production function ($\epsilon = 1$) case of the form

$$C(k,1) \doteq C(k,0) + C_\epsilon(k,0) + C_{\epsilon\epsilon}(k,0)/2 + C_{\epsilon\epsilon\epsilon}(k,0)/6 + \cdots \tag{28}$$

Note that this approximation is an approximation at all k, and theory tells us that it is good only for small ϵ. To determine how good this approximation is for $C(k,1)$ we could substitute it into the Euler equation and check to see if the Euler equation errors are small. They often turn out to be acceptable, but we will see below that even if (28) does not solve (27) well, the $C_\epsilon(k,0)$, $C_{\epsilon\epsilon}(k,0)$, etc., functions can still turn out to be very useful.

Single-Sector, Stochastic Growth. We next take the deterministic model above, add uncertainty, and show how to use the approximation to the deterministic policy function around k^{ss} in the deterministic case to compute an approximate policy function in the model with a small amount of uncertainty. While the assumption of small shocks may seem limiting, it is sensible in many applications, such as macroeconomic and related financial analysis.

The stochastic problem is

$$\begin{aligned} V(k) &= \sup\ E\{\textstyle\int_0^\infty e^{-\rho t}\, u(c)\, dt\} \\ & \qquad dk = (f(k) - c) + \sqrt{2\epsilon\sigma(k)}\ dz \end{aligned} \tag{29}$$

The Bellman equation becomes

$$0 = \max_c \left[-\rho V(k) + u(c) + V_k(k)\ (f(k) - c) + \epsilon\sigma\,(k)\, V_{kk}(k)\right]$$

It is straightforward to show that $C(k)$ solves

$$0 = \alpha(k)u'''(C(k)) + \phi(k)\ u''(C(k)) + \gamma(k)\ u'(C(k)) \tag{30}$$

where

$$\begin{aligned} \alpha(k) &= \epsilon\sigma(k)\ [C'(k)]^2 \\ \phi(k) &= [f(k) - C(k) + \epsilon\sigma'(k)]\ C'(k) + \epsilon\sigma(k)\ C''(k) \\ \gamma(k) &= f'(k) - \rho \end{aligned}$$

Formally, we are again looking for the terms of the Taylor expansions of C ,

$$\begin{aligned} C(k,\epsilon) \doteq\ & C(k^{ss},0) + C_k(k^{ss},0)(k - k^{ss}) + C_\epsilon(k^{ss},0)\epsilon \\ & + C_{kk}(k^{ss},0)(k - k^{ss})^2/2 + C_{\epsilon k}(k^{ss},0)\epsilon(k - k^{ss}) \\ & + C_{\epsilon\epsilon}(k^{ss},0)\epsilon^2/2 + ... \end{aligned} \tag{31}$$

Before proceeding as before, we should note that the validity of these simple methods in this case is surprising. Note that (30) is a second order differential equation when $\epsilon \neq 0$, but that it degenerates to a first-order differential equation when $\epsilon = 0$. Changing ϵ from zero to a nonzero value is said to induce a singular perturbation in the problem because of this change of order. Normally much more subtle and sophisticated techniques must be used to use the $\epsilon = 0$ case as a basis of approximation for nonzero ϵ. The remarkable feature of stochastic control problems, proved by Fleming[48], is that this is not the case, that perturbations of ϵ, the instantaneous variance, can be analyzed as a regular perturbation in ϵ [9].

With Fleming's analysis in hand, we will now proceed. We assume that we know all the k derivatives of C at $k = k^{ss}$ and $\epsilon = 0$. This is what the previous section on deterministic problems produced. We now move to computing C_ϵ by differentiating (30) with respect to ϵ. When we impose the deterministic steady state conditions $f(k^{ss}) = C\,(k^{ss})$, $f'(k^{ss}) = \rho$, and $\epsilon = 0$, we arrive at a linear equation which implies that

[9] A more modern analysis of this problem relying on viscosity methods instead of probabilistic methods as in Fleming and Souganides[49]. Their approach is also more general, possibly including distorted economies.

$$C_\epsilon = \frac{u'''C_k^2 + C_{kk}}{u''C_k}\sigma(k) + \sigma'(k) \tag{32}$$

where all the derivatives of C are evaluated at $k = k^{ss}$ and $\epsilon = 0$. Note that the solution for C_ϵ is a function not only of the deterministic steady state value of u, u', and u'', it also depends on u''', and C_{kk}, which in turn depends on f'''. If u were quadratic, f linear, and $\sigma'(k) = 0$, then (32) shows that $C_\epsilon = 0$, as we expect from the certainty equivalence results for linear-quadratic control. Again, continued differentiation of (30) with respect to ϵ and k leads to solutions for $C_{\epsilon\epsilon}$, $C_{\epsilon k}$, $C_{k\epsilon\epsilon}$, etc. Judd and Guu [75] present Mathematica programs for computing these coefficients. They also show that the approximations are valid over a substantial range of values for ϵ and k.

Dynamic Programming. The optimal growth examples above are just special cases of dynamic programming problems. Albrecht et al. [1] showed that one could differentiate the Bellman equation with respect to an exogenous parameter. Even the higher-order aspects of the computations above can be justified. Blume, Easley, and O'Hara[14] discuss when dynamic programming solutions are smooth in the state variables. Bensoussan also provides a general treatment.

Adjustment Cost Models. The problems above were based essentially on first-order conditions. We can apply perturbation methods to other problems which are not as simple. Dixit **??** studied the dynamics of models where a controller wants to keep the state of a system close to some optimal value and incurs a fixed adjustment costs whenever he adjusts the state. This leads to (S, t, s) rules; that is, when the state moves up to S or down to s the controller incurs the adjustment cost and pushes the state to a target t. There are many models which fit this description, but they seldom have analytical solutions to the problem of determining S, s, and t in terms of structural parameters. Dixit used perturbation methods to derive algebraic formulas for S and s in terms of structural parameters. He also demonstrates that first-order approximations yield very good approximations when the adjustment cost is empirically reasonable. On the qualitative side, he makes rigorous the fact that the region of inaction, $S - s$, is quite large for small variance; more precisely, he proves a fourth–power law which states that $S - s \propto \epsilon^{1/4}$ when the adjustment cost is ϵ. This result is quite important since it says that the region of inaction is quite large relative to the cost for small costs. Dixit**??** discusses a number of applications of this result. This is an excellent example of how one can use the perturbation method to get an analytically simple rule of thumb which provides important intuition about a problem.

Stochastic Equilibrium Analysis Without Pareto Efficiency. Many equilibrium problems do not reduce to optimal control problems, such as dynamic equilibria with taxation or money. While the discussion above concerned an optimal control problem, the same methods can be used to study the behavior of an economy distorted by taxation. The basic fact is that near the deterministic steady state, the linear approximation to law of motion in the stochastic model is

$$dx = A(x - (x^{ss} - \Delta))dt + \Sigma dz \tag{33}$$

where A is the linearization of the deterministic model and Σ is the covariance matrix of the shocks to the state. In the deterministic model, x^{ss} is the target state and A "pushes" the state towards the target. This expressions shows that the linear approximation to the

stochastic model involves the same linear law of motion locally but with a new target, where the adjustment Δ arises due to certainty nonequivalence.

With this observation, Balcer and Judd[4] studied the effects of taxation in a simple capital accumulation model where (33) reduces to

$$dk = \lambda(k - (k^{ss} - \Delta))dt + \Sigma dz \tag{34}$$

where λ is the negative eigenvalue of the linearization of the dynamic system describing the taxed equilibrium (similar to (10)). Therefore, the effects of taxation on business cycle fluctuations reduce to its effect on λ. They show how the level and the composition of the effective tax rate affect important business cycle statistics.

One can also compute an equilibrium value function under distortions. If we have a tax of τ on all income but have all revenues rebated in a lump-sum fashion, there is an equilibrium value function, $V^\tau(k)$, which gives the present value of future utility if the current capital stock is k. This value function will be quite different from the dynamic programming value function, which, in this notation, is $V^0(k)$. If $C^\tau(k)$ is the consumption policy function under the tax τ, the defining equation for $V^\tau(k)$ is

$$\rho V^\tau(k) = u(C^\tau(k)) + V_k^\tau(k)\ (f(k) - C^\tau(k)) + \epsilon\sigma(k)\, V_{kk}^\tau(k) \tag{35}$$

An optimal policy chooses consumption to maximize the right-hand side, but the equilibrium policy under taxation does not. To see the difference, recall that the deterministic steady state is the k^τ which satisfies $f'(k^\tau) = \rho/(1-\tau)$ in the deterministic case. At k^τ, saving is zero and $V(k^\tau) = u(C^\tau(k))/\rho$. Differentiating (35) with respect to k yields

$$\rho V_k^\tau(k) = u' C_k^\tau + V_k^\tau(k)\ (f' - C_k^\tau) + V_{kk}^\tau(f - C^\tau) \tag{36}$$

which reduces at k^τ to

$$u' - V_k = \frac{V_k^\tau(f' - \rho)}{C_k^\tau} = \rho \frac{V_k^\tau}{C_k^\tau} \frac{\tau}{1-\tau}$$

which shows that the social marginal value of capital deviates from marginal utility of consumption when the tax rate is not zero.

This fact is important when we come to evaluate the impact of uncertainty on the equilibrium value function. Differentiating (35) at $\epsilon = 0$ and $k = k^\tau$ we find

$$\rho V_\epsilon^\tau = (u' - V_k^\tau) C_\epsilon^\tau + \sigma\, V_{kk}^\tau(k) \tag{37}$$

which implies that the true first-order approximation to V^τ around the deterministic steady state is

$$V(k,\epsilon) \doteq u(C^\tau(k,0))/\rho + (k - k^\tau) V_k^\tau(k,0) + \epsilon((u' - V_k^\tau) C_\epsilon(k,0) + \sigma\, V_{kk}^\tau(k))/\rho \tag{38}$$

shows that the impact of uncertainty on the equilibrium value function depends on the degree of certainty nonequivalence, C_ϵ, when the tax rate is nonzero, and that dependence increases with increasing taxation. This is an example of a question where certainty–equivalent methods of approximating the stochastic economy will not produce reliable, first–order accurate answers.

Multidimensional, High–Order Approximations. The examples explored above have only one dimension. These methods can be extended to multidimensional problems, yielding high-order approximations to multidimensional problems. Bensoussan[10] discusses these problems for the finite–horizon case, and Judd [74] presents these procedures for the infinite-horizon case. A nontrivial difficulty in dealing with the higher–order approximations is the messy notation associated with multivariate versions of Taylor's theorem. Judd[74] extends the Einstein tensor notation, which was introduced to drastically simplify expressions in general relativity theory, to make these higher–order approximation techniques in optimal control contexts more tractable. Judd [74], following Fleming[48], further extends the multidimensional case to include uncertainty. The basic fact is that all the higher–order terms of the Taylor series expansion, even in the stochastic multidimensional case, are solutions to linear problems once one computes the first–order term in the state variables. This indicates that the higher–order terms are easy to compute. Initial experiments indicate that they are also good approximations well beyond the steady state values. These procedures have not been exploited much, but can be obviously applied to problems in the real business cycle, finance, public finance, and dynamic general equilibrium literatures.

The other development is the work of Fleming and Souganides[49]. They derive asymptotic results for problems written in viscosity form. One advantage of these problems is that they can handle infinite–horizon problems, whereas the results described in Bensoussan are proven mostly for finite-horizon cases. While discussing viscosity, a relatively recent advance in nonlinear partial differential equations, is beyond the scope of this chapter, we should note that these methods surely cover the equations which arise in dynamic programming, and might generalize to cover equilibrium problems.

Dynamic Games. Perturbation techniques can also be used to analyze dynamic games. Because of the notational burden of a formal treatment, I will here just give the basic idea behind the perturbation approach. Suffice it to say here that we are discussing dynamic game equilibrium concepts which can be written as solutions to ordinary or partial differential equations, or some similar system of functional equations.

As with any perturbation method, we begin with a "point" (possibly in a function space) where we know the solution. In game theory, such cases do arise. For example, suppose that we have two players who each influence their own state variables, but that the payoff functions and the laws of motion are such that neither player is affected by the actions of the other. This would, for example, be the case of two differentiated duopolists where the cross-elasticity of demand is zero, and the state variable of the game is the vector of the firm's capital stocks. Then the equilibrium of such a "game" is trivial, reducing to an optimal control problem for each player. Using the techniques above, we can compute local approximations for each player's strategy around steady states of the degenerate game.

Now suppose that the payoffs and/or laws of motion are slightly perturbed so that each player now cares about the other's actions. By differentiating the functional equations which characterize equilibrium with respect to the perturbation parameter and imposing the implicit function theorem and Taylor's theorem, we will be able to compute how equilibrium is affected by the alteration.

Another kind of starting point is to specify a game with general interactions, but make some parametric assumption such that the players have no interest in the dynamics. This is the case when the interest rate is infinite. In such cases, the dynamic game reduces to

a static game and, in equilibrium, neither player expends any effort to affect the future. With this degenerate case in hand we can then compute expansions in the inverse of the interest rate to determine what happens as the firms begin to care about the future. There are two examples of papers using these methods.

Judd[69] applied Theorem 4 to a patent race model. He assumed a duopoly model where the players had two kinds of research strategies and it is necessary to complete a sequence of steps. Analytic solution of such a general problem is clearly impossible. He began by assuming that the patent race had a zero prize for the winner, which, of course, implies a Nash equilibrium of no effort. This is also equivalent to the infinite interest rate case. He then proved local existence of equilibrium as well as constructed local linear and quadratic approximations.

Budd et al.[23] contains the most complex perturbation analysis of a dynamic game. They analyzed a stochastic market share duopoly game. Specifically, current profits for each firm is a function of firm one's market share, s, which is the state of the game. Each player expends effort to increase his share, which moves stochastically. The result is a stochastic dynamic game. The two degenerate cases they use are the infinite interest rate case and the case of infinite instantaneous variance of random movements in s. In these cases the firms either don't care about future market share or essentially have no control over future market share, implying a Nash equilibrium of zero effort. They compute asymptotic expansions in the inverses of the interest rate and the disturbance variance. With these expansions they are able to examine the dynamics of competition, determining when, for example, the laggard firm will work hard to catch up, when the leading firm will work hard to keep its advantage, etc.

There have been few applications of perturbation methods to game analyses thus far, but they do indicate the potential of the method. Srikant and Basar [118] develops regular perturbation methods for a large class of dynamic games different from those examined in Judd and Budd et al. Given the general applicability of these methods, the increased interest in dynamic econometric analyses, and the difficulties of game theory computation, one suspects that these procedures will become increasingly popular.

The Macroeconomic "Linear–Quadratic Approximation". The perturbation methods described have been used to approximate a wide variety of optimal control and economic equilibrium problems, and can be used much more extensively. While many macroeconomists have also studied stochastic growth models, many have eschewed the procedures above and instead use ad hoc procedures which replace nonlinear growth models with hopefully similar linear-quadratic models. Since the latter strategy bears some similarity to perturbation methods and often uses similar terminology, we will next describe it and discuss the many differences between it and perturbation methods.

As discussed in Magill[93] and Kydland and Prescott[84, 85], one basic idea is to replace a stochastic nonlinear control problem with a "similar" linear-quadratic control problem which "approximates" the nonlinear model, and then apply linear-quadratic methods to solve the model[10]. This procedure is described precisely in McGrattan[95][11]. She takes

[10] The procedure described here is applicable only to optimal control problems, and those equilibrium problems which reduce to optimal control problems.

[11] While many have used the "linear-quadratic" method, McGrattan's is the only *precise* statement of how to apply the procedure to the *general* discrete-time multidimensional control problem which I have seen in the published literature.

the nonlinear stochastic optimal control problem

$$\begin{aligned} V(x_0) &\equiv \max_{u_t} E\left\{\sum_{t=0}^{\infty} \beta^t \pi(u,x)\right\} \\ & x_{t+1} = g(x_t, u_t, \epsilon_t) \end{aligned} \tag{39}$$

where x is a vector of state variables, u is a vector of controls, and π is concave. She solves for the steady state of the deterministic version of (39), and replaces (39) with the linear regulator problem

$$\begin{aligned} V(x_0) &\equiv \max_{u_t} E\{\sum_{t=0}^{\infty} \beta^t \left(x_t' Q x_t + u_t' R u_t + 2 x_t' W u_t\right)\} \\ & x_{t+1} = A x_t + B u_t + C \epsilon_t \end{aligned} \tag{40}$$

where $x'Qx + u'Ru + 2x'Wu$ is the second-order Taylor expansion of π, and $Ax + Bu$ is the first-order Taylor expansion of g, both taken at the deterministic steady state[12].

The linear-quadratic procedure outlined in McGrattan[95] differs from the perturbation method in its approach, objective, and results. Despite using the term "linear approximation," the objective is not to compute a locally valid Taylor series for the equilibrium behavior rules. In fact, this procedure may produce an "approximation" which differs substantially from the Taylor series produced by perturbation methods. This is immediately seen by applying it to (29): $f''(k^{ss})$ appears in the solution to $C'(k^{ss})$ in (25) but appears nowhere in (40) after we apply McGrattan's procedure to (29); therefore, the linear decision rule computed by McGrattan's method applied to (29) would not be the linear approximation of the true decision rule at the steady state, (22), even in the deterministic model. In fact, those who use this procedure generally make no claim that they are computing the linear approximation of the true decision rule.

If one were to use investment instead of consumption as the decision rule in (29) then the result from McGrattan's procedure does yield the true linear approximation in many cases (further work is needed to see how general this fact is). This does not say that McGrattan's procedure is correct. Instead it points out an undesirable sensitivity to economically inessential details in the formulation of the problem. In contrast, perturbation methods are not sensitive to such changes.

No matter how one proceeds, the approach in McGrattan, Magill, and Kydland and Prescott, makes no adjustment for variance. The approximation is a certainty equivalent approximation even though the true problem is generally not certainty equivalent. At best, this procedure computes the first two terms of (31) above but drops the third term. The result is only half of the true linear approximation at the deterministic steady state since the approximation includes the linear Taylor expansion term for the state variables but excludes all Taylor expansion terms for the variance terms. Multidimensional generalizations of the rules computed in Judd and Guu [75] have no such problems.

This intuitive way of approaching the problem can lead to some conceptual problems in thinking about approximations. The linear-quadratic intuition behind (40) says to

[12] Kydland and Prescott use a slightly different procedure. They choose linear rules which satisfy the Euler equation at a collection of points near the steady state, where the collection is determined by the variance of the shocks. In this respect, their procedure is similar to the projection method we discuss below. They comment that in the case they examine, the differences are slight, but there is no reason to believe that this is always true. Their procedure results in linear-quadratic approximations which are affected by the variance. I include their procedure in this section since their stated goal is to simplify "the determination of the equilibrium process by reducing it to solving a linear-quadratic maximization problem" .

replace a nonlinear problem with a similar linear–quadratic problem because the latter is solvable. Suppose that you wanted a higher-order approximation of the optimal decision rule. This approach suggests that the way to compute a quadratic approximate decision rule would be to take a third-order polynomial approximation of the objective around the deterministic steady state and solving exactly the resulting cubic optimal control problem. Of course, there is no exact solution in general for third-order problems, making it appear difficult to compute a quadratic approximation to the decision rule. In contrast, the perturbation methods described above show that the higher-order terms are in fact easy to compute.

Christiano adopts a different approach to the "linear–quadratic" approach. He writes down the Euler equations for the nonlinear model in the form (18), and then linearizes these equations around the steady state to create a linear system of the form (19). Two comments are in order. First, this essentially reduces the problem to a calculus of variations problem. Since this cannot be done for all optimal control problems, this approach is limited. However, it is of the form justified by the stable manifold theorem. Second, he also imposes certainty equivalence on his approximation to stochastic models. Therefore, he also ends up with only a "half-linear" approximation.

Dotsey et al.[45], Christiano[30], and McGrattan[95] have documented the quality of some implementations of the macroeconomic linear-quadratic approach. The results follow what one would expect from the perturbation analysis. The Christiano and McGrattan implementations of the linear–quadatic method do fairly well when it comes to modeling movements of quantities, but not as well with asset prices. This is expected since perturbation methods show that the linear approximation of quantity movements depend on only linear–quadratic terms whereas asset pricing movements are more likely to involve higher–order terms. In particular, the extra terms produced in Judd and Guu[75] show that the deviations from certainty equivalence depend on higher–order derivatives of the utility function. The linear–quadratic approximation also does less well as the variance of the productivity shocks increases since the linear-quadratic approach ignores the effects of the variance on the decision rules.

The linear–quadratic scheme in (40) is used to solve for equilibria which solve a social planning problem. Macroeconomists have devised complex iterative schemes to compute equilibria of distorted economies. They also revolve around linear–quadratic approximations of the individual agents' problems (see, for example, Cooley and Hansen[37]). These procedures are offered without any rigorous justification, and offer no reason why they should be used instead of the earlier linearization methods derived from standard mathematical methods. As pointed out above, the standard perturbation methods used by Laitner, Judd, Bovenburg, and described in Stokey and Lucas[119] will compute first-order valid linear approximations in nonlinear equilibrium models, and do so in a nonrecursive, hence much faster, fashion.

Furthermore, the problems with this macroeconomic approach are even greater when dealing with distorted models. These approximations also ignore the impact of variance. The point of many of these exercises is to compute the welfare effects of various policies. This requires the computation of an equilibrium value function. We saw above that the first-order approximation to such functions, (38) includes the deviation from certainty equivalence when taxes are present. Therefore, their computations of utility are not reliable. Another example of the inadequacy of the macroeconomic approach is in Chari et al.[27]. They show that the resulting "linear approximation" does poorly relative to a global nonlinear procedure. Since they do not take an explicit perturbation approach

(that is, formulate it as an application of the implicit function theorem or one of its generalizations, and compute an appropriate number of terms), this is not evidence against the use of perturbation methods, only against the ad hoc approach they use.

Linear Model Computation. The linear system approach could also be used, but has been overlooked, when it comes to the analysis of linear-quadratic models. The idea is simple: if one has a model with a deterministic steady state and globally linear equilibrium behavioral rules, then the linear rules which are locally valid near the steady state are the globally valid rules. All of the perturbation methods outlined above are direct, noniterative, methods in contrast to the complex, iterative procedures often used by economists to solve linear dynamic economic models.

5. Bifurcation Methods

Sometimes we will want to compute an approximation to an implicitly defined function at a point where the conditions of the Implicit Function theorem do not hold, in particular when $H_y(x_0, y_0)$ is singular. In some cases, there is additional structure which can be exploited by *bifurcation* methods, to which we now turn.

Suppose that $H(x, \epsilon)$ is C^2. One way to view the equation $H(x, \epsilon) = 0$ is that for each ϵ it defines a collection of x which solves the equation. We say that ϵ_0 is a *bifurcation point* if the number of solutions to $H(x, \epsilon) = 0$ changes as ϵ passes through ϵ_0. Two situations are summarized in the following theorem.

Theorem 5. *(Bifurcation Theorem) Suppose $H(x, 0) = 0$ for all x, where $H : R^2 \to R$. Furthermore, suppose that*

$$H_x(x_0, 0) = 0 = H_\epsilon(x_0, 0),\ H_{x\epsilon}(x_0, 0) \neq 0$$

for some $(x_0, 0)$. Then, if $H_{\epsilon\epsilon}(x_0, 0) \neq 0$, there is an open neighborhood $\mathcal{N}$ of $(x_0, 0)$ and a function $h(\epsilon), h(\epsilon) \neq 0$ for $\epsilon \neq 0$, such that

$$H(h(\epsilon), \epsilon) = 0 \ \ on \ \ \mathcal{N}$$

and $H(x, \epsilon)$ is locally diffeomorphic to $\epsilon(\epsilon - x)$ or $\epsilon(\epsilon + x)$. Otherwise, if $H_{\epsilon\epsilon}(x_0, 0) = 0 = H_{\epsilon\epsilon\epsilon}(x_0, 0)$, then there is an open neighborhood $\mathcal{N}$ of $(x_0, 0)$ and a function $h(\epsilon), h(\epsilon) \neq 0$ for $\epsilon \neq 0$, such that

$$H(h(\epsilon), \epsilon) = 0 \ \ on \ \ \mathcal{N}$$

and $H(x, \epsilon)$ is locally diffeomorphic to $\epsilon^3 - x\epsilon$ or $\epsilon^3 + x\epsilon$. In both cases, $(x_0, 0)$ is a bifurcation point.

It may seem that Theorem 5 has limited applicability given its low-dimension character. Fortunately, there is a procedure, the Lyapunov-Schmidt method, which is used to transform high dimension (even infinite dimension) problems into appropriate low dimension problems at which point one applies the procedures above. This greatly increases the applicability of this approach.

Theorem 5 also seems limited in that H has domain R^2. Theorem 5 generalizes to $H : R^{n+1} \to R^n$. Furthermore, the Bunch Theorem (see Zeidler) generalizes the bifurcation methods in Theorem 5 to allowing both x and ϵ to be in Banach spaces. Space limitations prevent our discussing these generalizations here, but below we will see that economic applications are obvious.

There are many kinds of bifurcations; the simple ones in Theorem 5 are referred to as the transcritical and pitchfork bifurcations. Another, more complex, bifurcation which arises naturally in economics is the Hopf Bifurcation. We present the statement in Benhabib and Nishimura[9].

Theorem 6. *(Hopf Bifurcation) Suppose that* $\dot{x}= F(x,\mu)$, $x \in G \subset R^n$, $\mu \in [-c,c] \subset R$, $F \in C^k$. *Suppose that there exists stationary solutions, that is, for* $|\mu| < c$, *there is* $\widetilde{x}(\mu)$ *such that* $F(\widetilde{x}(\mu),\mu) = 0$. *Suppose that the Jacobian* $F_x(\widetilde{x}(\mu),\mu)$ *has a parametric pair of eigenvalues which can be expressed as* $\alpha(\mu) \pm \beta(\mu)i$ *where* $\alpha(0) = 0$, $\beta(0) \neq 0$, *and* $\alpha'(0) \neq 0$. *Then there exists a family of parametric solutions* $x(t,\epsilon)$ *and* $\mu(\epsilon)$ *of* $\dot{x}\ (t,\epsilon) = F(x(t,\epsilon),\mu(\epsilon))$ *such that* $x(t,0) = \widetilde{x}(0)$ *but* $x(t,\epsilon) \neq \widetilde{x}(\mu(\epsilon))$ *for* $\epsilon \neq 0$. *Furthermore,* $\mu(\epsilon)$ *is* C^{k-1} *and the period of the cycle is* $2\pi/\left|\beta(0)\right|$.

The result stated above is just a first-order result; higher order approximations are available. There are also conditions which guarantee that the periodic solutions are stable orbits. There are further generalizations of the Bifurcation Theorem which covers cases where there are many nondegenerate branches passing through a bifurcation point. Such cases may correspond naturally to multiple equilibria in economic models. For a more complete discussion of these issues see Zeidler[128], Chow and Hale [34], and Golubitsky and Schaeffer[53].

5.1. Applications of the Hopf Bifurcation to Dynamic Economic Theory. The Hopf Bifurcation Theorem has been extensively used to study the possibility of deterministic cycles in economic models. Benhabib and Nishimura[9] explored the possibility of cycles in multisector growth models. They showed how to use the Hopf Bifurcation Theorem to check for the presence of Hopf bifurcations and offered plausible numerical examples of Hopf Bifurcations in optimal growth models. Zhang[131] presented a simplified version of this analysis and also showed how to compute expressions for the period of such cycles and how to check their stability.

The Hopf bifurcation has also been used in Industrial Organization theory. Feichtinger[46] used the Hopf bifurcation in a dynamic model of advertising to argue that cycles in advertising expenditure were quite plausible. The Hopf bifurcation has also been used to analyze general equilibrium with financial imperfections. The theme of these papers is that while the equilibrium dynamics of an economy may be stable with perfect capital markets, capital market imperfections may lead to more complex dynamics. Again, the Hopf bifurcation is used to demonstrate the existence of stable cycles. These papers include Franke[50], who investigated a Keynes-Wicksell model with adaptive expectations for inflation and found that periodic orbits were possible.

There are possibly many other applications of the Hopf bifurcation. Most dynamic analyses examine only the steady state, not its local dynamic structure. At the least, one can check the local linear structure to see if the number of stable and unstable eigenvalues is consistent with local asymptotic stability. Since stable cycles are often associated with unstable steady states, the presence of too many unstable eigenvalues should lead an analyst to check for the possibility of a nearby Hopf bifurcation. Since this checking is purely an algebraic exercise, easily done by symbolic computation methods, such checks should become standard.

5.2. Gauge Functions. The methods described above, commonly referred to as regular perturbations, compute expansions of the form $\sum_{i=1}^{\infty} a_i \epsilon^i$. There are many cases where

we will want to compute different expansions. In general, a system of *gauge functions* is a sequence of functions, $\{\delta_n(\epsilon)\}_{n=1}^{\infty}$, such that

$$\lim_{\epsilon \to 0} \frac{\delta_{n+1}(\epsilon)}{\delta_n(\epsilon)} = 0$$

An *asymptotic expansion* of $f(x)$ near $x = 0$ is denoted

$$f(x) \sim f(0) + \sum_{i=1}^{\infty} a_i \, \delta_i(x)$$

where, for each n,

$$\lim_{x \to 0} \frac{f(x) - (f(0) + \sum_{i=1}^{n} a_i \, \delta_i(x))}{\delta_n(x)} = 0 \, .$$

In regular perturbations, the sequence of gauge functions is $\delta_k(\epsilon) = \epsilon^k$. Another example of a gauge system is $\delta_k(\epsilon) = \epsilon^{k/2}$. In many problems, part of the problem is in determining the correct gauge system. The next sections present examples of this more general problem.

5.3. Bifurcation Applications to Stochastic Modelling. Whereas the Hopf bifurcation is useful in analyzing deterministic systems, the simpler pitchfork and transcritical bifurcations examined above in Theorem 5 can be used to study stochastic problems. In this section, we will illustrate this with two examples. First we will discuss the details of a simple portfolio problem. Second, we will discuss a much more sophisticated application.

Portfolio Choices with Small Risks. Suppose that an investor has W in wealth to invest in two assets. The safe asset yields R per dollar invested and the risky asset yields $\tilde{Z}$ per dollar invested. If a proportion ω of his wealth is invested in the risky asset, final wealth is $\tilde{Y} = W((1-\omega)R + \omega\tilde{Z})$. We assume that he chooses ω to maximize $E\{u(\tilde{Y})\}$ for some concave utility function $u(\cdot)$.

One way in which economists have gained insight into this problem is to approximate u with a quadratic function and solve the resulting quadratic optimization problem. It is argued that this is valid for small risks. The bifurcation approach allows us to examine this rigorously. We first create a continuum of portfolio problems by assuming

$$\tilde{Z} = R + \epsilon\tilde{z} + \epsilon^2\pi$$

At $\epsilon = 0$, $\tilde{Z}$ is degenerate and equal to R. If $\pi > 0$, we model the intuitive case of risky assets paying a premium. Note that we multiply $\tilde{z}$ by ϵ and π by ϵ^2. Since the variance of ϵz is $\epsilon^2 \sigma_z^2$, this models the observation that risk premia are the same order as the variance.

Judd and Guu[77] applied the Bifurcation Theorem to this problem, producing a procedure which involves solving only linear equations and which can be used for noncompact distributions. We will briefly outline their analysis. The first-order condition for ω is

$$0 = E\{u'(R + \omega(\epsilon z + \epsilon^2\pi))\,(z + \epsilon\pi)\} \equiv G(\omega, \epsilon) \tag{41}$$

We want to analyze this problem for small ϵ. We cannot apply the implicit function theorem since $0 = G(\omega, 0)$ for all ω implying that ω is indeterminate at $\epsilon = 0$. Since we

want to solve for ω as a function of ϵ near 0, we first need to compute which of these ω values is the "correct" solution to the $\epsilon = 0$ case; specifically, we want to compute

$$\omega_0 \equiv \lim_{\epsilon \to 0} \omega(\epsilon)$$

Implicit differentiation of (41) implies

$$0 = G_\omega \omega' + G_\epsilon \tag{42}$$

Differentiating G we find

$$G_\epsilon = E\{u''(\tilde{Y})\,(\omega z + 2\omega\epsilon\pi)\,(z + \epsilon\pi) + u'(\tilde{Y})\pi\}$$

$$G_\omega = E\{u''(\tilde{Y})\,(z + \epsilon\pi)^2\epsilon\}$$

At $\epsilon = 0$, $G_\omega = 0$. $\omega'(0)$ can be well-defined in (42) only if $G_\epsilon(\omega, 0) = 0$ also. Therefore, we look for a bifurcation point, ω_0, defined by $0 = G_\epsilon(\omega_0, 0)$. At $\epsilon = 0$, this reduces to $0 = u''(R)\,\omega_0\sigma_z^2 + u'(R)\pi$, which implies

$$\omega_0 = -\frac{u'(R)}{u''(R)}\,\frac{\pi}{\sigma_z^2}$$

This is the simple portfolio rule indicating that ω is the product of risk tolerance and the risk premium per unit variance. If ω_0 is well-defined, then this must be its value. Since the conditions of the Bifurcation Theorem are satisfied at $(\omega_0, 0)$, there is a function $\omega(\epsilon)$ which goes satisfies (41) and goes through $(\omega_0, 0)$.

Note that this is not an approximation to the portfolio choice at any particular variance. Instead, ω_0 is the limiting portfolio share as the variance vanishes. Some authors treat this as an approximation to the true solution, $\omega(\epsilon)$, for small ϵ. That, however, is not the case. If we want the linear approximation of $\omega(\epsilon)$ at $(\omega_0, 0)$, we must go one more step since the linear approximation is $\omega(\epsilon) \doteq \omega(0) + \epsilon\,\omega'(0)$. To calculate $\omega'(0)$ we need to do one more round of implicit differentiation. Differentiating (42) with respect to ϵ yields $0 = G_{\omega\omega}\,\omega'\omega' + 2G_{\omega\epsilon}\,\omega' + G_\omega\omega'' + G_{\epsilon\epsilon}$. At $(\omega_0, 0)$, $G_{\epsilon\epsilon} = u'''(R)\omega_0^2\,E\{z^3\}$, $G_{\omega\omega} = 0$, $G_{\omega\epsilon} = u''(R)\,E\{z^2\}$. Therefore,

$$\omega'(0) = -\frac{1}{2}\,\frac{u'''(R)}{u''(R)}\,\frac{E\{z^3\}}{E\{z^2\}}\,\omega_0^2\ .$$

This formula tells us how the share of wealth invested in the risky asset changes as the riskiness increases, highlighting the importance of the third and second derivatives of utility and the ratio of skewness to variance. If the distribution of the risky asset is symmetric, then $E\{z^3\} = 0$, and the constant ω_0 is the linear approximation of $\omega(\epsilon)$. This is also true if $u'''(R) = 0$, such as in the quadratic utility case. However, if the utility function not quadratic and the risky return is not symmetrically distributed, then $\omega'(0) \neq 0$, and the linear approximation is a nontrivial function. Note that this says that a linear approximation to $\omega(\epsilon)$ requires a cubic approximation to the utility function and third moments of $\tilde{z}$. This fact, also demonstrated in Samuelson[114], shows how the simple approach of using only a quadratic approximation to the objective function does not produce a valid linear approximation for $\omega(\epsilon)$. The advantage of the bifurcation approach demonstrated hear is that the structure of the problem indicates exactly what information is needed.

Samuelson [114] earlier analyzed this problem in a lesss formal fashion. Also, his formal analysis was limited to $\tilde{Z}$ with compact distributions, that is, random variables whose support goes to a point as ϵ goes to zero, a detail which substantially limits practical interest. The perturbation arguments used above make no such restriction. While the Samuelson method worked in this example, using Theorem 5 allows us to proceed in a more general fashion, and provides the necessary formal justification for these calculations.

One example of where the true problem has a bifurcation structure and standard linear approximation procedure is unacceptable is Huffman [63]. Huffman examined an overlapping generations model of capital accumulation, and tried to examine the impact of a business cycle shock on asset trading. He computed the deterministic steady state and computed the impulse function for individual wealth and asset trading arising from unanticipated shocks to endowments and output. Since this was all done in an otherwise deterministic, perfect foresight model, the tacit assumption, appropriate for such models, was that equity was the only traded asset. However, Huffman then interpreted the impulse functions from the deterministic model as impulse response functions for a stochastic model, where all agents know that these shocks occur frequently. In the stochastic case, there would also be demand for trade in bonds as well as equity, and business cycle shocks would generate disturbances to bond holdings as well as to equity holdings. This was all ignored by Huffman, who implicitly assumed that even in the stochastic model the only asset was equity. Such a capital market imperfection will have important impacts on the predicted asset trading, and is not an appropriate assumption for the U.S. capital market. This is not an appropriate approximation assumption; just because one of the deterministic equilibria has no bond trading does not mean that the absence of bonds is an appropriate approximation for the stochastic model, even one with shocks with small variance. The importance of including bond trading into the analysis depends on the question being investigated. It is likely that the welfare loss is small when the variance is small. However, including bond trading could have substantial impact on the volume of trade in various assets.

Bifurcation and Sunspots. A particularly sophisticated application of bifurcation techniques appeared in Chiappori et al.[29]. They analyzed the existence of stationary sunspot equilibria near steady states of overlapping generations model of arbitrary dimension. They show that when a steady state has indeterminate local deterministic dynamics, i.e., there exists a continuum of perfect foresight paths converging to the steady state, then there exists a continuum of sunspot equilibria which have support in neighborhoods of the steady state. They also are able to determine the possible qualitative character of the sunspot equilibria. This paper displays a very sophisticated and rich application of the ideas behind Theorem 5. These methods will also allow researchers to assess quantitative aspects of sunspot equilibria.

6. Asymptotic Expansions of Integrals

In economic and econometric problems, integrals frequently take the form

$$I(\lambda) \equiv \int_D e^{-\lambda g(x)} \, f(x) \, dx \tag{43}$$

where λ is a large parameter. Simply differentiating (43) with respect to λ at $\lambda = \infty$ will not work here. Laplace's method provides a useful way to approximate (43). The basic idea is that the major contribution of the integrand is at the minimum of $g(x)$. Suppose

$g(x)$ is minimized at $x = a$. For large λ, if $x \not= a$ then $e^{-\lambda g(x)} \ll e^{-\lambda g(a)}$. As long as $f(x)$ does not offset this for $|x-a| \gg 0$, $I(\lambda)$ is determined largely by the behavior of the integrand, $e^{-\lambda g(x)} f(x)$, for x near a.

The one-dimensional case is easy to state. Assume that g and f satisfy the asymptotic series

$$g(x) \sim g(a) + \sum_{i=0}^{\infty} a_i (x-a)^{i+\mu}, \quad f(x) \sim \sum_{i=0}^{\infty} b_i (x-a)^{i+\alpha-1}$$

Under modest assumptions (see Wong, [126], or Bleistein and Handelsman, [13]) if the integral $I(\lambda) = \int_a^b f(x) e^{-\lambda g(x)}\, dx$ converges absolutely for sufficiently large λ, and if g is minimized on $[a,b]$ at a, then

$$I(\lambda) \sim e^{-\lambda g(a)} \sum_{i=0}^{\infty} \Gamma\left(\frac{i+\alpha}{\mu}\right) \frac{c_i}{\lambda^{(i+\alpha)/\mu}} \tag{44}$$

where $\Gamma(\lambda) \equiv \int_0^\infty e^{-x} x^{\lambda-1}\, dx$ is the gamma function, and the c_i depend on the a_i and b_i. In particular

$$c_0 = \frac{b_0}{\mu a_0^{\alpha/\mu}}, \quad c_1 = \left(\frac{b_1}{\mu} - \frac{(\alpha+1) a_1 b_0}{\mu^2 a_0}\right) a_0^{-(\alpha+1)/\mu}$$

To compute these coefficients and others one essentially expands the integrand in terms of λ and matches like powers[13]. One of the byproducts of this theorem is the construction of an integrand which is close to $e^{-\lambda g(x)}\ f(x)$ but also integrable. This approximation to the integrand is then integrated to produce Laplace's approximation. Note that the gauge functions of λ in (44) depend on the asymptotic expansions of f and g.

One elementary application of Laplace's method is Stirling's formula for $n!$. Recall that $n! = \Gamma(n+1)$. We would like to approximate $\Gamma(n)$ for large n. To use Laplace's method, let $x = y\lambda$; then

$$\Gamma(\lambda) = \lambda^\lambda \int_0^\infty e^{-\lambda(y - \ln y)} y^{-1}\, dy$$

The minimum of $y - \ln y$ is at $y = 1$. Break the integral into two integrals over $[0,1]$ and $[1,\infty)$, and add the two one-sided approximations to get the two-term approximation

$$\Gamma(\lambda) \sim \sqrt{2\pi}\ \lambda^{\lambda - \frac{1}{2}} e^{-\lambda} \left(1 + \frac{1}{12\lambda}\right)$$

Stirling's formula is just the one-term expansion, $n! \simeq \sqrt{2\pi}(n+1)^{n+\frac{1}{2}} e^{-(n+1)}$.

While the operating assumption in Laplace's approximation is that λ is a "large" parameter, practical use of such methods rely on what "large" means quantitatively. Fortunately, these expansions may do very well even when λ is actually small. For example, Stirling's approximation for 1! is .9595 and the two–term expansion yields .9995.

There is a multidimensional extension of Laplace's method. Suppose $D \subset R^n$, $f, g \in C^2[D]$. Suppose the minimum of $g(x)$ for $x \in D$ is achieved at x_0 in the interior of D. Then the leading term of the expansion is

$$I(\lambda) \doteq \frac{e^{-\lambda g(x_0)}}{|H|^{1/2}} \left(\frac{2\pi}{\lambda}\right)^{n/2} f(x_0)$$

[13] Bender and Orszag gives an intuitive presentation of this procedure.

where $H \equiv (g_{x_i x_j})$ is the Hessian of g. Higher order terms can also be computed. While computing higher–order terms would be very tedious to construct, symbolic languages such as Mathematica, Maple, and Macsyma, are ideally suited to do this.

6.1. Econometric Applications of Asymptotic Methods. Asymptotic methods in econometrics are essentially perturbation methods where the properties of an estimator are computed in terms of the size of the data set and the expansion is around the case of an infinite sample size. In some cases, the asymptotic problem can be handled by relatively simple procedures, such as Edgeworth expansions.

In other cases, the full power of Laplace's method is needed to compute asymptotic properties of statistics. In this case, the integral is the likelihood function, and it is written in the form (43) where the parameter λ is the sample size. Phillips [104] used Laplace's method to approximate small sample marginal densities of instrumental variables estimators. Ghysels and Lieberman [52] use Laplace's method to compute small sample biases which arise from using filtered data in dynamic regressions. Laplace's method has been more popular among statisticians; see the citations in [52]. Holly and Phillips [62] use the related saddlepoint procedure. These methods work well, but are not often used, possibly because their implementation requires much algebra. One suspects that a more intensive use of symbolic computational tools would make them more accessible.

6.2. Theoretical Applications of Laplace's Method. While the theoretical applications of asymptotic methods for evaluating integrals are few currently, they are likely to increase. Brock[19] discusses where Laplace's method is useful in evaluating statistical mechanical systems adapted to economic issues. As this modelling approach matures, it is likely that Laplace's method and related asymptotic procedures will be quite useful.

7. The Mathematics of L^p Approximations

We will often want to approximate functions over a broad range of values with relatively uniform accuracy. In this case, we turn to L^p approximations. *L^p approximations* finds a "nice" function g which is "close to" a given function f in the sense of a L^p norm. To compute an L^p approximation of f, one ideally needs the entire function, an informational requirement which is generally infeasible. *Interpolation* is any procedure which finds a "nice" function which goes through a collection of prescribed points. When using interpolation, the objective is to assure that if the data comes from a function g then the interpolant is close to g. *Regression* lies between L^2 approximation and interpolation in that the amount of data used exceeds the number of free parameters, producing an approximation which "best" fits the data. In all cases, we need to formalize the notions of "nice" and "close to."

7.1. Orthogonal Polynomials. We will next use basic vector space ideas to construct representations of functions which will lead to good approximations. Since the space of continuous functions is spanned by the polynomials, x^n, it is natural to think of the ordinary polynomials as a basis for the space of continuous functions. However, recall that good bases for vector spaces possess useful orthogonality properties. We will develop those orthogonality ideas to construct *orthogonal polynomials.*

Definition 7. *A weighting function, $w(x)$, on $[a, b]$ is any function which is positive and has a finite integral on $[a, b]$. Given a weighting function $w(x)$, we define an inner product*

on integrable functions over $[a, b]$*:*

$$\langle f, g\rangle = \int_a^b f(x)\, g(x)\, w(x)\, dx$$

The family of polynomials $\{\varphi_n(x)\}$ *are mutually orthogonal with respect to the weighting function* $w(x)$ *if and only if*

$$\langle \varphi_n, \varphi_m\rangle = 0, \quad n \neq m$$

There are several examples of orthogonal families of polynomials, each defined by a different weighting function and interval. Some common ones useful in economics are Legendre, Chebyshev, Laguerre, and Hermite polynomials. Legendre polynomials assume $w(x) = 1$ on the interval $[-1, 1]$; the n'th Legendre polynomial is

$$P_n(x) \equiv \frac{(-1)^n}{2^n n!} \cdot \frac{d^n}{dx^n}\left[(1-x^2)^n\right]$$

The Chebyshev polynomials arise from $w(x) = (1-x^2)^{-\frac{1}{2}}$ on $[-1, 1]$; the n'th Chebyshev polynomial is

$$T_n(x) \equiv \cos(n \cos^{-1} x)$$

The Chebyshev and Legendre polynomials are useful in solving problems which live on compact sets since a linear change of variables will transform in compact interval into $[-1, 1]$. The Laguerre polynomials correspond to $w(x) = e^{-x}$ on $[0, \infty)$; the n'th member is

$$L_n(x) \equiv \frac{e^x}{n!} \frac{d^n}{dx^n}\left(x^n e^{-x}\right)$$

Laguerre polynomials are useful when one needs to approximate time paths of variables in a deterministic analysis. Hermite polynomials arise from $w(x) = e^{-x^2}$ on $(-\infty, \infty)$; the n'th member is

$$H_n(x) \equiv (-1)^n e^{x^2} \frac{d^n}{dx^n}\left(e^{-x^2}\right)$$

Hermite polynomials are used to approximate functions of normal random variables.

7.2. Least-Squares Orthogonal Polynomial Approximation. Given $f(x)$ defined on $[a, b]$, one approximation concept is least-squares with respect to the weighting function $w(x)$. That is, given $f(x)$, the least-squares polynomial approximation of f with respect to weighting function $w(x)$ is the degree n polynomial which solves

$$\min_{\deg\ (p) \leq n} \int_a^b \left(f(x) - p(x)\right)^2 w(x)\, dx.$$

In this problem, the weighting function $w(x)$ indicates how we care about approximation errors as a function of x. For example, if one has no preference over where the approximation is good (in a squared-error sense) then we take $w(x) = 1$. If one cared more about the error around $x = 0$ we should choose a $w(x)$ which is larger near 0.

The connections between orthogonal polynomials and least-squares approximation are immediately apparent in solving for the coefficients of $p(x)$ in the least-squares approximation problem. If $\{\varphi_n\}_{n=1}^{\infty}$ is an orthogonal sequence with respect to $w(x)$, and we

define $\langle f, g\rangle \equiv \int_a^b f(x)\, g(x)\, w(x)\, dx$ the induced metric is $\| f \| \equiv \langle f, f\rangle$, the least-squares solution minimizes $\| f - p \|$, and can be expressed

$$p(x) = \sum_{i=0}^{n} \frac{\langle f, \varphi_i\rangle}{\langle \varphi_i, \varphi_i\rangle}\, \varphi_i(x).$$

Note the similarity between least-squares approximation and linear regression. The formula for $p(x)$ is essentially the same as regressing the function f on $n+1$ orthogonal regressors; the coefficient of the i'th "regressor," $\varphi_i(x)$, equals the "covariance" between f and the i'th "regressor" divided by the variance of the i'th regressor. This is no accident since regression is a least-squares approximation.

Chebyshev Approximation. We will next describe some of the features of Chebyshev approximation since they play an important role in many applications.

Theorem 8. *(Chebyshev Approximation Theorem) Assume $f \in C^k\,[-1, 1]$. Let*

$$C_n(x) \equiv \frac{1}{2}\, c_0 + \sum_{j=1}^{n} c_j T_j(x)$$

where

$$c_j \equiv \frac{2}{\pi} \int_{-1}^{1} \frac{f(x)\, T_j(x)\, dx}{\sqrt{1-x^2}}.$$

Then there is a b such that, for all $n \geq 2$

$$\| f - C_n \|_\infty \leq \frac{b \ln n}{n^k}$$

Hence $C_n \to f$ uniformly as $n \to \infty$. Furthermore, there is a constant c such that

$$|c_j| \leq c/j^k, \quad j \geq 1$$

This theorem has many useful aspects. First, if we compute a n-term Chebyshev approximation, we need to assess the likelihood of it being "nearly" as good as the full approximation. If the last few terms of the n-term approximation do not appear to be dropping at the j^{-k} rate indicated in the theorem, we would take this as evidence for adding more terms; if the coefficients are dropping at the indicated rate we feel more comfortable in accepting the n-term approximation. Note that, even though the construction is a least-squares approach, the convergence is uniform, a far stronger form of convergence. Since uniform approximation is a more difficult problem, we instead use Chebyshev approximation which, according to Theorem 8, will work nearly as well in the uniform norm.

7.3. Interpolation. Interpolation is any method which takes a finite set of pointwise restrictions and finds a function $f : R^n \to R^m$ satisfying those restrictions.

Lagrange Interpolation. Lagrange interpolation takes a collection of n points in R^2, (x_i, y_i), $i = 1, \cdots, n$, where the x_i are distinct, and finds a degree $n-1$ polynomial, $p(x)$, such that $y_i = p(x_i)$, $i = 1, \cdots, n$. The Lagrange formula demonstrates that there is such interpolating polynomial. Define

$$\ell_i(x) = \prod_{j \neq i} \frac{x - x_j}{x_i - x_j}$$

Note that $\ell_i(x)$ is unity at $x = x_i$ and zero at $x = x_j$ for $i \neq j$. This property implies that the polynomial

$$p(x) = \sum_{i=1}^{n} y_i \ell_i(x)$$

interpolates the data, that is, $y_i = p(x_i)$, $i = 1, \cdots, n$. Furthermore, this is the unique solution.

Hermite Interpolation. We may want to find a polynomial p which fits slope as well as level requirements. Suppose we have data

$$p(x_i) = y_i,\ p'(x_i) = y_i'\ ,\ i = 1, \cdots, n$$

where the x_i are distinct. Since we have $2n$ conditions, we are looking for at least a degree $2n-1$ polynomial which satisfies the conditions above.

We will construct the unique solution, $p(x)$. First define the functions

$$\tilde{h}_i(x) = (x - x_i)\, \ell_i(x)^2$$

$$h_i(x) = \left(1 - 2\ell_i'(x_i)\,(x - x_i)\right) \ell_i(x)^2$$

The critical facts are that h_i is a function which is zero at all x_j nodes except at x_i, where it is unity, and its derivative is zero at all x_j, and the reverse is true for $\tilde{h}_i(x)$. The unique solution to the Hermite interpolation problem is

$$p(x) = \sum_{i=1}^{n} y_i\, h_i(x) + \sum_{i=1}^{n} y_i'\, \tilde{h}_i(x)$$

7.4. Approximation Through Interpolation. Interpolation is extremely powerful since it uses a minimal amount of information to construct an approximation. It is also dangerous since the number of free parameters equal the amount of data. Furthermore, we want the approximation to be valid generally, not just at the interpolation nodes. This is not generally true for interpolation schemes. Consider the function $f(x) = \frac{1}{1+x^2}$ over the interval $[-5, 5]$. Let $p_n(x)$ be the n'th degree polynomial which agrees with f at the $n+1$ uniformly spaced (including the endpoints) nodes. Not only does p_n not converge to f, but for $|x| > 3.64$, $\limsup_{n\to\infty} |f(x) - p_n(x)| = \infty$. Therefore, for a seemingly well-behaved C^∞ function, interpolation at the uniformly spaced nodes does not improve as we use more points.

Interpolation Error. The last example may discourage one from approximating a function through interpolation. While the example does indicate that caution is necessary, with care we can reduce the likelihood of perverse behavior by interpolants. To see what we can do, we examine the general interpolation error. Recall that the Lagrange polynomial interpolating f at points x_i is $p_n(x) = \sum_{i=1}^{n} f(x_i)\ell_i(x)$. Define

$$\Psi(x; x_1, \cdots, x_n) = \prod_{k=1}^{n} (x - x_k).$$

The following theorem provides a bound on the interpolation error of the Lagrange interpolant.

Theorem 9. *Assume $a = x_0 < x_1 < \cdots < x_n = b$. Then*

$$\sup_{x \in [a,\, b]} \mid f(x) - p_n(x) \mid \leq \parallel f^{(n+1)} \parallel_\infty (n!)^{-1} \sup_{x \in [a,\, b]} \Psi(x; x_1, \cdots, x_n) \tag{45}$$

This bound decomposes the interpolation error into three pieces. The first two are independent of any analysts choice. However, the third term depends on the choice of interpolation points. By making good choices for the x_i we can substantially affect the interpolation error.

Here we see a significant difference between the problem facing a numerical analyst and the problems of an econometrician. An econometrician must take the values of f evaluated at whatever points some data generating process provides. In contrast, in approximation problems we get to choose where to evaluate f. In general, interpolation would be a bad procedure for econometricians since there is in general no assurance that our data comes from a good choice of x's. When we can choose the points, there is some hope that we can choose them to keep down the interpolation error. Furthermore, econometricians have to deal with significant error in the observations of f, whereas in numerical contexts we evaluate f with high accuracy.

Chebyshev Interpolation. We will next determine a good collection of interpolation nodes. Note that our choice of $\{x_i\}_{i=1}^{n}$ affects only the maximum value of $\Psi(x)$, which in turn does not depend on f. So if we want to choose interpolation points so as to minimize their contribution to (45), the problem is

$$\min_{x_1, \cdots, x_n} \quad \max_{x} \quad \Pi_{k=1}^{n} (x - x_k)$$

The solution to this problem on $[-1, 1]$ is

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

which are the zeros of $T_n(x)$. Therefore, the interpolation nodes which minimize the error bound (45) are the zeros of a Chebyshev polynomial adapted to the interval. This shows that the Chebyshev interpolant is the best in terms of minimizing the worst-case error. Furthermore, it also keeps the maximum error, $\parallel f - p_n \parallel_\infty$, acceptably small, as the next theorem shows.

Theorem 10. *(Chebyshev Interpolation Theorem) Suppose $f \in C^k[a,b]$. If I_n^f is the degree n Chebyshev interpolant, then there is some d_k such that for all n*

$$\| f - I_n^f \|_\infty \leq \left(\frac{2}{\pi} \log(n+1) + 2 \right) \frac{d_k}{n^k} \| f^{(k)} \|_\infty .$$

This theorem says that the Chebyshev interpolant converges to f rapidly as we use more Chebyshev zeros. Furthermore, if f has k derivatives, then the convergence rate is $O(n^{-k}\log(n+1))$. If $f \in C^\infty$, then we have $O(n^{-k}\log(n+1))$ convergence for all k; of course, the proportionality constants, d_k, are also increasing in k. Convergence may seem to be an unremarkable property, but recall that interpolation at uniformly spaced points does not necessarily converge. Given these properties, Chebyshev methods are valuable whenever the approximated function is smooth.

7.5. Approximation Through Regression. Another way to approximate a function arbitrarily well is to use regression. In regression, one evaluates the function $f(x)$ at m points, and use the resulting evaluations to choose a parametric approximation with n parameters, $n \ll m$, which minimizes some loss function. The methods closest in spirit to the material above are the seminonparametric methods, reviewed in Hardle[?]. A key asymptotic result in the seminonparametric literature is that if m and n grow at appropriate rates, then the approximation converges to $f(x)$ as $n \to \infty$. While regression methods can be used, they are based on "random" choices of the x_i, whereas other approximation methods make efficient choices of the x_i and will generally dominate regression.

7.6. Piecewise Polynomial Interpolation. Lagrange interpolation computes a C^∞ function to interpolate the given data. An alternative is to construct a function which is only piecewise smooth. Two common schemes are Hermite polynomials and splines.

Step Function Approximation. One common approximation strategy in economics is to use step functions. Step function approximations on $[a,b]$ are generated by a basis of step functions, $\{\varphi_i : i = 1, \cdots, n\}$ where $h = \frac{a-b}{n}$ and

$$\varphi_i(x) = \begin{cases} 0, & a \leq x \leq a + (i-1)h \\ 1, & a + (i-1)h \leq x \leq a + ih \\ 0, & a + ih \leq x < b \end{cases}$$

If the interpolation data are (x_i, y_i) and $\varphi_i(x_i) = 1$, then the step function $\sum_i y_i \varphi_i(x)$ interpolates the data. To get better approximations, one increases n.

Piecewise Linear Approximation. Piecewise linear approximations take a sequence of data, (x_i, y_i), and creates a piecewise linear function which interpolates the data. If the x_i are uniformly distributed, then they are generated by a basis of *tent functions*, that is, functions of the form, for $i = 0, \cdots, n$,

$$\varphi_i(x) = \begin{cases} 0, & a \leq x \leq a + (i-1)h \\ (x - (a + (i-1)h))/h, & a + (i-1)h \leq x \leq a + ih \\ 1 - (x - (a + ih))/h, & a + ih \leq x \leq a + (i+1)h \\ 0, & a + (i+1)h \leq x \leq b \end{cases}$$

These are called tent functions since $\varphi_i(x)$ is zero to the right of $a+(i-1)h$, rises linearly to a peak at $a+ih$, and then falls back to zero at $a+(i+1)h$, and remains zero. While both

step function and piecewise linear approximations fit into our general linear approach, they differ in that the basis elements are zero over most of the domain, and at each point in the domain most basis functions are zero. This is the defining feature of *finite element* approaches to approximation. While the resulting bases are not strictly orthogonal, they are close to being so since the inner product of most distinct pairs of basis elements is zero.

Hermite Interpolation Polynomials. Next, suppose that we have both level and slope information at $x_1, \cdots, x_n$. Within each $[x_i, x_{i+1}]$ interval, we construct the Hermite interpolation polynomial given the level and slope information at x_i and x_{i+1}. The collection of interval-specific Hermite interpolations constitute a piecewise polynomial approximation. The resulting function is a cubic polynomial almost everywhere. However, at the interpolation nodes, it is only C^1. This lack of smoothness is often undesirable and is addressed by splines.

Splines. Another piecewise smooth scheme is to construct a *spline*. A spline is any smooth function which is piecewise polynomial but also almost as smooth where the polynomial pieces connect. Formally, a function $s(x)$ on $[a, b]$ is a spline of order k if s is C^{k-2} on $[a, b]$, and there is a grid of points (called nodes) $a = x_0 < x_1 < \cdots < x_n = b$ such that s is a polynomial of degree at most $n-1$ on each subinterval $[x_i, x_{i+1}]$, $i = 0, \ldots, n-1$. Note that order 2 splines are just the common piecewise linear functions.

The cubic spline (that is, of order 4) is popular. Suppose that we have Lagrange interpolation data $\{(x_i, y_i) \mid i = 0, \cdots, n\}$. The x_i will be the nodes of the spline, and we want to construct a spline, $s(x)$, such that $s(x_i) = y_i, i = 0, \cdots, n$. On each interval $[x_i, x_{i+1}]$, $s(x)$ will be a cubic $a_i + b_i\, x + c_i\, x^2 + d_i\, x^3$. The definition of a cubic spline together with the Lagrange data provides us with $4n-2$ conditions on the $4n$ coefficients. Various splines are differentiated by the two additional conditions imposed. One way to fix the spline is to pin down $s'(x_0)$ and $s'(x_n)$. For example, the *natural spline* imposes $s'(x_0) = 0 = s'(x_n)$. *Hermite* splines give $s'(x_0)$ and $s'(x_n)$ values $f'(x_0)$ and $f'(x_n)$ when these are known.

In general, degree k splines with data at n nodes will yield $O(n^{-(k+1)})$ convergence for $f \in C^{k+1}[a, b]$. Splines are excellent for approximations for two general reasons. First, evaluation is cheap since splines are locally cubic. To evaluate a spline at x you must first find which interval $[x_i, x_{i+1}]$ contains x, then find the coefficients for the particular cubic polynomial used over $[x_i, x_{i+1}]$, and evaluate that cubic at x. The second reason for using splines is that good fits are possible even for functions which are not C^∞ or have regions of large higher-order derivatives, situations where orthogonal polynomials do not do as well since global approximation schemes have difficulties in dealing with small regions of high curvature. On the other hand, if a function is well–behaved, orthogonal polynomials will generally do better.

7.7. Shape-Preserving Interpolation. Above we have focused on the pointwise convergence properties of various approximation schemes. Sometimes we will want to both interpolate data and preserve some shape in the data. For example, if the interpolation data indicates an increasing function, we may want to compute an approximation which is increasing everywhere, not only node-to-node but also between the interpolation nodes. Even though a scheme which converges pointwise will asymptotically preserve shape, these methods are not satisfactory since we will want to preserve the shape when we have a small amount of data, not just when we have large amounts of data. It is on this dimension where

the difference between orthogonal polynomials and piecewise polynomial approximations are important since orthogonal polynomials will not generally preserve shape.

Schumaker[117] presents a particularly simple way to construct shape-preserving quadratic splines. Suppose we want to find a function $s \in C^1[t_1, t_2]$ such that

$$s(t_i) = z_i,\ s'(t_i) = s_i,\ i = 1, 2$$

and, furthermore, suppose $z_1 < z_2, s_1 > s_2$, implying that the data are consistent with a concave function. The task is to find an interpolating function s which is also concave. Schumaker accomplishes this by adding one interpolation node $\xi_i \in [t_i, t_{i+1}]$ and constructs quadratic functions over $[t_i, \xi_i]$ and $[\xi_i, t_{i+1}]$ which together make a concave C^1 function s on $[t_i, t_{i+1}]$. In general, if the data on $[t_i, t_{i+1}]$ are consistent with monotonicity, concavity, convexity, or nonnegativity, then one can construct a piecewise quadratic function which is monotone, concave, convex, or nonnegative on $[t_i, t_{i+1}]$. The nontrivial fact here is such a ξ_i exists for any interpolation data. By piecing together these functions over subintervals, we can preserve shape globally. If one does not have slope information, one need only to choose the slope parameters so as to be consistent with the shape of the data. Schumaker also shows how to make judicious estimates of the slopes.

There are many papers on this topic; see Judd[74] for several references.

7.8. Multidimensional Approximation. Most economic problems involve several dimensions – physical and human capital, capital stocks of competitors, wealth distribution, etc. When we attempt to approximate functions of several variables, many difficulties present themselves. We will discuss multidimensional interpolation and approximation methods, first by generalizing the one-dimensional methods via product formulations, and then by constructing inherently multidimensional schemes.

Tensor Product Bases. *Tensor product* methods build multidimensional basis functions up from simple one-dimensional basis functions. If $\{\varphi_i(x)\}_{i=1}^{\infty}$ is a basis for functions of one real variable, then the set of pairwise products, $\{\varphi_i(x)\varphi_j(y)\}_{i,j=1}^{\infty}$ is the tensor product basis for functions of two variables. To handle n dimensional problems in general, one can take all the n-wise products, and create the *n-fold tensor product* of a one-dimensional basis. The tensor approach can extend orthogonal polynomials and spline approximation methods to several dimensions. One advantage of the tensor product approach is that if the one-dimensional basis is orthogonal in a norm, the tensor product is orthogonal in the product norm. The disadvantage is that the number of elements increases exponentially in the dimension.

Complete Polynomials. There are many ways to form multidimensional bases and avoid the "curse of dimensionality." One way is to use *complete polynomial* bases, which grow only polynomially as the dimension increases. To motivate the complete polynomials, recall Taylor's theorem for R^n in Theorem 1 above. Notice the terms used in the k'th degree Taylor series expansion. For $k = 1$, Taylor's theorem uses the linear functions $\mathcal{P}_1 \equiv \{1, x_1, x_2, \cdots, x_n\}$. For $k = 2$, Taylor's theorem uses

$$\mathcal{P}_2 \equiv \mathcal{P}_1 \cup \{x_1^2, \cdots, x_n^2, x_1x_2, x_1x_3, \cdots, x_{n-1}x_n\}.$$

$\mathcal{P}_2$ contains some cross-product terms, but not all; for example, $x_1x_2x_3$ is not in $\mathcal{P}_2$. In general, the k'th degree expansion uses functions in

$$\mathcal{P}_k \equiv \{x_1^{i_1} \cdots x_n^{i_n} \mid \sum_{\ell=1}^{n} i_\ell \leq k,\ 0 \leq i_1, \cdots, i_n\}.$$

The set $\mathcal{P}_k$ is the *complete set of polynomials of total degree k.*

Complete sets of polynomials are often superior to tensor products for multivariate approximation. The n-fold tensor product of $\{1, x, \cdots, x^k\}$ contains $(k+1)^n$ elements, far more than $\mathcal{P}_k$. For example, $\mathcal{P}_2$ contains $1 + n + n(n+1)/2$ elements compared to 3^n for the tensor product. Taylor's Theorem tells us that many of the tensor product elements add little to the approximation, saying that the elements of $\mathcal{P}_k$ will yield a k'th order approximation near x^0, and but that the n-fold tensor product of $\{1, x, \cdots, x^k\}$ can do no better than k'th order convergence since it does not contain all degree $k+1$ terms. This suggests that a complete family of polynomials will give us nearly as good an approximation as the tensor product of the same order, but with far fewer elements.

Finite Element Approaches. Finite element methods use bases whose elements have small support. One simple example is bilinear interpolation. Suppose we have the values of $f(x,y)$ at $(x,y) = (\pm 1, \pm 1)$. Then, the following 4 functions form a cardinal interpolation basis:

$$\begin{array}{ll} \varphi_1(x,y) = \frac{1}{4}(1-x)(1-y), & \varphi_2(x,y) = \frac{1}{4}(1+x)(1-y) \\ \varphi_3(x,y) = \frac{1}{4}(1+x)(1+y), & \varphi_4(x,y) = \frac{1}{4}(1-x)(1+y) \end{array} \tag{46}$$

The bilinear approximation to f on the square $[-1,1] \times [-1,1]$, which is an example of an *element*, is

$$f(-1,-1)\phi_1(x,y) + f(1,-1)\phi_2(x,y) + f(1,1)\phi_3(x,y) + f(-1,1)\phi_4(x,y).$$

The approximation is linear at each edge, but generally has a saddlepoint curvature on the interior. To interpolate data on a two-dimensional lattice, we create the bilinear approximation on each square.

Finite element methods consist of partitioning a domain into several elements, and patching together the local approximations on the elements, but this is not easy. Since we generally want the result to be a continuous function, care must be taken that resulting approximation is continuous across element boundaries. With bilinear interpolation, this will hold since any two approximations overlap only at the edges of rectangles, and on those edges the approximation is the linear interpolant between the common vertices. If we know that we are approximating a smooth function, then the kinks at the edges of the elements may make bilinear approximation unappealing. Assuring smoothness at element boundaries is an increasingly difficult problem as we increase the desired degree of differentiability and the dimension. The bilinear finite element scheme is just the simplest of a large number of finite element approximation schemes. There is a large literature on finite element approximations of multidimensional functions (see Burnett[22]).

Neural Networks. The previous approximation procedures are based on linear combinations of polynomial functions. Neural networks provide us with an alternative and inherently nonlinear functional form for approximation. A *single-layer* neural network is a function of the form

$$F(x;\beta) \equiv h(\sum_{i=1}^{n} \beta_i g(x_i)) \tag{47}$$

where $x \in R^n$ is the vector of inputs and h and g are scalar functions. A common form chooses $g(x) = x$, reducing (47) to the form $h(\beta^T x)$. A *single hidden-layer feedforward*

network is the form

$$F(x;\beta,\gamma) \equiv f(\sum_{j=1}^{m} \gamma_j h(\sum_{i=1}^{n} \beta_i^j g(x_i))). \tag{48}$$

Note the simplicity of the functional forms; this simplicity makes neural network approximations easy to evaluate.

The data for a neural network consists of (y,x) pairs such that y is supposed to be the output of a neural network if x is the input. This requirement imposes conditions on the parameters β in (47) and β and γ in (48). One fits single-layer neural networks by finding β to solve

$$\min_{\beta} \sum_{j} (y_j - F(x^j;\beta))^2$$

and the objective of a single hidden-layer feedforward network is to solve

$$\min_{\beta,\gamma} \sum_{j} (y_j - F(x^j;\beta,\gamma))^2.$$

which are just instances of nonlinear least squares fitting.

The approximating power of neural network approximation is indicated by theorems of Horni, Stinchcombe and White (see White [122] for a wide-ranging discussion of neural networks and their properties). Let G be a continuous function, $G : R \to R$, such that $\int_{-\infty}^{\infty} G(x)dx$ is finite and nonzero and G is L^p for $1 \leq p < \infty$. Let

$$\Sigma^n(G) = \{g : R^n \to R \mid \quad g(x) = \textstyle\sum_{j=1}^{m} \beta_j G(w^j \cdot x + b_j),\ b_j, \beta_j \in R\,,$$

$$w^j \in R^n, w^j \neq 0,\ m = 1,2,\cdots\}$$

be the set of all possible single-hidden layer feedforward neural networks using G as the hidden layer activation function.

Theorem 11. *Let $f : R^n \to R$ be continuous. Then for all $\epsilon > 0$ probability measure μ, and compact sets $K \subset R^n$, there is a $g \in \Sigma^n(G)$ such that*

$$\sup_{x \in K} |f(x) - g(x)| \leq \epsilon$$

and

$$\int_R |f(x) - g(x)|\, d\mu \quad \leq \epsilon$$

This also holds when G is a squashing functions, i.e., $G : R \to [0,1]$, G is nondecreasing, $\lim_{x\to\infty} G(x) = 1$, and $\lim_{x\to-\infty} G(x) = 0$.[14]

These are universal approximation results which justifies the use of neural network approximation, and helps explain its success. There is some evidence that neural network approximation methods may be particularly efficient at multidimensional approximation in the sense of needing relatively few free parameters; see Barron[6] for a recent result. The theoretical development of neural networks is proceeding, but is inherently difficult because of the nonlinearity of this approach.

[14] Note that a squashing function is a cumulative distribution function and vice-versa. A common choice for G is the sigmoid function, $G(x) = 1/(1+e^{-x})$.

8. Applications of Approximation to Dynamic Programming

Approximation methods are a key part of most numerical procedures. They are particularly important in discrete-time dynamic programming problems. These problems are among the most useful and basic of dynamic economic problems, with well-understood theoretical properties. We will briefly discuss them and the approximation aspects of numerical dynamic programming.

Let $\pi(u, x)$be profit flow if the state is x and the control is u. Suppose the law of motion is

$$x_{t+1} = g(x_t, u_t)$$

Then the value function, $V(x)$, solves

$$V(x) = \max_u \ \pi(u, x) + \beta V(g(x, u)) \equiv (TV)(x) \tag{49}$$

The standard theoretical procedure is to iterate on the basic functional equation, (49). If we could handle arbitrary functions, we would start with a guess, V_0, and then compute the sequence $\{V_n\}$ generated by

$$V_n = TV_{n-1} \tag{50}$$

This procedure converges when viewed as a mapping in the space of value functions.

On the computer, however, one cannot store arbitrary functions. There are several details which need to be decided to compute approximations to (50). Since we cannot deal directly with the space of continuous functions, we focus on a finite-dimensional subspace. We will approximate $V(x)$ as a finite linear sum of basis functions.

$$V(x) \doteq \sum_{i=1}^{N} a_i \varphi_i(x) \equiv \hat{V}(x\ ,\ \vec{a}) \tag{51}$$

Numerical procedures construct a $\hat{V}(x)$ which approximately satisfies the Bellman equation, (49). More specifically, the objective is to find a vector, $\vec{a} \in R^N$, such that $\hat{V}$ solves (49) as closely as possible.

The basic task is to replace T, an operator mapping continuous functions to continuous functions, with a finite-dimensional approximation, $\hat{T}$, which maps functions of the form in (51) to functions of the same form. We construct $\hat{T}$ in two steps. First, we choose a finite collection, X, of points x, and evaluate $(T\hat{V})(x)$ at $x \in X$. We will refer to this as the *maximization step* since it is the maximization problem in (49) at x. The resulting values are points on the function $T\hat{V}$. Since we are approximating a continuous value function, we use that information to choose a value function of form (51) which "best" summarizes the information generated concerning $T\hat{V}$. This is the critical *approximation step*, and we denote the result $\hat{T}\hat{V}$. In essence, $\hat{T}$ takes a function of form (51), $\hat{V}$, and maps it to another function of the same form, and is therefore a mapping in the space of coefficients, and the objective is to find a fixed point for $\hat{T}$ in the space of coefficients. We can also view $\hat{T}$ as a mapping from continuous functions to the finite-dimensional subspace representable as $\hat{V}(x\ ,\ \vec{a})$, in which case the problem is to find a fixed point of $\hat{T}$ in the space of functions of form (51).

The details of the approximation aspects of this procedure – choosing a basis for the expression of V, choosing points X to evaluate $T\hat{V}$, and fitting the data – are important. We next discuss some basic approaches.

8.1. Discretization Methods. The simplest approximation procedure is to discretize the state space, that is, they replace the problem on a continuous state space with one with a finite number of points. This has the advantage of reducing the problem to one of finite matrices. The other advantage is that the resulting analysis exactly solves some similar economic problem. See Rust[112] for a discussion of numerical dynamic programming procedures. Even in the case of discrete–state dynamic programming, projection solution ideas come into use. The key computation in the discrete–state approach is the solution of a large linear system. This can be accomplished approximately by using the GMRES method (see Saad and Schulz[113]) which essentially computes a few directions and finds an approximate solution which is spanned by these directions and minimizes a loss function.

While the discretization method does not obviously fit the description above, it is generally equivalent to approximating the value function with a step function[15]. However, step functions are highly inefficient ways to approximate a smooth value function. Because of this, the discretized state space method is unlikely to be of much value in economic analysis outside of naturally discrete problems, one–dimensional problems, or problems where the solutions are so nonsmooth that discretization is competitive with smooth approximation schemes. The impracticality of discretization is indicated by the fact that supercomputers are often used. Multidimensional problems are practically impossible, even for supercomputers, since the "curse of dimensionality" is particularly vexing for this method; if N points are used for a one-dimensional problem, then N^d points will be used for a d -dimensional problem. There are several ingenious methods for making discrete state problems more efficient; see Rust[112] for a description of these algorithms. We will focus on the alternatives presented by the application of approximation ideas.

8.2. Multilinear Approximation. While the discretization approach has been popular in macroeconomics, many other economists and the operations research literature in general have moved instead to continuous approximations of the value function. The simplest example of this is the DYGAM package discussed in Dantzig, et al. [41] which used multilinear interpolation on hypercubes when computing $\hat{V}_{n+1}$ from the information generated by $T\hat{V}_n$. In economics, Zeldes[130] has used piecewise linear approximations in life-cycle modelling.

This procedure has several advantages. Far fewer nodes are needed compared to a discretization method since the continuity of V is being exploited. There are some difficulties. First, the kinks make the optimization step more difficult, and are unrepresentative of V if V is C^2. Second, multilinear approximation generates curvature properties which may cause multiple local optima in the optimization step. The problem is that the interpolation may not have the same shape as the data.

8.3. Polynomial Approximations. If a little continuity is good, then more should be better if V is sufficiently continuous. In this spirit, Bellman et al.[7] proposed the use of polynomials, Daniel[38] discussed the use of splines, and Johnson et al[64] report computing experience with a variety of approximation schemes. Judd[74] presents an example of using a tensor-product basis of Chebyshev polynomials to solve a three-dimensional

[15] After computing the solution to (50), many users then use linear interpolation to estimate the value function at points not part of the discretized state space. Since this linear interpolation is done only after the value iteration is completed, it does not affect these comments and it's contribution to improving the algorithm's accuracy is limited.

optimal growth problem. The advantages of polynomial approximations are that fewer points are evaluated and increased smoothness makes the optimization step more rapid.

There are, however, some problems which may arise with polynomial approximation which don't arise with discretization or multilinear approximation. The difficulty is that many interpolation schemes do not preserve shape. Even if we use the best possible interpolation scheme, the resulting approximation may not be good in between the nodes in X, and can lead to instabilities in the value iteration algorithm. To deal with this, Judd and Solnick[78] proposes the use of shape-preserving polynomials to construct value function approximations, and computes upper bounds on the error which are superior to those from the discretization approach. In particular, this leads to convergence proofs for value function iteration with shape-preserving approximation, an important fact in itself since there can be no such proof for value iteration with polynomial approximation schemes in general.

9. Projection Methods

Our discussion of dynamic programming indicated that approximation ideas may be useful in solving the operator equations which arise in dynamic programming. We next discuss how these ideas from approximation theory naturally lead to algorithms for solving many of the operator equations which arise in economics. They are called *projection methods,* also known as *weighted residual methods.* We will describe the general projection approach for solving general operator problems. In fact, most of the techniques currently used by economists are also projection methods when viewed from the general perspective.

The first important observation is that in many economic models, equilibrium can be expressed as a collection of functions. In dynamic programming problems, that unknown function will be either the value function and/or the optimal policy function. In dynamic games, the unknown functions are the agents' strategy functions. In optimal growth models, the unknown function may be the optimal policy function. In dynamic equilibrium models, the unknown functions would include functions which indicate consumption demand, labor supply, asset trading strategies, and asset and commodity prices, all as functions of the underlying state variables. For specificity, consider the following simple deterministic growth problem:

$$\sum_{t=0}^{\infty} \beta^t \, u(c_t)$$

where capital obeys the law of motion

$$k_{t+1} = f(k_t) - c_t$$

To calculate the optimal consumption policy (and competitive equilibrium consumption function), $h(k)$, it is enough to focus on the Euler equation,

$$0 = u'(h(k)) - \beta \, u'(h(f(k) - h(k))) \, f'(f(k) - h(k)) \equiv \; (\mathcal{N}(h))(k) \tag{52}$$

The basic idea of projection techniques is to first express equilibrium conditions on these functions as a zero of an operator, $\mathcal{N} : B_1 \to B_2$, where B_1 and B_2 are function spaces. In (52) above, the Euler equation error is defined to be that operator, where B_1 and B_2 are both equal to the space of continuous functions on $[0, \infty)$. In general, the operator $\mathcal{N}$ can be an ordinary differential equation, as in optimal control problems,

a partial differential equation, as in continuous-time dynamic programming, or a more general functional equation, as in Euler equations expressing necessary conditions for recursive equilibria (as formulated in Prescott and Mehra[26]). Of course, space and time limitations make it impossible for computers to store and evaluate all possible elements of B_1. To make the problem tractable, projection methods focus on a finite-dimensional subspace of candidates in B_1 which can be easily represented on a computer and is likely to contain elements "close" to the true solution. The selection of this finite-dimensional space naturally exploits approximation methods. It may be difficult for the computer to evaluate $\mathcal{N}$, in which case we find a computable operator, $\hat{\mathcal{N}}$, which is "similar" to $\mathcal{N}$. Within the finite-dimensional space of candidate solutions, we find an element which is "almost" a zero of $\hat{\mathcal{N}}$.

While the basic idea is natural, there are many details. The key details are specifying the approximation method we will use, the finite-dimensional subspace within which we look for an approximate solution, and the computer representation of $\mathcal{N}$, defining what "close" and "almost" mean, and finding the approximate solution. By studying these details, we will see how to implement these ideas efficiently to solve numerically interesting dynamic nonlinear economic problems.

9.1. General Projection Algorithm. We next describe the projection method in a general context. One begins with an operator equation representation of the problem, that is, one reduces the economic problem to finding an operator $\mathcal{N}$ and a function f such that equilibrium is represented by the solution to

$$\mathcal{N}(f) = 0$$

where $f : D \subset R^N \to R^M$, $\mathcal{N} : B_1 \to B_2$, and the B_i are function spaces. Typically $\mathcal{N}$ is a composition of algebraic operations, differential and integral operators, and functional compositions, and is frequently nonlinear. We shall show how to implement the canonical projection technique in a step-by-step fashion. We first give an overview of the approach, then highlight the critical issues for each step, and discuss how the steps interact.

The first step is to decide how to represent approximate solutions. One general way is to assume that our approximation, $\hat{f}$, is built up as a linear combination of simple functions. We will also need a concept of when two functions are close. Therefore, the first step is to choose a basis and an appropriate concept of distance:

Step 1) Choose bases, $\Phi_j = \{\varphi_i\}_{i=1}^{\infty}$, and inner products, $< \cdot, \cdot >_j$, over B_j, $j = 1, 2$.

The basis over B_1 should be flexible, capable of yielding a good approximation for the solution, and the inner products should induce useful norms on the spaces.

Next, we decide how many basis elements to use and how to implement $\mathcal{N}$:

Step 2) Choose a degree of approximation n for f, a computable approximation $\hat{\mathcal{N}}$ of $\mathcal{N}$, and a collection of n functions from B_2, $p_i : D \to R^M$, $i = 1, \cdots, n$.

The approximate solution will be $\hat{f} \equiv \Sigma_{i=1}^{n} a_i \varphi_i(x)$. The convention is that the φ_i increase in "complexity" and "nonlinearity" as i increases, and that the first n elements are used. The best choice of n cannot be determined *a priori*. Generally, the only "correct" choice is $n = \infty$. Larger n should yield better approximations, but one is most interested in the smallest n which yields an acceptable approximation. One initially begins with small n and increases n until some diagnostic indicates that little is gained by continuing. Similar

issues arise in choosing $\hat{\mathcal{N}}$. Sometimes we can take $\hat{\mathcal{N}} = \mathcal{N}$, but more generally some approximation is necessary. The p_i are the projection directions we will use to determine $\vec{a}$.

Step 1 lays down the topological structure of our approximation and Step 2 fixes the flexibility of the approximation. Once we have made these basic decisions, we begin our search for an approximate solution to the problem. Since the true solution f satisfies $\mathcal{N}(f) = 0$, we will choose as our approximation some $\hat{f}$ which makes $\hat{\mathcal{N}}(\hat{f})$ "nearly" equal to the zero function. Since $\hat{f}$ is parameterized by $\vec{a}$, the problem reduces to finding a coefficient vector $\vec{a}$ which makes $\hat{\mathcal{N}}(\hat{f})$ nearly zero. This search for $\vec{a}$ is the focus of Steps 3–5.

Step 3) For a guess $\vec{a}$, compute the approximation, $\hat{f} \equiv \Sigma_{i=1}^{n} a_i \varphi_i(x)$, and the residual function,

$$R(x\,;\,\vec{a}) \equiv (\hat{\mathcal{N}}(\hat{f}))\,(x).$$

The first guess of $\vec{a}$ should reflect some initial knowledge about the solution. After the initial guess, further guesses are generated in Steps 4 and 5, where we see how we use the inner product, $< \cdot, \cdot >_2$, defined in the space B_2, to define what "near" means.

Step 4) For each guess of $\vec{a}$, compute the n projections,

$$P_i(\vec{a}) \equiv \langle R(\cdot\,;\,\vec{a})\,, p_i(\cdot)\rangle_2\,, i = 1, \cdots, n.$$

Step 5) By making a series of guesses over $\vec{a}$ and iterating over steps 3 and 4, find $\vec{a}$ which sets the n projections equal to zero.

This general algorithm breaks the numerical problem into several distinct steps. It points out the many distinct techniques of numerical analysis which are important. First, in Steps 1 and 2 we choose the finite-dimensional space wherein we look for approximate solutions, hoping that within this set there is something "close" to the real solution. These steps require us to think seriously about approximation theory methods. Second, Step 4 will involve numerical integration if we cannot explicitly compute the integrals which define the projections. Third, Step 5 is a distinct numerical problem, involving the solution of a nonlinear set of simultaneous equations or the solution of a minimization problem. We shall now consider each of these numerical problems in isolation.

Choice of Basis and Inner Product. There are many criteria which the basis and inner product should satisfy. The full basis Φ_1 for the space of candidate solutions should be "rich"; in particular, it should be complete in B_1. We will generally use inner products of the form

$$\langle f(x)\,,\, g(x)\rangle_1 \equiv \int_D f(x)g(x)w(x)\,dx$$

for some weighting function $w(x) \geq 0$.

Computational considerations also play a role in choosing a basis. The φ_i should be simple to compute. They should be similar in size to avoid scaling problems. While asymptotic results such as the Chebyshev Interpolation Theorem may lull one into accepting polynomial approximations, practical success requires a basis where only a few elements will do the job. This requires that the basis elements should "look something like" the solution. In particular, our discussion of approximation methods above shows

that we should use smooth functions to approximate smooth functions, but use splines to approximate functions which may have kinks or other extreme local behavior. We will also see that the use of orthogonal bases will enhance efficiency and accuracy. Because of its special properties, a generally useful choice is the Chebyshev polynomial family. If, on the other hand, one has a basis which is known to efficiently approximate the solution, one should use that instead or combine it with the Chebyshev polynomials. A good, problem–specific, choice of basis can substantially improve algorithmic performance over the generic approximation methods discussed above. However, the generic approaches are usually quite acceptable if one has no apparent problem–specific alternative.

Choice and Evaluation of Projection Conditions. Projection techniques include a variety of special methods. Generally we use $< \cdot, \cdot >$ to measure the "size" of the residual function, $R(x; \vec{a})$. The general strategy is to find an $\vec{a}$ which makes $R(x; \vec{a})$ small. There are several ways to proceed.

First, we have the *least-squares* approach which chooses $\vec{a}$ so as to minimize the "weighted sum of squared residuals":

$$\min_{\vec{a}} \quad \langle R(x; \vec{a}),\ R(x; \vec{a}) \rangle.$$

This replaces an infinite-dimensional operator equation with a nonlinear minimization problem in R^n. The standard difficulties may arise; for example, there may be local minima which are not global minima. The objective may be poorly conditioned. However, there is no reason for these problems to arise more often here than in any other context, such as maximum likelihood estimation, where extremal problems are solved numerically.

While the least-squares method is a direct approach to making $R(x; \vec{a})$ small, most projection techniques find approximations by fixing n projections and choosing $\vec{a}$ to make the projection of the residual function in each of those n directions zero. Formally, these methods find $\vec{a}$ such that $\langle R(x; \vec{a}), p_i(x) \rangle_2 = 0$ for some specified collection of functions, p_i. Different choices of the p_i defines different implementations of the projection method.

One such technique is the *Galerkin* method. In the Galerkin method we use the first n elements of the basis for the projection directions. Therefore, $\vec{a}$ is chosen to solve the equations:

$$P_i(\vec{a}) \equiv \langle R(x; \vec{a}),\ \varphi_i(x) \rangle = 0, \qquad i = 1, \cdots, n$$

Notice that here we have reduced the problem of solving a functional equation to solving a finite set of finite–dimensional nonlinear equations. In some cases in physics, the Galerkin projection equations are the first-order conditions to some least-squares minimization problem, in which case the Galerkin method is also called the *Rayleigh-Ritz* method. This is not as likely to happen in economics problems because of the inherent nonlinearities.

There are obviously many ways to implement the projection idea. A collocation method takes n points from the domain D, $\{x_i\}_{i=1}^n$, and chooses $\vec{a}$ to solve

$$R(x_i\,; \vec{a}) = 0, \qquad i = 1, \cdots, n$$

This is a special case of the projection approach since $R(x_i\,; \vec{a})$ equals the projection of $R(x; \vec{a})$ against the Dirac delta function at x_i, $\delta(x - x_i)$. This projection is written $\langle R(x; \vec{a}),\ \delta(x - x_i) \rangle_2$. *Orthogonal collocation* chooses the collocation points in a special way. The chosen x_i are the zeros of the n'th basis element, where the basis elements are orthogonal with respect to the inner product. The Chebyshev Interpolation Theorem suggests its power. Suppose we have found an $\vec{a}$ such that $R(z_i^n\,; \vec{a}) = 0$, $i = 1, \cdots, n$,

where the z_i^n are the n zeros of T_n. As long as $R(x;\vec{a})$ is smooth in x, the Chebyshev Interpolation Theorem says that these zero conditions force $R(x;\vec{a})$ to be close to zero for all x, and that these are the best possible points to use if we are to force $R(x;\vec{a})$ to be close to zero. Even after absorbing these considerations, it is not certain that even orthogonal collocation is a reliable method; fortunately, its performance turns out to be surprisingly good.

Choosing the projection conditions is a critical decision since the major computational task is the computation of those projections. The collocation method is fastest in this regard since it only uses the value of R at n points. More generally, the projections will involve integration. In some cases one may be able to explicitly perform the integration. This is generally possible for linear problems, and possible for special nonlinear problems. However, our experience is that this will generally be impossible for nonlinear economic problems. We instead need to use numerical quadrature techniques to compute the integrals associated with evaluating $\langle \cdot, \cdot \rangle$. A typical quadrature formula approximates $\int_a^b f(x)\, g(x)\, dx$ with a finite sum $\sum_{i=1}^n \omega_i f(x_i)$ where the x_i are the quadrature nodes and the ω_i are the weights. Since these formulas also evaluate $R(x;\vec{a}\,)$ at just a finite number of points, x_i, quadrature-based projection techniques are essentially weighted collocation methods. The advantage of quadrature formulas over collocation is that information at more points is used to compute the approximation, hopefully yielding a more accurate approximation of the projections.

Finding the Solution. Step 5, which determines $\vec{a}$ by solving the projection conditions computed in Step 4, uses either a minimization algorithm (in the least-squares approach) or a nonlinear equation solver to solve the system $P(\vec{a}) = 0$. Many alternatives exist, including successive approximation, Newton's method, and homotopy methods, all of which have been used in the economics applications of the projection method.

10. APPLICATIONS OF PROJECTION METHODS TO RATIONAL EXPECTATIONS MODELS

Most methods used in numerical analysis of economic models fall within the general description above. We will see this below when we compare how various methods attack growth problems. The key fact is that the methods differ in their choices of basis, fitting criterion, and quadrature techniques. With the general method laid out, we will now report on a particularly important application to show its usefulness.

10.1. Discrete-Time Deterministic Optimal Growth. We examine optimal growth problems in discrete time and show how projection techniques can be adapted to calculate solutions. The stochastic case is one which has been studied by many others with various numerical techniques. In fact, one point we make below is that most of these procedures are really projection methods. By recognizing the common projection approach underlying these procedures, we can better understand their differences, particularly in accuracy and speed. We conjecture that the comparative performances of these various implementations of projection ideas in the discrete-time stochastic optimal growth problem is indicative of their relative value in other future problems.

We first examine the deterministic growth problem described above which is characterized in (52). We shall now describe the details of a projection approach to that problem. The domain D of our approximation will be $[k_m, k_M]$. k_m and k_M are chosen so that the solution will have k confined to $[k_m, k_M]$. In particular, $[k_m, k_M]$ must contain the steady

state, a point which we can determine before calculations begin. Our approximation to h is parametrically given by

$$\hat{h}(k\,;\,\vec{a}) = \sum_{i=1}^{n} a_i \psi_i(k)$$

where n is the number of terms used. Common choices include the Chebyshev polynomials $\psi_i(k) \equiv T_{i-1}(2\frac{k-k_m}{k_M-k_m} - 1)$, the tent functions, or the ordinary polynomials.

In this problem, $\mathcal{N}$ is a simple operator using only arithmetic operations and composition. Therefore, we can take $\hat{\mathcal{N}} = \mathcal{N}$. Since h is continuous, we define $\mathcal{N}$ to have domain and range in $C^0[k_m, k_M]$. Hence, $B_1 = B_2 = C^0[k_m, k_M]$, the continuity of $\mathcal{N}$ in the L^∞ norm following from the u, f, and $\hat{h}$ being C_1 in all their arguments. Given the Euler equation, (52), the residual function becomes

$$R(k\,;\vec{a}) = u'(h(k\,;\vec{a})) - \beta u'(h(f(k) - h(k\,;\vec{a});\vec{a}))f'(f(k) - h(k\,;\vec{a})) = \mathcal{N}(\hat{h})$$

To compute $\vec{a}$, we can do one of several things. First, we consider orthogonal collocation. We choose n values of k, denoted by k_i, $i = 1, ..., n$. We then choose $\vec{a}$ so that $R(k_i\,;\vec{a}) = 0$ for each i . Orthogonal collocation chooses the k_i to be the n zeros of ψ. The Chebyshev Interpolation Theorem strongly argues for using Chebyshev polynomials in this case. If $R(k_i\,;\vec{a}) = 0$ for each k_i, then we would like to conclude that $R(k_i\,;\vec{a})$ is the zero function on the domain D. The Chebyshev Interpolation Theorem says that this is most justified if the k_i were the Chebyshev zeros, and that if we use Chebyshev zeros, we are very likely to $R(k_i\,;\vec{a})$ to be nearly zero.

We could also implement the Galerkin method. If we use Chebyshev polynomials as a basis, then we use projections with the inner product

$$\langle f(k),\, g(k)\rangle \equiv \int_{k_m}^{k_M} f(k)g(k)w(k)dk$$

where

$$w(k) \equiv (1 - (2\frac{k - k_m}{k_M - k_m} - 1)^2)^{-\frac{1}{2}}.$$

With this choice of inner product, the basis is orthogonal. The Galerkin method computes the n projections

$$P_i(\vec{a}) \equiv \int_{k_m}^{k_M} R(k;\vec{a})\,\psi_i(k)\,w(k)\,dk, \qquad i = 1, \ldots, n$$

and chooses $\vec{a}$ so that $P(\vec{a}) = 0$. Here the difficulty is that each $P_i(\vec{a})$ is an integral which needs to be computed numerically. The form of $w(k)$ implies the use of Gauss-Chebyshev quadrature. That is, we approximate $P_i(\vec{a}) = 0$ conditions with

$$0 = \sum_{j=1}^{m} R(k_j;\vec{a})\,\psi_i(k_j)$$

for some $m > n$, with the k_j being the m zeros of ψ_{m+1}.

When we have calculated our estimate of $\vec{a}$, we would like to check if this procedure yields reliable approximations. Several diagnostics can be used to see if the proposed solution is acceptable. First, the a_k coefficients decline rapidly in k, as predicted by the

Chebyshev approximation theorem. Second, the low-order coefficients should be insensitive to the choice of n. While these facts do not prove that the approximation is good, we would be uncomfortable if the high-order coefficients were not small, or if the coefficient estimates were not stable as we increase n. We also want to examine test cases to see if the results from the projection method agree with the answer from another method known to be accurate. Judd[72] performs these tests on a variety of empirically interesting cases, finding that the projection method applied to this model is very accurate and very fast.

Table 2 (which is taken from Judd[72]) indicates the kind of accuracy which can be achieved. We assume that $f(k)$ was Cobb–Douglas with capital share of .25, and that the steady state capital stock is $k = 1$. We first solved the problem with an 800,000 point discretization over the range $[.5, 1.3]$. We then used the projection method to solve the problem. The entry under PROD indicates the output at k, and CONS indicates the optimal consumption as computed by a 800,000 point discretization method. The entries under $n = 9, 6, 4, 2$ columns indicate the error of the degree n polynomial approximation. The notation $a(-m)$ means $a \times 10^{-m}$. The results indicate that even a low-order approximation does quite well.

Table 2: Errors in Consumption Policy Function

k	$PROD$	$CONS$	$n = 9$	$n = 6$	$n = 4$	$n = 2$
0.50	0.1253211	0.1147611	3(−7)	3(−7)	0.01	−1(−4)
0.70	0.1401954	0.1335954	−3(−7)	−3(−7)	−1(−6)	1(−4)
0.80	0.1465765	0.1421165	−2(−7)	−1(−7)	−5(−6)	2(−4)
0.90	0.1524457	0.1501957	4(−7)	0.04	−5(−6)	2(−4)
1.00	0.1578947	0.1578947	0	−0.01	−3(−6)	2(−4)
1.10	0.1629916	0.1652816	−2(−7)	−2(−7)	2(−6)	9(−5)
1.30	0.1723252	0.1792852	2(−7)	2(−7)	4(−6)	−1(−4)

The tent function approach was used in Bizer and Judd[11] in a similar model. There the interpolation nodes were chosen to be uniformly distributed in D. The advantage of the piecewise linear approximation is that the resulting interpolation is shape-preserving. This may be useful since we know that h is monotone increasing. However, the shape–preservation considerations turn out not to be important relative to the differentiability considerations which argue for Chebyshev polynomials. The policy functions computed in Judd[72] using Chebyshev polynomials were monotone increasing, and using tent functions substantially reduced the algorithm's efficiency.

10.2. Stochastic Optimal Growth. We next turn to a stochastic optimal growth model. This example will show us how to handle multidimensional problems and the conditional expectations which arise in stochastic dynamic problems. We will also be able to describe the parameterized expectations method of solving rational expectations models.

More specifically, we examine the problem

$$\begin{aligned} \max \quad & E\left\{\Sigma_{t=0}^{\infty} \beta^t \, u(c_t)\right\} \\ & k_{t+1} = \theta_t f(k_t) - c_t \\ & \ln \theta_{t+1} = \rho \ln \theta_t + \epsilon_{t+1} \end{aligned} \tag{53}$$

where θ_t is a stationary AR(1) multiplicative productivity parameter. We will assume that the productivity shocks $\epsilon_t \sim N(0,\sigma^2)$ are independent. In this problem, both the beginning-of-period capital stock and the current value of θ are needed for a sufficient description of the state. Hence, consumption is a function of both k and θ, $h(k,\theta)$, and the Euler equation is

$$u'\left(h(k,\theta)\right) = \beta\, E\left\{u'(h(\theta f(k) - h(k,\theta),\tilde{\theta}))\,\tilde{\theta} f'(\theta f(k) - h(k,\theta)) \mid \theta\right\} \tag{54}$$

At this point, we will rewrite the Euler equation to make it more linear. We know that projection algorithms work well for linear problems. Perhaps our algorithm will do better if we make it more like a linear problem. To that end, rewrite (54) as

$$0 = h(k,\theta) - (u')^{-1}(\beta\, E\left\{u'(h(\theta f(k) - h(k,\theta),\tilde{\theta}))\,\tilde{\theta} f'(\theta f(k) - h(k,\theta)) \mid \theta\right\}) \tag{55}$$

Note that (55) has two terms, one linear in $h(k,\theta)$, and the other is similar to a CRTS function of next period's potential consumption values. This is the problem investigated in Judd[72] and in the Taylor–Uhlig symposium. We shall now describe and compare the various methods.

The procedure for finding $\hat{h}$ is similar to the deterministic case. First of all, we need to approximate the policy function. Judd[72] and Coleman[36] use approximations of the form

$$\hat{h}(k,\theta;\mathbf{a}) = \sum_{i=1}^{n_k}\sum_{j=1}^{n_\theta} a_{ij}\psi_{ij}(k,\theta)$$

where the ψ_{ij} functions are products of Chebyshev functions of k and θ in Judd, and tent functions of $\ln k$ and $\ln\theta$ in Coleman[36]. Judd[72] also considered complete polynomials. Comparisons followed the considerations outlined above. Since the policy function is smooth, the smooth approximation procedures did better with the complete polynomial approach doing best, that is, the greatest accuracy per unit of computer time. Coleman's choice of a finite element approach reduced efficiency since it used far more basis elements; furthermore, it cannot switch to a complete polynomial approach. These differences between the spectral approach advocated in Judd and the finite element approach used in Coleman become even larger as we move to higher dimensions.

In their approach to the stochastic growth model, den Haan and Marcet [57] parameterized the policy function to be

$$h(k,\theta) = (k^{\delta_2}\theta^{\delta_3}e^{\delta_1})^{1/\gamma} = (\exp\{\delta_1 + \delta_2\ln k + \delta_3\ln\theta\})^{\frac{1}{\gamma}} \tag{56}$$

that is, they assume that log consumption is a linear function of $\ln k$ and $\ln\theta$. However, this basis is not orthogonal. When they tried to improve the approximation to a quadratic form in $\ln k$ and $\ln\theta$, the lack of orthogonality lead to difficulties which prevented them from improving on the linear approximation. They argue that the collinearity of their basis elements is "a fortunate situation" and justifies their focus on low-order polynomial approximations. In contrast, the use of orthogonal bases in Judd and the use of a finite element approach in Coleman leads to no difficulties in finding substantially better approximations beyond low-order polynomials.

The comparisons of the Coleman, den Haan and Marcet, and Judd approaches to solving 54 illustrates the importance of approximation ideas. den Haan and Marcet, and Judd use polynomials to approximate what is presumed to be a smooth function. Coleman's contrasting use of finite elements introduces kinks in the approximation which forces

him to use many elements. The finite element approach and the orthogonal polynomial approach avoids the multicollinearity problems which limited den Haan and Marcet to low–order approximations. As these papers discuss, these differences lead to considerable differences in computational speed and accuracy in the final result.

10.3. Problems with Inequality Constraints. The optimal growth problem described above was simple in that the equilibrium was described in terms of an Euler equation which always had an interior solution. In some problems, constraints mean that the first-order conditions must be complemented with complementary slackness conditions. This was the nature of the problems which were the first to lead to numerical solutions of nonlinear rational expectations equilibria. Gustafson [56] investigated the problem of equilibrium storage of a storable commodity. He assume that output in period t is an exogenous random variable, $\widetilde{x}_t$, which is divided between a change in storage, $S_{t+1} - S_t$ (S_t is the beginning–of–period-t stock), and consumption, c_t. In equilibrium, price is a function of total stock, $p(S_t + x_t)$, and obeys the conditions

$$\begin{aligned} p(S_t + x_t) - E\{p(S_{t+1} + x_{t+1})\} &\geq 0 \\ (p(S_t + x_t) - E\{p(S_{t+1} + x_{t+1})\})S_{t+1} &= 0 \end{aligned} \tag{57}$$

where $S_{t+1} = S_t + x_t - D(p(S_t + x_t))$ and $D(p)$ is the demand function.

In some states of the world, the equilibrium storage level will be zero and the price function will not be a smooth. Gustafson[56] used a piecewise linear approximation of the equilibrium price function in his solution method. Piecewise linear approximations are relatively inefficient because many pieces are necessary to get a good approximation. In their analysis of the problem (which also generalized Gustafson to handle endogenous output) Williams and Wright solved for $E\{p_{t+1} \mid S_{t+1}\}$ as a function of S_{t+1},expressing this year's expectation of next year's price conditional on the amount stored for next year. This function determines the current price, future supply, and current stockpiling through an Euler equation similar to (57), but is smooth because it is a conditional expectation. Hence, they found that a low–order polynomial approximation was sufficient to solve the problem. Miranda and Helmburger [98] also used this insight in their analysis of stockpiling. Christiano and Fisher[32] applied the Wright–Williams technique for handling the inequality constraint to a constrained version of (53) and found similar advantages to using a smooth approximation.

This discussion points out two facts when dealing with inequality constraints. First, we can still use the same approximation ideas but we may have to adapt to handle the kinks which may arise. Second, skillful construction of the problem may result in finding a smooth function which characterizes equilibrium and allows us to use the more efficient smooth approximation schemes. Again, approximation ideas can be exploited to produce superior methods.

10.4. Dynamic Games. Methods which are useful for dynamic programming are also naturally useful for computing closed–loop (also known as Markov) equilibria of dynamic games. This holds since each player solves a dynamic programming problem, and equilibrium can be expressed as a coupled collection of Bellman equations for each player's dynamic programming problem. Kotlikoff, Shoven, and Spivak[80] used ordinary polynomials in their study of strategic saving and bequests. Miranda and Rui[99] computed

closed–loop equilibria for dynamic stockpiling games among commodity producing countries. They used Chebyshev polynomial approximations to players' value functions and projection methods to determine equilibrium value functions. In both cases, equilibrium was computed with relative ease.

10.5. Continuous Time Problems. The examples above have been of discrete-time systems. Projection methods have also been used to solve continuous–time models. One simple example is the canonical continuous-time optimal growth problem described above in (21), which reduced to solving the differential equation:

$$0 = C'(k)\,(f(k) - C(k)) - \frac{u'(C(k))}{u''(C(k))}(\rho - f'(k)) \equiv \mathcal{E}\,(k; C)$$

Judd[71] used a basis of Chebyshev polynomials to approximate $C(k)$ with $\hat{C}(k,a) \equiv \sum_{i=0}^{n-1} a_i\,T_i(k)$, on a large interval of capital stocks. Again, the performance of the algorithm was very good, independent of the details of the implementation. In fact, it easily outperformed the more commonly used shooting approach to the problem. Judd also extended this model to allow for taxation and uncertainty in continuous time. In all cases, accurate results were obtained quickly. Since projection methods were initially developed to deal with continuous–time systems represented by ordinary and partial differential equations, this is not surprising. One suspects that continuous–time systems in general will be readily computed with projection methods.

10.6. Models with Asymmetric Information. Many of the examples discussed above reduced to applying the projection method to standard mathematical problems – ordinary and partial differential equations and integral equations. To demonstrate the flexibility of the projection method, we next examine a very different kind of problem – equilibrium where individual agents have different information. These problems do not reduce to any of the standard operator problems discussed in applied mathematical literature. However, one can attack them successfully with the projection method. We will first describe an application to asset markets with asymmetric information. We will then discuss other economic problems where these methods have potential.

Information and Asset Markets. Asset market equilibrium with imperfect information have been rigorously studied in recent years. Grossman[54] and Grossman and Stiglitz[55] began a long literature on the partial equilibrium analysis of security markets with asymmetric information. However, much of this literature makes very special and simple assumptions about the distribution of returns, the information asymmetries, investor tastes, and asset structure. The restrictions substantially limit the generality of the results and the range of questions which can be addressed.

Recently, Judd and Bernardo[76] applied projection methods to analyze these models without special functional form assumptions. A simple one-period investment problem illustrates the method. Suppose each investor invests in two assets. The safe asset pays out R dollars per dollar invested, and the basic risky asset (we will call it stock) pays out $\tilde{Z}$ dollars per share. If an investor begins the first period with W dollars in cash and θ_0 shares of stock, and ends the first period with θ shares of stock which trade at a price of p dollars per share, his second, and final, period consumption will be

$$\tilde{c} = (W - (\theta - \theta_0)p)\,R + \theta\tilde{Z}. \tag{58}$$

The first-order condition for the choice of θ will be

$$0 = E\{u'(\tilde{c})\,(\tilde{Z} - pR) \mid I\} \tag{59}$$

where I is the investor's information set.

Computing Conditional Expectations. The conditional expectation in (59) implies that our equilibrium concept involves a conditional expectation. Numerical implementation of the conditional expectation conditions is the most challenging aspect of this problem. To solve this problem, Judd and Bernardo used the following definition of conditional expectation:

$$Z(X) = E\{Y \mid X\}$$

if and only if

$$E\{(Z(X) - Y)f(X)\} = 0$$

for all bounded measurable functions, $f(X)$, of X. Intuitively, this says that the prediction error of the conditional expectation, $E\{Y \mid X\}$, is uncorrelated with any measurable function of the conditioning information, X. This definition replaces the conditional expectation with an infinite number of unconditional expectation conditions.

Computing an Asymmetric Information Rational Expectations Equilibrium. We now show how to compute an equilibrium. Assume three types of investors, with type i investors observing y_i. The state of the market includes all private signals, $y = (y_1, y_2, y_3)$, but each investor sees only the market-clearing price and his own information. Therefore, a rational expectations equilibrium includes a price function $p(y)$ and type-specific demand policy functions, $\theta_i(p, y_i)$ for $i = 1,2,3$, such that given $p(y)$, θ_i solves (59) for $i = 1,2,3$, and $\sum_{i=1}^{3} \theta_i(y_i, p(y)) = 1$ for all states y.

In their solution, Judd and Bernardo [76] approximate the price law, $p(y_1, y_2, y_3)$, and the demand rules, $\theta_i(p(y), y_i)$, with multivariate polynomials. To determine the unknown coefficients in those polynomials, they impose projection conditions on the investors' first-order conditions. The first-order-condition for a type i investor is given by

$$E_{y,Z}\{u'(\tilde{c}_i)(\tilde{Z} - pR) \mid y_i, p\} = 0, \quad i = 1,2,3. \tag{60}$$

Using the definition of conditional expectation given above they impose projection conditions of the form

$$E_{y,Z}\{u'(\tilde{c}_i)(\tilde{Z} - p(y)R)p(y)^j y_i^k\} = 0, \quad . \tag{61}$$

for various choices of $j, k \geq 0$. The condition in (61) states that the product of the excess return and the marginal utility of consumption for a type i agent is uncorrelated with polynomials in $p(y)$ and y_i.

After imposing a sufficient number of such conditions, the result is a system of projection conditions constituting a finite nonlinear system of algebraic equations. This reduces an infinite dimensional functional problem to a finite-dimensional algebraic problem. The projection conditions given in equations (61) are only part of the conditional expectation condition given in equation (60). The hope is that a small number of projections can yield a useful approximation. Judd and Bernardo document the accuracy for this approximation method for a variety of distributions. Overall, their experience is that this method is reliable and reasonably fast.

10.7. Convergence Properties and Accuracy of Projection Methods. When using numerical procedures, it is desirable to know something concerning its errors. An important focus of theoretical numerical analysis is the derivation of bounds on errors. Two kinds of error results are desirable. First, it is desirable to derive an upper bound on the error for a given level of approximation. Second, if such upper bounds are not possible, it may still be valuable to know that the error goes to zero asymptotically, that is, as one lets the degree of approximation become arbitrarily large. The first kind of error information is rarely available. More typical in numerical algorithms for differential equations are asymptotic results. There has been little work on proving that the algorithms used by economists are asymptotically valid. Fortunately, there are general theorems concerning the consistency of the Galerkin method. Zeidler [128] and Krasnosel'skii and Zabreiko[81] demonstrate consistency under a variety of conditions. Even though it remains to be seen whether these theorems do cover our problems, they do indicate that projection methods are potentially valid for our economic problems.

Even if one had a convergence theorem for a method, it is clear that one cannot just blindly accept any answer one gets from a computation. Asymptotic theorems have a nasty feature of telling you only that the error goes to zero as your computational effort approaches infinity, but generally not telling you at what finite level of effort you may stop. Therefore, a more pragmatic approach is to ignore convergence theorems and instead use diagnostics to ask whether a solution is acceptable. We actually did that above in our construction of (26). There the issue was how well a perturbation expansion solved a continuous-time Euler equation. We constructed the approximation, substituted it in the Euler equation, and used the result, (26) to measure the amount of irrationality an economic agent is guilty of in equilibrium if each agent followed our approximate rule. If that number is small, say a dollar per million spent, then we argue that the approximate rule is as reasonable a prediction for behavior as the "true" equilibrium since people generally do not optimize beyond one part in a million.

In economic problems we can generally compute such diagnostics and measure the level of implied "irrationality." This diagnostic approach to evaluating a candidate solution can be applied independent of the computational method which produced the candidate solution. It does not rely on convergence; in fact, even if one uses a convergent method, one should still use such diagnostics to make sure that one did not stop the method too early. Furthermore, if one uses a method for which there is no convergence theorem, but it produces a solution which passes such diagnostics, the lack of a convergence theorem is irrelevant.

11. Hybrid Perturbation-Galerkin Method

We have discussed both perturbation and projection methods for solving economic models. While they are different approaches to approximation problems, we will next describe a method, the *hybrid perturbation-Galerkin* method which synergistically exploits their differences and similarities.

Suppose that there are a continuum of problems to be solved indexed by a parameter ϵ with the form

$$\mathcal{N}\left(f(x,\epsilon);\epsilon\right) = 0$$

Suppose that we can solve the $\epsilon = 0$ instance. The result of applying perturbation methods

near the $\epsilon = 0$ solution is the calculation of a series of the form

$$f(x,\epsilon) \sim \sum_{i=0}^{n} \delta_i(\epsilon)\,\varphi_i(x) \tag{62}$$

where the $\varphi_i(x)$ functions are computed by the perturbation calculations and the $\delta_i(\epsilon)$ are the generally prespecified gauge functions. Similarly, the result of a projection approach is an approximation of the form

$$f(x,\epsilon) \simeq \sum_{i=0}^{n} a_i(\epsilon)\varphi_i(x) \tag{63}$$

where the $\varphi_i(x)$ functions are the prespecified basis elements of the approximation system and the $a_i(\epsilon)$ coefficients are computed by the projection method. The strength of perturbation methods is that the approximations are quite good (in fact, asymptotically optimal) for small ϵ, but the weakness is that the quality may not hold up as ϵ increases. The projection approach tries to be good for any ϵ, but the difficulty is finding good bases which will allow the series in (63) to be short. Therefore, the strengths and weaknesses of these methods are complementary.

This observation turns out to be substantive. The idea of the hybrid perturbation-Galerkin method is to use the $\varphi_i(x)$ functions from perturbation calculations as the basis functions to be used in a projection method. We know that these functions constitute an optimal basis for small ϵ, and that the optimal weight on $\varphi_i(x)$ is $\delta_i(\epsilon)$ for small ϵ. The conjecture is that the φ_i functions still form a good basis for approximating $f(x,\epsilon)$ but that the weight on φ_i should not be the prespecified $\delta_i(\epsilon)$ but rather should be computed by (63).

Our continuous-time growth model gives a simple example of this approach. Recall the continuum of problems represented (27) and the related expansion (28). The objective there was to use a perturbation method to solving (21). We will use the results of the perturbation approach to develop a projection approach to solving (21). The first perturbation was the function

$$C_\epsilon(k) = k^\alpha \rho(\alpha^{-1} - \gamma) + (\gamma - \rho)k \tag{64}$$

Note that this function has a singularity at $k = 0$, a feature which is possibly also true of the solution to $C(k,1)$. This feature is absent in the orthogonal bases we discussed above. We see here already that this procedure has produced a basis element which has some advantages. To see if this is a good basis element, one can compare the basis $\{1, k, C_\epsilon(k)\}$ with the basis $\{1, k, k^2\}$. Computations show that the custom-made basis lead to solutions which had much smaller errors.

Note what (64) really suggests. Since the function k is already in the basis, $C_\epsilon(k)$ essentially adds the production function, k^α, to the basis. This reflects the general idea that the basis should be augmented by functions which are natural to the problem. Above we used differentiability properties to motivate basis elements. The hybrid approach attacks more precisely the problem of developing problem–appropriate bases.

Continuing the perturbation approach will generate a series of functions which can be used as a basis for a projection approach. For example, $C_{\epsilon\epsilon}(k)$ is a complicated function which essentially adds $k^{2\alpha-1}$ to the basis $\{1, k, C_\epsilon(k)\}$. This additional element is not as intuitive as $C_\epsilon(k)$, showing that the perturbation method will bring in elements other

than obvious ones. Again, computations show that the basis $\{1, k, C_\epsilon(k), C_{\epsilon\epsilon}(k)\}$ does much better than the basis $\{1, k, k^2, k^3\}$ in solving

These additional basis elements will possibly be collinear with previous elements. However, for any specified inner product, we can use a standard Gram–Schmidt procedure to construct a basis which spans the same space and is orthogonal. In this way, we can combine the conditioning advantages of orthogonal bases with the desirable shape properties of the perturbation functions.

Judd[73] discusses further the usefulness of this approach to producing bases. The example above just hints at the method's potential. In this example, reducing the number of basis elements is not important since the basis size is not a limiting factor in one-dimensional problems with smooth well-behaved solutions. However, basis size is a very important consideration in multidimensional problems. In those problems, a few well-chosen basis elements may allow for drastic reduction in basis size. One suspects that the hybrid perturbation-Galerkin method has substantial potential in multidimensional problems where economizing on the basis size is important.

The hybrid perturbation-Galerkin method also points out the value of combining methods. Since economics problems do not fit into standard mathematical classifications, it is likely that skillful combinations of various techniques will prove to be a powerful technique.

12. Conclusions

In this chapter we have reviewed a collection of approximation ideas which have proved themselves useful in computational analyses of economic models. We have also shown that a general class of techniques from the numerical partial differential equations literature can be usefully applied and adapted to solve nonlinear economic problems. Despite the specificity of the applications discussed here, the general description makes clear the general usefulness of perturbation and projection methods for economic problems, both theoretical modelling and empirical analysis. The application of perturbation and projection methods and the underlying approximation ideas have already substantially improved the efficiency of economic computations. Further exploitation of these ideas will surely lead to further progress.

References

[1] Albrecht, James W., Bertil Holmlund, and Harald Lang. "Comparative Statics in Dynamic Programming Models with an Application to Job Search," *Journal of Economic Dynamics and Control* **15** (No. 4, October, 1991), 755-769.

[2] Anderson, Gary S. "Symbolic Algebra Programming for Analyzing the Long-Run Dynamics of Economic Models," in *Economic and Financial Modeling with Mathematica*, ed., Hal Varian. Springer-Verlag: New York.

[3] Araujo, Aloisio, and Jose A. Scheinkman. "Smoothness, Comparative Dynamics, and the Turnpike Property," *Econometrica* **45** (April, 1977): 601-620.

[4] Balcer, Yves, and Kenneth L. Judd. "Dynamic Effects of Tax Policy," mimeo, 1985.

[5] Barro, Robert. "On The Determination of Public Debt," *Journal of Political Economy* 87 (October 1989), 940–971.

[6] Barron, A. R. "Universal Approximation Bounds for Superpositions of a Sigmoidal Function," *IEEE Transactions on Information Theory*," **39** (3, 1993), 930–945.

[7] Bellman, Richard, R. Kalaba, and B. Kotkin. "Polynomial Approximation: A New Computational Technique in Dynamic Programming: Allocation Processes," *Mathematics of Computation* **17** (1963), 155–161..

[8] Bender, Carl M., Steven A. Orszag. *Advanced Mathematical Methods for Scientists and Engineers*, McGraw-Hill, New York, 1978.

[9] Benhabib, Jess, and Kazuo Nishimura. "The Hopf Bifurcation and the Existence and Stability of Closed Orbits in Multisector Models of Optimal Economic Growth," *Journal of Economic Theory* **21**(3, December 1979), 421-444.

[10] Bensoussan, Alain. *Perturbation Methods in Optimal Control*, John Wiley and Sons, 1988.

[11] Bizer, David S., and Kenneth L. Judd, "Uncertainty and Taxation", *American Economic Review*, **79** (1989), 331-336.

[12] Blanchard, Olivier Jean and Charles M. Kahn. "The Solution of Linear Difference Models under Rational Expectations" *Econometrica* **48** (5, 1980), 1305–1311.

[13] Bleistein, N., and R. A. Handelsman. *Asymptotic Expansions of Integrals*, New York, Holt, Rinehart & Winston, 1976.

[14] Blume, Lawrence, David Easley, and Maureen O'Hara. "Characterization of Optimal Plans for Stochastic Dynamic Programs," *Journal of Economic Theory* **28** (1982), 221-234.

[15] de Boor, C. *A Practical Guide to Splines*, New York, NY: Springer, 1978.

[16] Bovenberg, A. Lans. "The Effects of Capital Income Taxation on International Competitiveness and Trade Flows," *American Economic Review* **79** (December, 1989), 1045–1064.

[17] Bovenberg, A. Lans. "Capital Income Taxation in Growing Open Economies," *Journal of Public Economics* **31** (December, 1986), 347–377.

[18] Bovenberg, A. Lans. The Corporate Income Tax in an Intertemporal Equilibrium Model with Imperfectly Mobile Capital," *International Economic Review* **29** (May, 1988), 321-340.

[19] Brock, William A. "Pathways to Randomness in the Economy: Emergent Nonlinearity and Chaos in Economics and Finance."

[20] Brock, William A., and L. J. Mirman, "Optimal Economic Growth and Uncertainty: The Discounted Case," *Journal of Economic Theory* **4** (1972), 479–513.

[21] Brock, William A., and Stephen J. Turnovsky, "The Analysis of Macroeconomic Policies in Perfect Foresight Equilibrium," *International Economic Review* **22** (February 1981), 179–209.

[22] D. S. Burnett, *Finite Element Analysis*, Addison-Wesley, Reading, Massachusetts, 1987.

[23] Budd, Christopher, Christopher Harris, and John Vickers. "A Model of the Evolution of Duopoly: Does the Asymmetry between Firms Tend to Increase or Decrease?" *Review of Economic Studies* **60** (No. 3, July 1993): 543–573.

[24] Caputo, Michael R. "How to do Comparative Dynamics on the Back of an Envelope in Optimal Control Theory," *Journal of Economic Dynamics and Control* 14 (1990), 655–683.

[25] Caputo, Michael R. "Comparative Dynamics via Envelope Methods in Variational Calculus," *Review of Economic Studies* 57(4, October 1990): 689-697.

[26] Chao, John C., and Peter C. B. Phillips, "Bayesian Posterior Distributions in Limited Information Analysis of the Simultaneous Equations Model," Yale University, mimeo, 1994.

[27] Chari, V. V., Lawrence J. Christiano, and Patrick J. Kehoe. "Policy Analysis in Business Cycle Models", mimeo, 1993.

[28] Cheney, E.W. *Introduction to Approximation Theory*, New York, NY: McGraw-Hill, 1966.

[29] Chiappori, Pierre Andre, Pierre Yves Geoffard, and Roger Guesnerie. "Sunspot Fluctuations around a Steady State: The Case of Multidimensional, One-Step Forward Looking Economic Models," *Econometrica* **60** (5, September 1992), 1097-1126.

[30] Christiano, Lawrence J. "Solving the Stochastic Growth Model by Linear-Quadratic Approximation and by Value-Function Iteration," *Journal of Business and Economic Statistics*, **8** (1990), 23–26.

[31] Christiano, Lawrence J. "Linear-Quadratic Approximation and Value-Function Iteration: A Comparison." *Journal of Business and Economic Statistics*, **8** (1990), 99–113.

[32] Christiano, Lawrence J. and J. D. M. Fisher. "Algorithms for Solving Dynamic Models with Occasionally Binding Constraints," mimeo, University of Western Ontario.

[33] Christiano, Lawrence J., and Victor Valdivia. "Notes on Solving Models Using a Linearization Method", Northwestern University, mimeo, 1993.

[34] Chow, Shui–Nee, and Jack K. Hale. *Methods of Bifurcation Theory*, Springer–Verlag, New York, 1982.

[35] Coddington, E. A., and N. Levinson, *Theory of Ordinary Differential Equations.* New York: McGraw-Hill. 1965.

[36] W. J. Coleman, II, "Solving the Stochastic Growth Model by Policy Function Iteration," *Journal of Business and Economic Statistics*, **8** (1990), 27–29.

[37] Cooley, and Gary Hansen. "The Inflation Tax in a Real Business Cycle Model," *American Economic Review* **79** (1989): 733–748.

[38] Daniel, J. W. "Splines and Efficiency in Dynamic Programming," *Journal of Mathematical Analysis and Applications* **54** (1976), 402–407.

[39] Danthine, Jean Pierre, and John B. Donaldson, "Stochastic Properties of Fast vs. Slow-Growing Economies," *Econometrica* **49** (1981), 1007–33.

[40] Danthine, Jean-Pierre, John B. Donaldson, and Rajiv Mehra, "On some computational aspects of equilibrium business cycle theory," *Journal of Economic Dynamics and Control* **13** (1989), 449–470.

[41] Dantzig, George B., R. P. Harvey, Z. F. Landowne, and R. D. McKnight. *DYGAM–A Computer System for the Solution of Dynamic Programs*, Control Analysis Corporation, Palo Alto, California. 1974.

[42] Dasgupta, Swapan, and Lionel W. McKenzie. "The Comparative Statics and Dynamics of Stationary States," *Preferences, Uncertainty, and Optimality: Essays in Honor of Leonid Hurwicz.* Boulder and Oxford: Westview Press, 1990, pages 280-303.

[43] Deaton, Angus, and Guy Laroque. "On the Behavior of Commodity Prices," Review of Economic Studies 59 (1992), 1-23.

[44] Dixit, Avinash. "Analytical Approximations in Models of Hysteresis," Review of Economic Studies **58** (No. 1, January, 1991), 141-151.

[45] Dotsey, Michael, and Ching Sheng Mao. "How Well Do Linear Approximation Methods Work?" *Journal of Monetary Economics* **29** (1992): 25–58.

[46] Feichtinger, Gustav. "Hopf Bifurcation in an Advertising Diffusion Model," *Journal of Economic Behavior and Organization* **17**(3, May 1992), 401-411.

[47] C. A. J. Fletcher, *Computational Galerkin Techniques*, Springer-Verlag: New York, 1984.

[48] Fleming, Wendell. "Stochastic Control for Small Noise Intensities," *SIAM Journal of Control* **9** (No. 3, August, 1971).

[49] Fleming, Wendell, and P. E. Souganides. "Asymptotic Series and the Method of Vanishing Viscosity," *Indiana University Mathematics Journal* **35** (2, 1986): 425-447.

[50] Franke, Reiner. "Stable, Unstable, and Persistent Cyclical Behaviour in a Keynes-Wicksell Monetary Growth Model," *Oxford Economic Papers* **44** (2, April 1992), 242-56.

[51] Friedman, Avner. "Stochastic Differential Games." *Journal of Differential Equations* **11** (1972): 79–108.

[52] Ghysels, Eric, and Offer Lieberman. "Dynamic Regression and Filtered Data Series: A Laplace Approximation to the Effects of Filtering in Small Samples," University of Montreal, 1993, mimeo.

[53] Golubitsky, Martin, and David G. Schaeffer. *Singularities and Groups in Bifurcation Theory: Volume I*, Springer–Verlag, New York, 1985.

[54] Grossman, Sanford. "On the Efficiency of Competitive Stock Markets Where Agents Have Diverse Information," *Journal of Finance* **18** (1976): 81–101.

[55] Grossman, Sanford J., and J. E. Stiglitz (1980): "On the Impossibility of Informationally Efficient markets," *American Economic Review*, **70**: 393–408.

[56] Robert L. Gustafson, Carryover levels for grains: a method for determining amounts that are optimal under specified conditions, USDA Technical Bulletin 1178, 1958.

[57] den Haan, Wouter, and Albert Marcet, Solving the stochastic growth model by parameterizing expedtations, *Journal of Business and Economic Statistics*, **8** (1990), 31–34

[58] Hansen, Gary D., and Edward C. Prescott. "Recursive Methods for Computing Equilibria of Business Cycle Models," IEM Discussion Paper 36, Federal Reserve Bank of Minneapolis, 1991.

[59] Hansen, Lars P., and Thomas J. Sargent. "Formulating and Estimating Dynamic Linear Rational Expectations Models," *Journal of Economic Dynamics and Control* **2** (1980), 7-46.

[60] Hansen, Lars P., and Thomas J. Sargent. *Recursive Linear Models of Dynamic Economies*, manuscript, Hoover Institution, 1990.

[61] Hansen, Lars P., and Kenneth Singleton. "Generalized Instrumental Variables Estimators of Nonlinear Rational Expectations Models," *Econometrica* **50** (1982), 1269–1286.

[62] Holly, A., and P.C.B. Phillips. "A Saddlepoint Approximation to the Distribution of the k-Class Estimator in a Coefficient in a Simultaneous System," *Econometrica* **47** (1979): 1527–48.

[63] Huffman, Gregory W. "A Dynamic Equilibrium Model of Asset Prices and Transaction Volume," *Journal of Political Economy* **95** (No. 1, February 1987): 138–159.

[64] Johnson, Sharon, Jery R. Stedinger, Christine A. Shoemaker, Ying Li, and Jose Alberto Tehada–Guibert. "Numerical Solution of Continuous–State Dynamic Programs using Linear and Spline Interpolation," *Operations Research* **41** (3, May–June 1993), 484–500.

[65] Judd, Kenneth L. "An Alternative to Steady-State Comparisons in Perfect Foresight Models," *Economics Letters* **10** (1982): 55–59.

[66] Judd, Kenneth L. "Short-run Analysis of Fiscal Policy in a Simple Perfect Foresight Model", *Journal of Political Economy* **93** (April, 1985), 298-319.

[67] Judd, Kenneth L. "Debt and Distortionary Taxation in a Simple Perfect Foresight Model", *Journal of Monetary Economics* **20** (1987): 51–72

[68] Judd, Kenneth L. "Welfare Cost of Factor Taxation in a Perfect Foresight Model", *Journal of Political Economy* **95** (August, 1987).

[69] Judd, Kenneth L. "Closed-Loop Equilibrium in a Multi–Stage Innovation Race," mimeo, 1985.

[70] Judd, Kenneth L. "Asymptotic methods in dynamic economic models," mimeo, Hoover Institution, October, 1990.

[71] Judd, Kenneth L. "Minimum weighted residual methods for solving dynamic economic models," mimeo, Hoover Institution, 1990.

[72] Judd, Kenneth L. "Projection Methods for Solving Aggregate Growth Models", *Journal of Economic Theory* **58** (No.2, December, 1992): 410–452.

[73] Judd, Kenneth L. "Hybrid Galerkin-Perturbation Methods Applied to Economic Problems," Hoover Institution, mimeo, 1995.

[74] Judd, Kenneth L. *Numerical Methods in Economics* , MIT Press (forthcoming).

[75] Judd, Kenneth L., and Sy-Ming Guu. "Perturbation Solution Methods for Economic Growth Models," in *Economic and Financial Modeling with Mathematica*, Hal Varian, ed., Springer-Verlag: New York, 1993.

[76] Judd, Kenneth L., and Antonio Bernardo, "Asset Market Equilibrium with General Securities, Tastes, Returns, and Information Asymmetries," Hoover Institution, mimeo, 1994.

[77] Judd, Kenneth L., and Sy-Ming Guu. "Bifurcation Approximation Methods Applied to Asset Market Equilibrium," Hoover Institution, mimeo, 1994.

[78] Judd, Kenneth L., and Andrew Solnick. "Numerical Dynamic Programming with Shape-Preserving Splines," mimeo, 1994.

[79] Kehoe, Timothy, and David Levine. "Comparative Statics and Perfect Foresight in Infinite Horizon Economies," *Econometrica* **53** (1985), 433–454..

[80] Kotlikoff, Lawrence, John Shoven, and Avi Spivak. "The Effect of Annuity Insurance on Savings and Inequality," *Journal of Labor Economics* **4** (3, part 2, 1986), S183–S207.

[81] Krasnosel'skii, M. A., and P. Zabreiko, "Geometrical Methods of Nonlinear Analysis," Springer-Verlag, 1984

[82] Kydland, Finn. "Equilibrium Solutions in Dynamic Dominant-Player Models," *Journal of Economic Theory* **15** (No. 2, August 1977), 307-24.

[83] Kydland, Finn. "Noncooperative and Dominant Player Solutions in Discrete Dynamic Games," *International Economic Review* **16** (No. 2, June, 1975), 321-35.

[84] Kydland, Finn E., and Edward C. Prescott. "A competitive theory of fluctuations and the feasibility and desirability of stabilization policy," in *Rational Expectations and Economic Policy*, ed. S. Fischer, Chicago: University of Chicago Press, pp. 169-198.

[85] Kydland, Finn E., and Edward C. Prescott. "Time to Build and Aggregate Fluctuations," *Econometrica* **50** (1982): 1345-1370.

[86] Lafrance, Jeffrey T., and L. Dwayne Barney, "The Envelope Theorem in Dynamic Optimization," *Journal of Economic Dynamics and Control* **15** (No. 2, April, 1991), 355–385.

[87] Laitner, John. "The stability of steady states in perfect foresight models," *Econometrica* **49** (1981), 319-333.

[88] Laitner, John. "Transition time paths for overlapping-generations models," *Journal of Economic Dynamics and Control* **7** (1984), 111-129.

[89] Laitner, John. "The Dynamic Analysis of Continuous-Time Life-Cycle Savings Growth Models," *Journal of Economic Dynamics and Control* **11** (1987), 331-357.

[90] Laitner, John. "Dynamic Determinacy and the Existence of Sunspot Equilibria," *Journal of Economic Theory* **47** (1, February 1989), 39-50.

[91] Laitner, John. "Tax Changes and Phase Diagrams for an Overlapping Generations Model," *Journal of Political Economy* **96** (No. 1, February 1990):193–220.

[92] Lucas, Robert E. "Optimal Investment Policy and the Flexible Accelerator," *International Economic Review*, **8** (February, 1967): 78-85.

[93] Magill, J. P. Michael. "A Local Analysis of N-Sector Capital Accumulation under Uncertainty," *Journal of Economic Theory*, **15** No. 1, June (1977), 211–219.

[94] Marcet, Albert. "Simulation Analysis of Dynamic Stochastic Models: Applications to Theory and Estimation," in C. Sims (ed.) *Advances in Econometrics Sixth World Congress*, Vol II., 91–118, Cambridge University Press.

[95] McGrattan, Ellen R. "Solving The Stochastic Growth Model by Linear-Quadratic Approximation." *Journal of Business and Economic Statistics* **8** (1, January 1990): 41–43.

[96] Merton, Robert C. "Optimal Consumption and Portfolio Rules in a Continuous-Time Model," *Journal of Economic Theory*, **3** (No. 4, December, 1971), 373–413..

[97] Milnor, T. *Topology from the Differentiable Viewpoint.* Charlottesville: University Press of Virginia. 1965.

[98] Miranda, Mario J., and Peter G. Helmberger. "The Effects of Commodity Price Stabiliization Programs." *American Economic Review* **78** (1988), 46–58.

[99] Miranda, Mario, and Xiongwen Rui. "Solving Dynamic Games via Polynomial Projection Methods: An Appplication to International Commodity Markets, Ohio State University, mimeo, 1994.

[100] Mortensen, Dale T. "Generalized Costs of Adjustment and Dynamic Factor Demand Theory," Econometrica **41** (July, 1973): 657-665.

[101] Oniki, Hajime. "Comparative Dynamics (Sensitivity Analysis) in Optimal Control Theory." *Journal of Economic Theory* **6** (June 1973): 265–283.

[102] Otani, Kiyoshi. "'Explicit Formulae of Comparative Dynamics." *International Economic Review* **23** (June 1982): 411–419.

[103] Phillips, Peter C.B. "Best Uniform and Modified Padé Approximants to Probability Densities in Econometrics" in *Advances in Econometrics*, Werner Hildebrand, ed., Cambridge University Press: Cambridge, 1982.

[104] Phillips, Peter C. B. "Marginal Densities of Instrument Variable Estimators in the General Single Equation Case", *Advances in Econometrics* **2**: 1-24.

[105] Phillips, Peter C. B. "To Criticize the Critics: An Objective Bayesian Analysis of Stochastic Trends," *Journal of Applied Econometrics* **6** (1991), 333–364.

[106] Prescott, Edward C., and R. Mehra, "Recursive competitive equilibrium: the case of homogeneous agents," *Econometrica* **48** (1980), 1365–1379.

[107] Rivlin, T.J. *An Introduction to the Approximation of Functions*, Waltham, Massachusetts: Blaisdell Publishing, 1969.

[108] Rice, John R. *Numerical Methods, Software, and Analysis*, Second Edition, Academic Press, Inc. Boston,1993.

[109] Rivlin, T. J. *Chebyshev Polynomials: From Approximation Theory to Algebra and Number Theory*, Wiley-Interscience, New York, 1990.

[110] Rust, John. "Optimal Replacement of GMC Bus Engines: An Empirical Model of Harold Zurcher," *Econometrica* **55** (5, 1987), 999–1033.

[111] Rust, John. "Maximum Likelihood Estimation of Discrete Control Processes," *SIAM Journal on Control and Optimization* **26** (1988), 1006-1023.

[112] Rust, John. "Numerical Dynamic Programming in Economics," this Handbook.

[113] Saad, Y., and M. H. Schulz. "GMRES: A Generalized Minimum Residual Algorithm for Solving Nonsymmetric Linear Systems," *SIAM Journal on Scientific and Statistical Computing* **7** (3, 1986), 856–869.

[114] Samuelson, Paul A. "The Fundamental Approximation Theorem of Portfolio Analysis in Terms of Means, Variances and Higher Moments," *Review of Economic Studies* **37** (1970), 537–542.

[115] Santos, Manual S., and Jean-Luc Vila, "Smoothness of the Policy Function in Continuous Time Economic Models: The One-Dimensional Case," mimeo, 1988.

[116] Sargent, Thomas J. "Estimation of Dynamic Labor Demand Schedules Under Rational Expectations," *Journal of Political Economy* **86** (1978), 1009-1044.

[117] Schumaker, Larry L. "On Shape-Preserving Quadratic Spline Interpolation," *SIAM Journal of Numerical Analysis* **20** (No. 4, August 1983): 854–864.

[118] Srikant, R., and T. Basar. "Iterative Computation of Noncooperative Equilibria in Nonzero-Sum Differential Games with Weakly Coupled Players," *Journal of Optimization Theory and Applications*, **71** (No. 1, October 1991): 137–168.

[119] Stokey, Nancy L., and Robert E. Lucas, Jr. *Recursive Methods in Economic Dynamics*, Harvard University Press: Cambridge. 1989.

[120] Taylor, John B., and Harald Uhlig, Solving nonlinear stochastic growth models: a comparison of alternative solution methods, *Journal of Business and Economic Statistics*, **8** (1990), 1–18.

[121] Treadway, A. B. "Adjustment Costs and Variable Inputs in the Theory of the Competitive Firm," *Journal of Economic Theory* **2** (December, 1970): 329-347.

[122] White, Halbert. *Artificial Neural Networks,* Blackwell Publishers: Cambridge. 1992.

[123] Wright, Brian D., and Jeffrey C. Williams. "The Economic Role of Commodity Storage," *Economic Journal* **92** (1982), 596-614.

[124] Wright, Brian D., and Jeffrey C. Williams. "The Welfare Effects of the Introduction of Storage," *Quarterly Journal of Economics* **99** (1984), 169-182.

[125] J. Williams, and B. Wright, *Storage and Commodity Markets*, Cambridge University Press, January, 1991.

[126] Wong, R. *Asymptotic Approximation of Integrals*, Academic Press: San Diego. 1989.

[127] Zadrozny, Peter. "Analytic Derivatives for Estimation of Dsicrete-Time, Linear Quadratic, Dynamic , Optimization Models," *Econometrica* **56** (1988), 467–472.

[128] Zeidler, E. *Nonlinear Functional Analysis and Its Applications: Volume I,* Springer-Verlag: New York, 1986.

[129] Zeidler, E. "Nonlinear Functional Analysis: Volume II," Springer-Verlag, New York, 1989.

[130] Zeldes, Stephen P. "Optimal Consumption with Stochastic Income Deviations from Certainty Equivalence," *Quarterly Journal of Economics* **104** (1989), 275–298.

[131] Zhang, Wei-Bin. "Hopf Bifurcations in Multisector Models of Optimal Economic Growth," *Economic Letters* **26** (4, 1988), 329-334.