Discussion Paper 107

Institute for Empirical Macroeconomics
Federal Reserve Bank of Minneapolis
250 Marquette Avenue
Minneapolis, Minnesota 55480-0291

December 1995

Accuracy Estimates for a Numerical Approach to Stochastic Growth Models

Manuel S. Santos*

Centro de Investigacion Economica
ITAM, Mexico

Jesus Vigo*

Facultad de Ciencias Economicas
Universidad de Salamanca, Spain

ABSTRACT

In this paper we develop a discretized version of the dynamic programming algorithm and derive error bounds for the approximate value and policy functions. We show that under the proposed scheme the computed value function converges quadratically to the true value function and the computed policy function converges linearly, as the mesh size of the discretization converges to zero. Moreover, the constants involved in these orders of convergence can be computed in terms of primitive data of the model. We also discuss several aspects of the implementation of our methods, and present numerical results for some commonly studied macroeconomic models.

*We would like to thank Jerry Bona, Richardo Cavalcanti, Larry Christiano, Pat Kehoe, Tim Kehoe, and Jinchao Xu for helpful conversations. We are also grateful for several discussions from seminar participants of the Federal Reserve Bank of Minneapolis and Universidad Carlos III de Madrid. The views expressed herein are those of the authors and not necessarily those of the Federal Reserve Bank of Minneapolis or the Federal Reserve System.

1. Introduction

In recent years, computational analysis has become an important tool for the study of economic models [see, for instance, recent survey work by Judd (1991), Kehoe (1991), Marcet (1994), Rust (1994), and Taylor and Uhlig (1990)]. This analysis generally provides an approximate solution of a given model, and such relevant quantitative information is not usually available from more traditional mathematical techniques. In order for numerical experiments to be effected rigorously, however, it is essential to have in hand error bounds or accuracy estimates of the computed solutions [cf., Bona and Santos (1994)]. As is commonly realized [e.g., Judd (1992, Sec. 7)], there has been very little theoretical work by economists on proving the accuracy of their numerical simulations.[1]

In this paper we consider a family of discrete-time stochastic growth models. As it is well known [cf. Stokey and Lucas (1989)], solutions to these models are most conveniently approached by the methodology of dynamic programming, as embodied in the value and policy functions. Via a value-function iteration algorithm, we develop a numerical discretization procedure to compute these fundamental functions. We prove that if h is the mesh size of the discretization then the approximation error for the value function is bounded by Mh^2, and the approximation error for the policy function is bounded by Nh, where M and N are positive constants that can be ascertained in terms of primitive data of the model as comprised in the objective function and technological constraints.

Our analysis is based upon differentiability properties of the value function. In deterministic models this function is of class C^2 under regular conditions [see Santos (1994) for a recent account of this topic]. For the class of models considered in this paper, a formal proof of the differentiability of class C^2 of the value function is not available. However, we shall show that a simple extension of the existing analysis is sufficient to validate such property in the present context.

[1] Li (1993) and Marcet and Marshall (1994) are notable exceptions to this trend. Li provides a rigorous derivation of error bounds for a simple monetary model. Marcet and Marshall prove certain asymptotic properties of a parameterized expectations algorithm in a standard growth model similar to our framework. Although Marcet and Marshall's convergence results are comforting, these results cannot be generally invoked in numerical computations since they only show that the approximation scheme reaches the exact solution as the computational cost goes to infinity.

Our approach to the differentiability problem yields upper estimates for the second-order derivatives of the value function in terms of primitive data of the model. These estimates can then be employed to bound the constants involved in the above orders of convergence. Much less is known, however, about higher-order derivatives of the value function; moreover, regular examples have been constructed where this function fails to be differentiable of class C^3 [cf. Araujo (1991) and Santos (1994)]. Hence, differentiability analysis suggests that without further specific assumptions higher orders of convergence for the value function beyond the quadratic one are not available.

Although our numerical procedures may yield fairly accurate solutions, they would be of limited interest if they cannot be implemented in practice. At a later stage of the paper we present some illustrative numerical computations, and show that our proposed methods perform relatively well in medium-scale models. Likewise, our accuracy estimates are also useful at several stages in the computation process, since they provide basic information for devising the discretization procedure, for testing the computer code, and for constructing continuation methods to speed up computations.

In implementing our methods we also discuss certain computational issues of some concern to us. More specifically, we focus on the following points: (a) tightness of our accuracy estimates and efficient ways to compute them numerically, (b) magnitude of the error involved in alternative discretizations, (c) efficiency of higher order approximations, (d) accuracy of some standard simulated results in real business cycle theory.

The remaining sections proceed as follows. In Section 2, we present a non-linear, stochastic model of economic growth. Our analysis begins in Section 3 with the differentiability properties of the value and policy functions. This is a necessary step to derive subsequently orders of convergence for the approximate value and policy functions computed from a dynamic programming algorithm. In Section 4 we implement our numerical procedures with the aid of some familiar examples. Some concluding comments follow in the final section.

2. The Model and Preliminary Considerations

We shall focus on a standard class of stochastic, reduced-form models of economic growth in which the solution to the optimal planning problem may be interpreted as the equilibrium law of motion of a decentralized economy. Our framework of analysis is encompassed in the class of economies set out in Stokey and Lucas (1989). The reader is referred to this monograph for several basic definitions and technical points raised in the course of our discussion.

Let $(K, \mathcal{K})$ and $(Z, \mathcal{Z})$ be measurable spaces, and let $(S, \mathcal{S}) = (K \times Z, \mathcal{K} \times \mathcal{Z})$ be the product space. The set K contains all possible values for the endogenous state variable, Z is the set of possible values for the exogenous shock, and S is the set of state values for the system. The evolution of the random component $\{z_t\}_{t\geq 0}$ is governed by a stochastic law defined by a function $\varphi : Z \times Z \to Z$ and an i.i.d. process $\{\varepsilon_t\}_{t\geq 1}$ where $z_t = \varphi(z_{t-1}, \varepsilon_t)$. It follows that the mapping φ induces a stationary transition function Q on $(Z, \mathcal{Z})$. Moreover, for each z_0 in Z one can define a probability measure $\mu^t(z_0, \cdot)$ on every t-fold product space $(Z^t, \mathcal{Z}^t) = (Z \times Z \times ... \times Z, \mathcal{Z} \times \mathcal{Z} \times ... \times \mathcal{Z})$ comprising all partial histories of the form, $z^t = (z_1, ..., z_t)$.

The technological constraints of the economy are summarized by a given feasible set $\Omega \subset K \times K \times Z$, which is the graph of a correspondence $\Gamma : K \times Z \to K$. The intertemporal objective is characterized by a one-period return function v on Ω and a given discount factor $0 < \beta < 1$. The optimization problem is to find a sequence of measurable functions $\{\pi_t\}_{t=0}^{\infty}$, $\pi_t : Z^{t-1} \to K$, as a solution to

$$W(k_0, z_0) = \sup_{\{\pi_t\}_{t\geq 0}} \sum_{t=0}^{\infty} \beta^t \int_{Z^t} v(\pi_t, \pi_{t+1}, z_t)\, \mu^t(z_0, dz^t)$$

$$\text{s. t. } (\pi_t, \pi_{t+1}, z_t) \in \Omega \tag{2.1}$$

$$z_t = \varphi(z_{t-1}, \varepsilon_t)$$

$$(k_0, z_0) \text{ fixed}, \pi_0 = k_0, \text{and } t = 0, 1, 2, ...$$

ASSUMPTION 1: *The set $K \times Z \subset R^l \times R^m$ has non-empty interior, and $\mathcal{K} \times \mathcal{Z}$ is the Borel-algebra of subsets of $K \times Z$. The set Ω is closed, and for each fixed z the projection $\Omega_z = \{(k,k') \mid (k,k',z) \in \Omega\}$ is convex and varies continuously with z.*

ASSUMPTION 2: *The mapping $v : \Omega \to R$ is bounded, continuous, and on the interior of its domain it is differentiable of class C^2 with bounded first- and second-order derivatives. Moreover, for all fixed z there exists some constant $\alpha > 0$ such that $v(k,k',z) + \frac{\alpha}{2}\|k'\|^2$ is concave as a function on (k,k').*

ASSUMPTION 3: *For each $(k_0,z_0) \in int(K \times Z)$ there exists an optimal solution to (2.1) such that every optimal realization $\{k_t,z_t\}_{t\geq 0}$ has the property that $(k_t,k_{t+1},z_t) \in int(\Omega)$ for each $t \geq 0$.*

ASSUMPTION 4: *The function $\varphi : Z \times Z \to Z$ is continuous, and for each fixed ε, the mapping $\varphi(\cdot,\varepsilon)$ is C^2 and the derivative functions $D_1\varphi(z,\varepsilon)$ and $D_{11}\varphi(z,\varepsilon)$ are bounded and jointly continuous over all points (z,ε) in $Z \times Z$. Also, there are non-negative constants $0 \leq \rho \leq 1$ and $C \geq 0$ such that the first- and second-order partial derivatives of z_t with respect z_0, $\frac{\partial z_t}{\partial z_0}$ and $\frac{\partial^2 z_t}{\partial z_0^2}$, have the property that $\left\|\frac{\partial z_t}{\partial z_0}\right\| \leq C\rho^t$ and $\left\|\frac{\partial^2 z_t}{\partial z_0^2}\right\| \leq C\rho^t$ for each $t > 0$.*

These assumptions are entirely standard and are usually presumed to hold over a certain compact domain which comprises the asymptotic dynamics of the optimal law of motion [cf., Santos (1994), and Stokey and Lucas (1989)]. In Assumption 2, the norm $\|k'\|$ is the usual Euclidean norm. Hence, such an assumption imposes a strong form of concavity on the second component of the utility function [cf., Montrucchio (1987)]. Observe that over compact sets the assumption is weaker than the conventional requirement that the Hessian matrix $D^2v_z(k,k')$ be negative definite[2] over all points (k,k',z) in Ω. The interiority postulate asserted in Assumption 3 is necessary to establish subsequently the smoothness of optimal paths. The examples below enhance the applicability of this condition. Regarding Assumption 4 note that $z_t = \varphi(\varphi(\cdots(\varphi(z_0,\varepsilon_1),\varepsilon_2)\cdots),\varepsilon_t)$.

[2]For functions v over a set $\Omega \subset R^l \times R^l \times R^m$, $Dv(k_0,k_1,z)$ will denote the (first-order) derivative of v evaluated at an interior point (k_0,k_1,z), and $D_iv(k_0,k_1,z)$, $i = 1,2,3$, will denote the partial derivative of v with respect to the ith component variable. Similarly, $D_{ij}v(k_0,k_1,z)$ will denote a second-order partial derivative of v with respect to the ith and jth components. Sometimes we will use the notation $D^2v_z(k_0,k_1)$ to represent the Hessian matrix of the mapping $v(\cdot,\cdot,z)$, where z is held fixed.

The assumption then requires the existence of some constants $0 \leq \rho \leq 1$ and $C \geq 0$ such that the matrix norms $\left\| \frac{\partial}{\partial z_0} \varphi\left(\varphi\left(\cdots\left(\varphi\left(z_0, \varepsilon_1\right), \varepsilon_2\right)\cdots\right), \varepsilon_t\right) \right\| \leq C\rho^t$ and $\left\| \frac{\partial^2}{\partial z_0^2} \varphi\left(\varphi\left(\cdots\left(\varphi\left(z_0, \varepsilon_1\right), \varepsilon_2\right)\cdots\right), \varepsilon_t\right) \right\| \leq C\rho^t$ for every realization $(\varepsilon_1, ..., \varepsilon_t)$ and $t > 0$.

The value function $W(k_0, z_0)$, given in (2.1), is well defined and jointly continuous on $K \times Z$ [cf., Stokey and Lucas (1989)]. Moreover, for each fixed z_0 the mapping $W(\cdot, z_0)$ is concave, and satisfies the Bellman equation

$$W(k_0, z_0) = \sup_{k_1} v(k_0, k_1, z_0) + \beta \int_Z W(k_1, z_1) Q(z_0, dz_1) \tag{2.2}$$

$$\text{s. t. } (k_0, k_1, z_0) \in \Omega$$

The optimal value $W(k_0, z_0)$ is attained at a unique point given by the policy function $k_1 = g(k_0, z_0)$. The policy function is also continuous. Furthermore, an iterated substitution on the right-hand side of (2.2) shows that the set of optimal contingency plans $\{k_t, z_t\}_{t \geq 0}$ is a Markov process determined by the optimal policy $k_{t+1} = g(k_t, z_t)$.

For the purposes of a later development, we recall that the value function W can be obtained as the unique fixed point of the following dynamic programming algorithm. Let $\mathcal{W}$ be the space of bounded, continuous functions V on the state space $K \times Z$ endowed with the norm $\|V\| = \sup_{(k,z) \in K \times Z} |V(k, z)|$. Define the (non-linear) operator $T : \mathcal{W} \to \mathcal{W}$, given by

$$T(V)(k_0, z_0) = \sup_{k_1} v(k_0, k_1, z_0) + \beta \int_Z V(k_1, z_1) Q(z_0, dz_1) \tag{2.3}$$

$$\text{s. t. } (k_0, k_1, z_0) \in \Omega$$

for $V \in \mathcal{W}$. It is well known that T is a contractive mapping on $\mathcal{W}$ with modulus $0 < \beta < 1$, i.e. $\|TV_0 - TV_1\| \leq \beta \|V_0 - V_1\|$ for $V_0, V_1 \in \mathcal{W}$. It follows that W is the unique fixed point under T, and $\|W - V_n\| \leq \beta^n \|W - V_0\|$ for $V_n = T^n V_0$, where T^n denotes the n-times composition of T.

Finally, we close these preliminaries with a discussion of the differentiability of class C^1 of W. It follows that under the above conditions [e.g., see Stokey and Lucas (1989)] at every interior point (k_0, z_0) the function W is differentiable with respect to k, and the partial derivative is given by the familiar envelope condition

$$D_1 W(k_0, z_0) = D_1 v(k_0, k_1, z_0) \tag{2.4}$$

where k_1 is the optimal point. Moreover, a straightforward calculation allows us to show that

$$D_2 W(k_0, z_0) = \sum_{t=0}^{\infty} \beta^t \int_{Z^t} \left[D_3 v(k_t, k_{t+1}, z_t) \cdot \frac{\partial z_t}{\partial z_0} \right] \mu^t(z_0, dz^t) \tag{2.5}$$

where again the right-hand side of (2.5) is evaluated at the optimal contingency plan $\{k_t, z_t\}_{t \geq 0}$. Hence, the function W is a C^1 mapping on $\text{int}(K \times Z)$. Consequently, at interior points the optimal policy $k_1 = g(k_0, z_0)$ can be characterized by the first-order condition

$$D_2 v(k_0, k_1, z_0) + \beta \int_Z D_1 W(k_1, z_1) Q(z_0, dz_1) = 0 \tag{2.6}$$

Under the foregoing assumptions one readily shows from a mere application of the implicit function theorem to (2.6) that if W is differentiable of class C^2 then g is a C^1 mapping. The converse result is not so straightforward as (2.5) involves an infinite series of partial derivatives.

3. Theoretical Results

We begin this section with an analysis of the second-order derivatives of the value function. Our method of proof yields a direct calculation of these derivatives in terms of primitive data of the model. We then consider a numerical procedure to compute the value function based upon a dynamic programming algorithm defined over a discrete state space of mesh size h. Under the maintained assumptions, we demonstrate that the computed value function W^h converges quadratically to the true value function W as h goes to zero. An analogous result involving linear convergence is obtained for the corresponding policy function g^h. The constants involved in these orders of convergence can be bounded from primitive parameters of the model.

3.1. Second-Order Differentiability of the Value Function

In a discrete-time stochastic growth framework, the second-order differentiability of the value function has been studied in Blume, Easley and O'Hara (1982) and Gallego (1993). The assumptions imposed in Blume, Easley and O'Hara are fairly strong for the present context. These authors postulate rather stringent separability and invertibility conditions over the stochastic process as well as smooth density functions. Gallego (1993) considers a broader class of stochastic innovations following the method of proof in Santos (1994). Here, random variables may be discrete, or may not posses smooth density functions, allowing thus for event trees and other commonly studied models with uncertainty. It should be stressed, however, that all these papers study the differentiability of the value function with respect to the endogenous state space K, and for the purposes of our present analysis we need to establish the joint differentiability of W over the entire state space $K \times Z$. This latter problem is closely related to the joint differentiability of W over a space of parameters [cf. Kehoe, Levine and Romer (1990) and Santos (1992)]. In Part I of the Appendix, we indicate how the arguments in Santos (1992) can be extended to validate the present results.

THEOREM 3.1: *Under Assumptions (1)-(4) the value function W is a C^2 mapping in int$(K \times Z)$.*

THEOREM 3.2: *Under Assumptions (1)-(4) the policy function g is a C^1 mapping in int$(K \times Z)$. Moreover there is a constant $L > 0$ such that* $\sum_{t=1}^{\infty} \beta^t \int_{Z^t} \|D_1 g^{t+1}(k_0, z_0)\|^2 \mu^t(z_0, dz^t) \leq \frac{L}{\alpha}$.

We shall now provide a direct method to bound the derivatives of these functions in terms of the underlying data of the model. As illustrated below, these computations are useful for specific applications. It is shown in the Appendix that the second-order own derivative of W with respect to the endogenous state variable k is determined by the following quadratic optimization problem

$$x_0 \cdot D_{11} W(k_0, z_0) \cdot x_0 = \sup_{\{\pi_t\}_{t \geq 0}} \sum_{t=0}^{\infty} \beta^t \int_{Z^t} [(\pi_t, \pi_{t+1}) \cdot D^2 v_{z_t}(k_t, k_{t+1}) \cdot (\pi_t, \pi_{t+1})] \mu^t(z_0, dz^t) \tag{3.1}$$

$$\text{s. t. } z_t = \varphi(z_{t-1}, \varepsilon_t),\ t = 1, 2, \ldots$$

Here the maximization proceeds over all measurable functions $\{\pi_t\}_{t=0}^{\infty}$, $\pi_t : Z^{t-1} \to R^l$, for $t \geq 1$ with $\pi_0 = x_0$ fixed, the one-period objective $D^2 v_{z_t}(k_t, k_{t+1})$ is the

Hessian matrix of the mapping $v(\cdot,\cdot,z_t)$ for z_t fixed, and $\{k_t, z_t\}_{t=0}^{\infty}$ is the optimal contingency plan to (2.1) for the initial value (k_0, z_0).

From this characterization, one readily proves [cf. Santos (1994)] that the optimal plan $\{\widehat{\pi}_t\}_{t=0}^{\infty}$ to maximization problem (3.1) determines the derivative of the policy function with respect to k_0. That is, $\widehat{\pi}_t = D_1 g^t(k_0, z_0) \cdot \pi_0$ for $t \geq 1$, where $D_1 g^t(k_0, z_0)$ denotes the derivative of the function $g(g(...g(k_0, z_0), ...), z_{t-2}), z_{t-1})$ with respect to k_0 for every possible realization $(z_1, z_2, ..., z_{t-1})$. Given that $(x_0, 0, 0, 0, ...)$ is a feasible solution to maximization problem (3.1), we obtain that

$$\|D_{11}W(k_0, z_0)\| \leq \|D_{11}v(k_0, k_1, z_0)\| \leq L \tag{3.2}$$

where $L = \sup_{(k_0,k_1,z_0)\in\Omega} \|D_{11}v(k_0, k_1, z_0)\|$. Moreover, if $\{\widehat{\pi}_t\}_{t=0}^{\infty}$ is an optimal solution to (3.1) with $\|\widehat{\pi}_0\| = 1$, then in view of the asserted concavity of $v(k, k', z)$ we must have

$$\sum_{t=0}^{\infty} \beta^t \int_{Z^t} [\widehat{\pi}_{t+1} \cdot \widehat{\pi}_{t+1}]\, \mu^t(z_0, dz^t) \leq \tfrac{L}{\alpha} \tag{3.3}$$

where $\widehat{\pi}_{t+1} \cdot \widehat{\pi}_{t+1}$ denotes the inner product multiplication at every possible value of the random vector $\widehat{\pi}_{t+1}$. Observe that (3.3) places an upper bound on the exponential growth factor of the derivative $\widehat{\pi}_t = D_1 g^t(k_0, z_0) \cdot \pi_0$. Indeed, this inequality implies that

$$\sum_{t=0}^{\infty} \beta^t \int_{Z^t} \left\| D_1 g^{t+1}(k_0, z_0) \right\|^2 \mu(z_0, dz^t) \leq \frac{L}{\alpha} \tag{3.4}$$

Differentiation of $D_2 W(k_0, z_0)$ in (2.5) with respect to k_0 yields

$$D_{12}W(k_0, z_0)^T = D_{21}W(k_0, z_0)$$

$$= \sum_{t=0}^{\infty} \beta^t \int_{Z^t} \left[\left(\tfrac{\partial z_t}{\partial z_0} \right) \cdot (D_{31}v(k_t, k_{t+1}, z_t) \cdot D_1 g^t(k_0, z_0) \right.$$

$$\left. + D_{32}v(k_t, k_{t+1}, z_t) \cdot D_1 g^{t+1}(k_0, z_0)) \right] \mu^t(z_0, dz^t)$$

Now, taking matrix norms we have

$$\left\|D_{21}W(k_0,z_0)^T\right\| = \|D_{12}W(k_0,z_0)\|$$

$$\leq \sum_{t=0}^{\infty}\beta^t \int_{Z^t}\left\|\frac{\partial z_t}{\partial z_0}\right\|\left[\|D_{31}v(k_t,k_{t+1},z_t)\|\,\|D_1g^t(k_0,z_0)\|+\right.$$

$$\left.\|D_{32}v(k_t,k_{t+1},z_t)\|\,\|D_1g^{t+1}(k_0,z_0)\|\right]\mu^t(z_0,dz^t)$$

$$\leq \left[1+\frac{2C\left(\frac{L}{\alpha}\right)^{\frac{1}{2}}}{1-\beta^{\frac{1}{2}}\rho}\right]G \tag{3.5}$$

where the last inequality follows from Assumption 4 and condition (3.4) for $G = \sup\{\|D_{31}v\|, \|D_{32}v\|\}$. Similar upper bounds can be obtained for $\|D_2g\|$ and $\|D_{22}W\|$. Observe from this analysis that in general variable z will have a more pronounced effect on the second-order derivatives of the value function.

3.2. Accuracy Estimates

Our purpose now is to present a computational method of W via a discretization of the state space, $S = K \times Z$, by means of finite-element techniques. The computational procedure is based on iterating the Bellman equation (2.2), and parallels some recent developments in the optimal control literature [e.g., Falcone (1987)]. Unlike the preceding literature, however, we shall establish the quadratic convergence of the sequence of approximate value functions, and devise a numerical scheme to implement the proposed method.

In the sequel we assume that the state space S is a polyhedron. This does not entail much loss of generality for most economic applications. Let $\{S^j\}$ be a family of simplices which conform a *triangulation* of S [i.e., $\cup_j S^j = S$ and $\text{int}(S^i) \cap \text{int}(S^j) = \phi$ for every pair of simplices S^i, S^j].[3] Let

$$h = \sup_j \text{ diam}\{S^j\}$$

[3]This kind of subdivision is not necessary for our results. For instance, rectangular subdivisions may sometimes be more suitable to certain applications. Both types of subdivisions are discussed in Brenner and Scott (1994).

Let (k^j, z^j) be a generic vertex of the triangulation. Consider the space of piecewise affine functions

$$\mathcal{W}^h = \left\{ V^h : S \to R \mid V^h \text{ is bounded, continuous and } DV^h \text{ is constant in int}\,(S^j) \text{ for each } S^j \right\}$$

Observe that $\mathcal{W}^h$ is a closed subspace of $\mathcal{W}$ equipped with the norm $\left\| V^h \right\| = \sup_{(k,z)\in K\times Z} \left| V^h(k,z) \right|$ for $V^h \in \mathcal{W}^h$.

Define the mapping $T^h : \mathcal{W} \to \mathcal{W}^h$, given by

$$\begin{aligned} T^h(V)(k_0^j, z_0^j) = \sup_{k_1} \; & v(k_0^j, k_1, z_0^j) + \beta \textstyle\int_Z V(k_1, z_1) Q(z_0^j, dz_1) \\ & \text{s. t. } (k_0^j, k_1, z_0^j) \in \Omega \end{aligned} \tag{3.6}$$

$$\text{for each vertex point } (k_0^j, z_0^j) \text{ and } V \in \mathcal{W}^h$$

Observe that the maximization and integration operations in (3.6) are performed exactly. Also, nodal values $T^h(V)(k_0^j, z_0^j)$ for all vertex points (k_0^j, z_0^j) yield a unique extension to the whole domain S over the space of piecewise affine functions compatible with a given triangulation $\{S^j\}$.

The following equation will play a central role in our analysis.

$$\begin{aligned} W^h(k_0^j, z_0^j) = \sup_{k_1} \; & v(k_0^j, k_1, z_0^j) + \beta \textstyle\int_Z W^h(k_1, z_1) Q(z_0^j, dz_1) \\ & \text{s. t. } (k_0^j, k_1, z_0^j) \in \Omega \end{aligned} \tag{3.7}$$

$$\text{for each vertex point } (k_0^j, z_0^j)$$

This is the corresponding discretized version of Bellman's functional equation (2.2).

LEMMA 3.3: *Under Assumptions (1)-(4) equation (3.7) has a unique solution* W^h *in* $\mathcal{W}^h$.

PROOF: The proof is the standard one [cf., Stokey and Lucas (1989)]. One immediately sees that T^h is a contraction mapping with modulus $0 < \beta < 1$. By

a well known fixed-point theorem, equation (3.7) has a unique fixed point W^h in $\mathcal{W}^h$.

LEMMA 3.4: *Let W be the value function defined in (2.1). Let $\gamma = \|D^2W\|$. Then under Assumptions (1)-(4) it must hold that $\left\|TW - T^hW\right\| \leq \frac{\gamma}{2}h^2$.*

PROOF: The proof is also standard, and it is based upon an application of Taylor's theorem. We shall follow closely the arguments sketched in Johnson (1987, Ch. 4).

Note that by the definitions of T and T^h we have that $TW(k^j, z^j) = W(k^j, z^j) = T^hW(k^j, z^j)$ for every vertex point (k^j, z^j), and that the function T^hW is piecewise affine. Consider now an arbitrary interior point (k, z) in a given simplex S^j with nodal points $\{(k^i, z^i)\}_{i=0}^{l+m}$. For convenience, let us temporarily use the notation, $x = (k, z)$ and $x^i = (k^i, z^i)$ for all i. Then x can be uniquely expressed as $x = \sum_{i=0}^{l+m} \lambda^i(x)x^i$ for some set of non-negative weights, $\lambda^i(x)$, such that $\sum_{i=0}^{l+m} \lambda^i(x) = 1$. Also, $T^hW(x) = \sum_{i=0}^{l+m} \lambda^i(x)W(x^i)$. Let $x^i(t) = x^i + t(x - x^i)$, for $0 \leq t \leq 1$. Then by Taylor's theorem

$$W(x^i) - W(x) = DW(x) \cdot (x^i - x) + \int_0^1 (x^i - x) \cdot D^2W(x^i(t)) \cdot (x^i - x)tdt$$

It follows then from the above definitions that

$$\begin{aligned} \left|W(x) - T^hW(x)\right| &= \left|\sum_{i=0}^{l+m} \lambda^i(x) \int_0^1 (x^i - x) \cdot D^2W(x^i(t)) \cdot (x^i - x)tdt\right| \\ &\leq \frac{\gamma}{2}\left(\sum_{i=0}^{l+m} \lambda^i(x) \left\|x^i - x\right\|^2\right) \end{aligned} \tag{3.8}$$

Now, one readily sees that a non-necessarily tight, upper estimate for (3.8) is $\frac{\gamma}{2}h^2$. The result is thus established.

THEOREM 3.5: *Let W be the fixed point of equation (2.2) and W^h be the fixed point of equation (3.7). Then under Assumptions (1)-(4) it must hold that $\left\|W - W^h\right\| \leq \frac{\gamma}{2(1-\beta)}h^2$.*

PROOF: Let T and T^h be as defined previously from (2.3) and (3.6), respec-

tively. Then

$$\left\|W - W^h\right\| = \left\|TW - T^hW^h\right\|$$

$$\leq \left\|TW - T^hW\right\| + \left\|T^hW - T^hW^h\right\|$$

$$\leq \left\|TW - T^hW\right\| + \beta\left\|W - W^h\right\|$$

where use is made in these computations of the triangle inequality and of Lemma 3.3. Therefore,

$$\left\|W - W^h\right\| \leq \tfrac{1}{1-\beta}\left\|TW - T^hW\right\|$$

Theorem 3.5 is now a direct consequence of Lemma 3.4.

COROLLARY 3.6: *Let $g(k^j, z^j)$ be the optimal policy for the original value function W at a vertex point (k^j, z^j), and let $g^h(k^j, z^j)$ be the optimal policy for the approximate value function W^h at vertex point (k^j, z^j). Then $\left\|g(k^j, z^j) - g^h(k^j, z^j)\right\| \leq \left(\frac{2\gamma}{\alpha(1-\beta)}\right)^{\frac{1}{2}} h$, for every vertex point (k^j, z^j).*

PROOF: Let $(k^j, z^j) \in S$ be a generic vertex point. Assume that $k_1 = g(k^j, z^j)$ and $k_1^h = g^h(k^j, z^j)$. For (k^j, z^j) fixed, let
$f(k) = v(k^j, k, z^j) + \beta \int_Z W(k, z')Q(z^j, dz')$. By Theorem 3.5, it must then hold true that

$$f(k_1) - f(k_1^h) \leq \tfrac{\gamma}{1-\beta}h^2 \tag{3.9}$$

Likewise,

$$\tfrac{\alpha}{2}\left\|k_1 - k_1^h\right\|^2 \leq f(k_1) - f(k_1^h) \tag{3.10}$$

as $Df(k_1) = 0$, and by Assumption 2 the function $f(k) + \frac{\alpha}{2}\|k\|^2$ is concave. Inequalities (3.9) and (3.10) combined together imply then that

$$\left\|k_1 - k_1^h\right\| \leq \left(\tfrac{2\gamma}{\alpha(1-\beta)}\right)^{\frac{1}{2}} h$$

Since (k^j, z^j) is an arbitrarily chosen vertex point in S, this proves the result.

REMARKS:

1. Inequalities (3.2)-(3.5) provide upper estimates for parameter $\gamma = \|D^2 W\|$. These estimates can be useful to bound the observed error in specific applications.

2. Constant $\dfrac{\gamma}{2(1-\beta)}$ becomes unbounded for $\beta = 1$. This singularity seems to be related to the fact that if the value of the approximation error in a single period may be up to $\dfrac{\gamma}{2}h^2$ (Lemma 3.4), then the cumulative error over the entire infinite horizon may extend up to $\dfrac{\gamma}{2(1-\beta)}h^2$.

3. Observe that in the above approximation, we have considered the space $\mathcal{W}^h$ of piecewise linear functions over a family $\{S^j\}$ of $(l+m+1)$-dimensional simplices. For certain applications, it may be more convenient to consider piecewise multi-linear functions over rectangular subdivisions. One can show that similar asymptotic results are also valid in this latter case.

4. If W and g are differentiable to a higher order, then it is possible to establish higher orders of convergence under piecewise polynomials of a higher degree.[4] Under regular assumptions, however, function W is not necessarily a C^3 mapping [Araujo (1991) and Santos (1994)]. Moreover, if higher-order derivatives become arbitrarily large, then more complex approximations do not necessarily yield better estimates [e.g., Stoer and Bulirsch (1993, p.51)].

5. It should be emphasized that for each of the vertex points (k^j, z^j) the integration and maximization operations in (3.6) must be accomplished exactly. If alternatively the maximization is carried out over a set of grid points $\{k^j\}$ with mesh size h, then one can show from a suitable reformulation of Lemma

[4] Of particular interest for certain applications is the approximation of function $W(k, z)$ by a C^1 mapping $\widehat{W}(k, z)$ with $\widehat{W}(k^j, z^j) = W(k^j, z^j)$ and $D\widehat{W}(k^j, z^j) = DW(k^j, z^j)$ at all vertex points (k^j, z^j). Observe that in the deterministic case such derivatives are easily computed from the one-period return function evaluated at the optimal solution [cf. eq. (2.4)]. Under this type of Hermite interpolant, convergence may be of order h^4 [Stoer and Bulirsch (1993, Th. 2.1.5.10)].

3.4 that the additional error involved in the computed value function is also of order h^2, since the first-order derivative at a maximizer k_1 is equal to zero. Likewise, if the integration is performed over a discretized space as an approximation for an underlying continuous-valued random variable, then the additional error involved in this approximation will depend on the integration scheme.[5] Our analysis therefore should be useful for making reasonable choices concerning discretizations of the state spaces K and Z, since the involved approximation errors may have different asymptotic behaviors. It also illustrates that commonly found computations which restrict the uncertainty space Z to very few states as compared to the space of capitals K [e.g., Christiano (1990), Danthine, Donaldson and Mehra (1989), and Dotsey and Mao (1992)] may obtain more accurate approximations for the same computational cost by considering more balanced grids over the whole space $K \times Z$.

6. The linear convergence of Corollary 3.6 can readily be extended to any point (k, z) in the domain $K \times Z$.

4. A Numerical Analysis of Some Growth Models

We shall now discuss a numerical algorithm to implement our computational approach in the context of some standard growth models. Most of our discussion will focus on the error involved in the orders of convergence and related theoretical aspects of our numerical procedure.

All data were generated using an DEC 2000 workstation, 300 ALPHA AXP with 358.1 MFLOPS, coded in standard FORTRAN 77, which in a double precision floating-point arithmetic allows for a sixteen-digit accuracy. The resulting programs constitute a package that we have named NIBS (Numerical Iterative Bellman Scheme). Subject to the aforementioned upper bound, our algorithm is in principle free from a fixed level of accuracy. Such level of accuracy is determined by the following parameter values:

[5]For instance, if the integration over the discretized uncertainty space Z mimics the trapezoidal rule, then the error is of order h^3 [cf., Press et al. (1992, p. 125)].

h = mesh size

$TOLW$ = Accuracy imposed in the iterative scheme: the program stops if given two consecutive value functions W_n^h and W_{n+1}^h the difference $\left\|W_n^h - W_{n+1}^h\right\| \leq TOLW$

$TOLI$ = Accuracy attained in integration

$TOLM$ = Accuracy attained in maximization

Of course, in order to implement efficiently the numerical algorithm, judicious choices of these parameters should be made in accordance with our theoretical analysis of the error involved in these approximations. For example, it would not be optimal to achieve an arbitrary high degree of accuracy in the maximization process ($TOLM$) for cases of a coarse discretization of the state space (large h). In our particular examples, however, in order to compare different numerical experiments, and analyze more neatly the evolution of the approximation error stemming from changes in h, we have always fixed both $TOLI$ and $TOLM$ to an eight-digit precision.

We would like to emphasize that the maximization and integration operations are performed with the aid of standard, well-tested codes. These procedures are explained in Part II of the Appendix. In our numerical experiments, all integrals are computed numerically under the same methodology, since all integrations involve a unique random variable. Regarding maximizations, however, we employ two different methods depending upon the dimensionality of the problem. For one-dimensional maximization we apply Brent's algorithm, whereas for multivariate maximization we use a version of the quasi-Newton method.

In order to effect the check for $TOLW$, the algorithm must provide at each iteration the difference between two consecutive value functions. Of course, by virtue of Lemma 3.3 –and modulo the approximation errors determined by both $TOLI$ and $TOLM$– this difference must decrease in each iteration by the factor $0 < \beta < 1$. This property is also embedded in our computations as an additional way to test the code. For simplicity, in all our computations, $TOLW$ has been set equal to h^2. Given the dynamic properties of our algorithm, this seems a reasonable value.

For illustrative purposes our study has been restricted to three standard growth models with closed form solutions. Example 1 considers a one-sector deterministic

growth model. A stochastic version of this model is the content of Example 2. Finally, in Example 3 the model is expanded to include leisure in the analysis.

Example 1. *A One-Sector Deterministic Growth Model.* If the sequence of random variables $\{\varepsilon_t\}_{t\geq 0}$ is degenerate, then our framework reduces to a deterministic growth model. Within this family of deterministic growth processes, we now consider a well known example in which the value and policy functions have closed form solutions. Such functional forms will serve as benchmark for our computations. The optimization problem is written as

$$W(k_0) = \max_{\{k_t\}} \sum_{t=0}^{\infty} \beta^t \log c_t$$

$$\text{s. t. } c_t = Ak_t^{\alpha} - k_{t+1}$$

$$k_0 \text{ fixed, } 0 < \beta < 1,\ 0 < \alpha < 1,\ A > 0, \text{ and } t = 0, 1, 2, ...$$

where c and k are positive numbers. Under these conditions the system has a unique interior steady state, $k^* > 0$, which is globally stable. Such steady-state value, k^*, is the solution of the following equation

$$\frac{1}{\beta} = \alpha A k^{*\alpha - 1} \tag{4.1}$$

As is well known [e.g., see Kehoe (1991, Example 5.2), and Manuelli and Sargent (1987, Ch. 1)] the value function $W(k_0)$ takes the form $W(k_0) = B + C \ln k_0$, where B and C are constants such that $C = \dfrac{\alpha}{1 - \alpha\beta}$. Likewise, the policy function $k_1 = g(k_0)$ takes the form, $k_1 = \alpha\beta A k^{\alpha}$. Moreover, these functions can be obtained as the limiting solutions of the following iterative process. Let

$$W_n(k_0) = T[W_{n-1}(k_0)] = \max_{k_1} \ \log c_0 + \beta W_{n-1}(k_1) \tag{4.2}$$

$$\text{s. t. } c_0 = Ak_0^{\alpha} - k_1$$

for $n = 1, 2, ...$ and $W_0 \equiv 0$. The aforementioned references also provide analytic solutions for the sequence of functions $\{W_n\}_{n\geq 1}$ and $\{g_n\}_{n\geq 1}$. Such functional forms are useful to verify our computations as a way to test the computer code.

But possibly the most exacting test of the whole exercise is to show that the sequence of computed value functions $\{W_n^h\}_{n\geq 1}$ and $\{g_n^h\}_{n\geq 1}$ from (3.6) behaves asymptotically as predicted by the error analysis

We initially consider parameter values, $A = 5$, $\beta = 0.95$, $\alpha = 0.34$. For such values the stationary state, $k^* = 4.214$. For the purposes of this exercise, the domain of possible capitals, K, is restricted to the interval $[0.1, 10]$. Under these conditions it is then easy to check that Assumptions (1)-(3) are all satisfied.

Over the feasible interval of capital stocks, $[0.1, 10]$, we consider a uniform grid of points k^j with step size h. In this simple univariate case our interpolations yield concave, piecewise linear functions. As specified in (3.6), we carry out the numerical experiment under the iterative process $W_{n+1}^h = T^h(W_n^h)$, starting with an initial value $W_0^h \equiv 0$. The exact maximization at vertex points k^j in (3.6) is effected by Brent's algorithm with $TOLM = 10^{-8}$, as reported in Part II of the Appendix. The computer program is instructed to stop once two consecutive value functions $W_{n+1}^h = T^h(W_n^h)$ satisfy the inequality $\left\|W_{n+1}^h - W_n^h\right\| \leq TOLW = h^2$. Since T^h is a contractive operator with modulus $0 < \beta < 1$, the fixed point $W^h = T^h(W^h)$ in (3.7) should then lie within a distance $\left\|W^h - W_n^h\right\| \leq \dfrac{h^2}{1-\beta}$.

We start this numerical exercise with $h = 10^{-1}$ and the initial condition $W_0 \equiv 0$. In computing the approximate value function W^h for $h = 10^{-1}$ the program stops after 99 iterations with a reported CPU time of 3 seconds. We then proceed in the same manner with $h = 10^{-2}$ and $h = 10^{-3}$. Figures 1(a)-1(c) depict the observed error between the true and computed value functions, $e_{\hat{n}}^h(k) = \left|W(k) - W_{\hat{n}}^h(k)\right|$ for $h = 10^{-j}$ with $j = 1, 2, 3$, where as specified above the terminal iteration $\hat{n}$ is determined by the value $TOLW = h^2$. Also, Table 1(a) summarizes further relevant information concerning our numerical computations, including the evolution of the error of the policy function.

Observe that the constant for the value function stemming from our numerical computations takes values around 18, whereas the corresponding constant for the policy function takes values around 3.6. Both functions converge as predicted by our error analysis. The relatively small value for the error of the optimal policy seems to be due to the simple analytical form of such function.

In order to compare these numerical estimates with our previous theoretical

analysis we first decompose the error in the following form

$$e^h_{\hat{n}}(k) = \left|W(k) - W^h_{\hat{n}}(k)\right| \leq \left|W(k) - W^h(k)\right| + \left|W^h(k) - W^h_{\hat{n}}(k)\right| \qquad (4.3)$$

where $W^h = T^h(W^h)$ is the fixed-point given in (3.7). Thus, $\left|W(k) - W^h(k)\right|$ is the error resulting from our numerical algorithm, and $\left|W^h(k) - W^h_{\hat{n}}(k)\right|$ is the error resulting from stopping the iteration process in finite time. Since optimal solutions converge at a relatively fast rate to the steady-state value $k^* = 4.124$, an appropriate estimate of the value γ referred to in Lemma 3.4 should be the absolute value of the second-order derivative at the steady state, k^*; that is, $|D^2W(k^*)| = 0.02828$. Hence, $\dfrac{\gamma}{2(1-\beta)} = 0.2828$. In situations where the value function does not feature a close form solution, one can use instead our estimate $\widehat{\gamma}$ of the second-order derivative given in (3.2). At the given steady state such value, $\widehat{\gamma} = 0.16697$. Hence, $\dfrac{\widehat{\gamma}}{2(1-\beta)} = 1.6697$

Therefore, the quantity $0.2828h^2$ is an upper estimate for the term $\left|W(k) - W^h(k)\right|$ in (4.3) (or $1.6697h^2$ if we were to ignore the functional form of the value function).[6] Likewise, as discussed above an upper estimate for the term $\left|W^h(k) - W^h_{\hat{n}}(k)\right|$ in (4.3) is the quantity $\dfrac{1}{1-\beta}h^2 = 20h^2$. Hence, our upper bound for the observed error $e^h_{\hat{n}}(k)$ is 21.6697, whereas the observed error is in fact around $18h^2$. Thus, $e^h_{\hat{n}}(k)$ falls in the range imposed by our theoretical analysis.[7]

This difference between the observed error and our upper estimates is something to be expected in practice, since our results are meant to bound the error on a worst-case basis. Moreover, in some cases this difference may be considerably large. Hence, for situations where the problem does not feature an analytic solution, a more operational approach is to appraise the error numerically –instead of using our upper estimates of the error [equations (3.2)-(3.5)]. This procedure is suggested, for instance, in Bona and Santos (1994). One computes the value function W from a very fine grid, and considers such computation as the true

[6] In the univariate case, a sharper version of Lemma 3.4 is available with $\|TW - T^hW\| \leq \dfrac{\gamma}{8}h^2$; e.g., Stoer and Bulirsch (1993, Th. 2.1.4.1). Hence, in this case a better estimate, $\left|W(k) - W^h(k)\right| \leq 0.0707h^2$.

[7] Observe from the figures that the observed error $e^h_{\hat{n}}(k)$ decreases with k. This seems to be due to the fact that the second-order derivative of the value function is decreasing in k.

solution. Then one analyzes the behavior of the error for several values of h. If the constants involved in the orders of convergence settle down around some given values, then one may take these values as the effective constants for the problem at hand.

We also considered further variations of the preceding exercise that we now pass to summarize.

1. *Discount factors β close to 1*: Our theoretical analysis suggests that as β approaches 1, the constants involved in the order of convergence may become unbounded. For $\beta = 0.99$ and all the above parameter values, Figure 1(d) displays the observed error of the value function for $h = 10^{-3}$. In this case, the constant goes up to 96, which is a 5.4-fold increase with respect to the preceding figures. This is roughly the ratio, $\frac{\frac{1}{(1-0.99)}}{\frac{1}{(1-0.95)}} = 5$, predicted by our error analysis, upscaled by a further increase in the second-order derivative of function W (since $W(k) = B + C \log k$ with $C = \frac{\alpha}{1-\alpha\beta}$). Hence, one should expect these constants to become unbounded as β approaches unity. Table 1(b) reports in an analogous way further information regarding this numerical experiment with $\beta = 0.99$.

2. *Higher-order approximations.* Instead of the space $\mathcal{W}^h$ of piecewise linear functions, one could focus on alternative finite-dimensional functional spaces involving higher-order approximations. An approximation that seems suitable for our deterministic model is to interpolate over the space of piecewise polynomials of order 3, under the conditions that at each vertex point k^j the operator $T^h(W_n^h(k^j))$ must satisfy (3.6) and additionally the resulting function W_{n+1}^h is C^1 with derivative at every vertex point equal to $DW_{n+1}^h(k^j) = \frac{\partial}{\partial k_0} \log(Ak_0^\alpha - k_1^{n+1})$ where k_1^{n+1} is the optimal choice at the $(n+1)$-th iteration. Observe that if the stationary state k^* is globally stable, then turnpike theory entails that the value function is infinitely differentiable [Araujo and Scheinkman (1977)]. For the case $\beta = 0.99$, we found that this approximation resulted in a saving of computer time.

3. *Continuation Methods*: In our previous iterated computational procedure, all iterations are started with the initial condition $W_0 \equiv 0$, regardless of the partition $\{S^j\}$. The following continuation method turns out to be com-

putationally faster while considering a nested sequence of finer intervals.[8] Such procedure seems especially attractive for very fine subdivisions of the state space or for high values for β. The method goes as follows. Let h_i be the mesh size of the partition $\{S^j_{h_i}\}$ for $i = 0, 1, ..., n$. Suppose that $h_0 > h_1 > ... > h_n$. Then, as before, start with the coarsest partition $\{S^j_{h_0}\}$ and compute the value function W^{h_0} from the above iterated method (3.6) with $W_0 \equiv 0$. Next, choose W^{h_0} as the initial condition for computing W^{h_1}. And follow the same procedure for subsequent refinements: pick $W^{h_{i-1}}$ as the initial choice for computing W^{h_i}, for $i = 1, 2, ..., n$.

For the same level of accuracy, this method reduces the computational burden, since earlier iterations are performed over coaser grids with less computational cost. Furthermore, if estimates of the approximation errors are available, then one can derive more efficient algorithms over these successive approximations. For instance, by Corollary 3.6 the search for the optimal policy can be restricted to a smaller region while proceeding with finer partitions. For the example under study, Table 1(c) replicates the computations of Table 1(b) under the proposed continuation method. It can be seen that in this case with $\beta = 0.99$ this iterative procedure leads to about a twenty-five per cent saving of CPU time.

Example 2: *A One-Sector Stochastic Growth Model.* We next consider a stochastic version of the preceding model in which the value and policy functions also have a closed form solution. The problem is written as

$$W(k_0, z_0) = \max_{\{k_t\}} \sum_{t=0}^{\infty} \beta^t \int_{Z^t} [\log c_t] \mu^t(z_0, dz^t)$$

$$\text{s. t. } c_t = z_t A k_t^{\alpha} - k_{t+1}$$

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

$$k_0 \text{ and } z_0 \text{ fixed, } 0 < \beta < 1,\ 0 < \rho < 1,\ 0 < \alpha < 1, \text{ and } t = 0, 1, 2, ...$$

where ε_t is an i.i.d. process with zero mean. Under these conditions the value function W has an analytical form given by $W(k_0, z_0) = B + C \ln k_0 + D \ln z_0$,

[8] Complexity properties of this method have been studied in Chow and Tsitsiklis (1991).

where $C = \dfrac{\alpha}{1-\alpha\beta}$ and $D = \dfrac{1}{(1-\alpha\beta)(1-\rho\beta)}$. Also, as previously the optimal policy is a constant fraction of total production, $k_{t+1} = \alpha\beta z_t A k_t^{\alpha}$.

We fix parameter values, $A = 5$, $\beta = 0.95$, $\alpha = 0.34$, $\rho = 0.90$. Also, we restrict the feasible domain so that $k \in [0.1, 10]$, $\varepsilon \in [-0.032, 0.032]$, and z is such that $\log z \in [-0.32, .32]$. As explained in the Appendix, we assume that the random process ε comes from a normal distribution, where the density has been rescaled in order to get a cumulative mass equal to unity. As in Prescott (1986) we assume a standard deviation $\sigma_\varepsilon = 0.008$. Observe then that the end-points of the domain of variable ε are four standard deviations away from the mean.

One can again check that under these restrictions Assumptions (1)-(4) are satisfied. Indeed, in this simple case one can show that the model has a globally stable invariant distribution. As the random shock has a small variance, all paths eventually fluctuate around the point $(k^*, \bar{z}) = (4.214, 1)$, where k^* is the state value of the deterministic model and $\bar{z}$ is the unconditional mean of the random process. Consequently, an appropriate estimate for γ of Lemma 3.4 should be the matrix norm of the second-order derivative of W at the point $(k^*, \bar{z})$. That is, $\|D^2 W(k^*, \bar{z})\| = \gamma = 10.1869$. Hence, for given $h > 0$ the observed error for the computed value function stemming from our interpolation procedure must be bounded above by $\dfrac{\gamma}{2(1-\beta)} h^2 = 101.869 h^2$. As in the preceding example, to this estimate we should add the other component of the observed error concerning the fact that the iteration process is stopped in finite time.

Over the feasible domain of state variables we set out a uniform grid of vertex points (k^j, z^j) with mesh size h. Our numerical procedure then follows the iterative process specified in (3.6) with an initial value $W_0 \equiv 0$, and where the integration in (3.6) is carried out under Simpson's rule (see the Appendix) with an eight-digit accuracy. Again, the one-dimensional maximization problem is effected following Brent's algorithm. The iterative process stops once two consecutive value functions satisfy the given inequality $\left\| W^h_{n+1} - W^h_n \right\| \leq h^2$.

Figures 2(a)-2(c) portray the observed error $e^h_n(k,z) = \left| W(k,z) - W^h_n(k,z) \right|$ for several values for h. Likewise, Table 2(a) presents further details of the numerical experiment. It should be observed that the error term $e^h_n(k,z)$ is always bounded by $21h^2$. Hence, the constant stemming from our computations is

bounded above by 21, whereas our estimate of the observed error,

$$\begin{aligned}\left|W^h(k,z) - W^h_{\hat{n}}(k,z)\right| &\leq \left|W(k,z) - W^h(k,z)\right| + \left|W^h(k,z) - W^h_{\hat{n}}(k,z)\right| \\ &\leq 101.869h^2 + 20h^2 \leq 121.869h^2\end{aligned}$$

We again emphasize that this substantial difference is a result to be expected in particular applications, since these estimates are by construction rough upper bounds of the maximum approximation error over the entire infinite horizon.

As in the preceding example, we considered alternative numerical experiments for several values for β, and replicated the original computations under the continuation method. Table 2(b) reports the corresponding numerical results for discount factor $\beta = 0.99$. As it is to be expected, the constant involved in the order of convergence for the value function is now five times larger. Likewise, Table 2(c) replicates the computations of Table 2(b) under the continuation method. It can be seen that the required CPU time gets down roughly to one half of that of the original experiment.

Example 3: *A One-Sector Stochastic Growth Model with Leisure*: The model is extended now to include work and leisure. In reduced form, the optimization problem is written as

$$W(k_0, z_0) = \max \tfrac{1}{3}\log c_0 + \tfrac{2}{3}\log l_0 + \beta \int_Z W(k_1, z_1) Q(z_0, dz_1)$$

$$\text{s. t. } c_0 = z_0 A k_0^{\alpha}(1 - l_0)^{1-\alpha} - k_1$$

$$\log z_1 = \rho \log z_0 + \varepsilon_1$$

where ε is again the i.i.d. process of Example 2. Under these conditions, the value of leisure, l, remains constant over the optimal solution, independently of the state variable (k, z). Hence, the value function W takes on the same functional form as in the previous example with the same values for constants C and D. Hence, the second-order derivatives of this function remain unchanged.

For the purposes of the numerical maximization, the problem involves leisure as an additional decision variable. Hence, Brent's algorithm is not longer suitable for our purposes. The one-period maximization step in now effected by the ZXMWD

subroutine from IMSL.[9] As previously, Figures 3(a)-3(c) illustrate the observed error involved in our numerical procedure. Also, Table 3 reports further relevant details of our numerical experiment.

It should be noticed that the observed error $e^h_{\hat{n}}(k,z)$ is of the same magnitude as that of our previous example. That is, in both stochastic examples, $e^h_{\hat{n}}(k,z) \leq 21h^2$. This result may come as no surprise, given that both value functions have identical second-order derivatives.

In addition to the analysis of the error, we have also examined the accuracy of quadratic approximations in this class of models. This is a topic of particular concern in the business cycles literature. As shown in (3.1), the second-order derivative $D_{11}W(k_0,z_0)$ can be characterized as the value function of a quadratic optimization problem. Moreover, the derivative of the policy function $D_1g(k_0,z_0)$ is defined by the linear decision rule of the quadratic model. On the other hand, the remaining partial derivatives of the value and policy functions depend in a nonlinear way on the uncertainty component [e.g., see equations (6.3) and (6.6) below]. Hence, we should expect solutions of the quadratic model to be good approximations of the non-linear dynamics for small stochastic perturbations near a deterministic steady-state. However, if these stochastic shocks are large then accuracy is no longer guaranteed.[10]

We should also point out that the usual quadratic approximations found in the literature consider the matrix of second-order derivatives $D^2v(k_t,k_{t+1},z_t)$ as the one-period objective function, whereas (3.1) shows that the right approximation is defined instead by the matrix of derivatives $D^2v_{z_t}(k_t,k_{t+1})$ with z_t held fixed. For large stochastic shocks both quadratic approximations may yield different solutions.

For the above example, the variance of the stochastic innovation is small. Hence, the quadratic approximation should mimic reasonably well the invariant distribution of the non-linear model with or without physical capital depreciation. For convenience of the analysis, our computations will be restricted to the above model with full physical capital depreciation, although similar results are available for cases of more realistic calibrations of capital depreciation.[11] As it is typical in

[9] In order to implement this subroutine, all functions are rewritten so as to depend on l, although for some functions such dependence is merely fictitious.

[10] Judd (1991) has made a similar point, but without a formal derivation of these derivatives.

[11] In cases where the value function has no analytic form comparisons were made with respect to the numerical solutions of our algorithm for arbitrary accuracy levels.

the business cycles literature, we consider the full matrix of second-order derivatives, $D^2v(k_t, k_{t+1}, z_t)$, as the one-period objective of the quadratic model [instead of the correct approximation (3.1)].

Table 4 reports the standard derivations and correlation coefficients for physical capital, k, consumption, c, work, $u = 1 - l$, and output, $y = Ak^{\alpha}u^{1-\alpha}$. These moments have been obtained from ten-thousand draws of the *i.i.d.* random process $\{\varepsilon_t\}$ that enabled us to construct a random path for state variable z. For the calibrated standard deviation of the stochastic innovation, these computations confirm the accuracy of the second-order moments obtained from the quadratic approximation. Also, we have checked that both models practically generate the same optimal paths, and so they yield similar predictions concerning the quantitative behavior of the level variables. Of course, such accuracy is not expected to hold for larger values of the standard deviation of the *i.i.d.* process $\{\varepsilon_t\}$.

5. Concluding Remarks

In this paper we have considered a standard class of models of economic growth. Via a discretized version of the dynamic programming algorithm, we have derived error bounds for the computed solutions. We have shown that the approximate value function converges quadratically to the true value function and that the approximate policy function converges linearly to the true policy function, as the mesh size of the discretization converges to zero. Also, with the aid of well-tested subroutines we have implemented numerically our computational method corroborating that the observed approximation error evolves asymptotically as predicted by our theoretical analysis.

Derivation of error bounds is a basic topic of theoretical numerical analysis. This topic is not as yet rooted in the economic literature, and relatively little is known about the accuracy properties of most of the algorithms used by economists. To the best of our knowledge, our study provides the first rigorous analysis of error bounds for the standard class of neoclassical growth models. This analysis guarantees arbitrarily accurate approximations, and should be important to settle controversial issues regarding subtle properties of a given solution and to appraise the accuracy of alternative numerical methods. Likewise, as illustrated in Section 4 these results are also useful for testing the computer code and for devising efficient computational procedures.

Although our algorithm has some desired accuracy properties, it may be relatively costly to implement in certain economic situations. For instance, for the simple class of models discussed in Section 4, most of the methods considered in Christiano and Fisher (1994), Judd (1992) and Taylor and Uhlig (1990) seem to provide accurate solutions at a much lower computational cost. We would like to emphasize that this somewhat controversial point does not undermine the relevance of our results. First of all, for the basic models implemented in this paper our algorithm can be effected in a reasonable computing time; in addition, technological developments in data storage and computation will facilitate the use of more reliable methods in a near future. Second, for large-scale models where our numerical scheme may become currently infeasible, it remains to be shown that the aforementioned algorithms will provide reasonable numerical solutions. In this research endeavor, our analysis of the error should be of guidance for conjecturing sensible approximation procedures for the advancement of faster algorithms in the simulation of more complex models. Third, as stressed in Rust (1994), recent developments in complexity theory suggest that if as in this paper comparisons are made on a worst-case error basis (the standard practice in numerical analysis), then it is not possible to circumvent the "curse of dimensionality". This implies that, unless specific properties of the model are embedded in the numerical scheme, one must face the decision of using reliable, costly algorithms, or else resorting to fast, *ad hoc* procedures that offer no guarantee about the accuracy of the solution.[12]

The numerical implementation of our algorithm has been achieved by standard, well-tested procedures, which enhance the applicability of our results. In terms of computational costs and other considerations, our method seems to fare well as compared to similar, rougher procedures found in the economic literature such as those devised by Christiano (1990) and Tauchen (1990). As it is customary at the initial stages of research of this nature, there should be wide room for further improvements with substantial savings in computing time. In this sense, our numerical scheme should be taken as a first attempt to vindicate the feasibility of the proposed algorithm.

The analysis of this paper has been restricted to a class of convex economies where optimal solutions may be thought as competitive allocations. These models are the simplest to study. It remains as a challenge for future economic research

[12]As a matter of fact, Chow and Tsitsiklis (1991) argue that, among the class of accurate methods, value-function iterated algorithms with continuation are roughly optimal.

the derivation of an analogous theory of error for models with money, externalities, adjustment costs, taxes, and other types of frictions where the equivalence between optimal and competitive solutions is not guaranteed. Incidentally, such a research program may entail the study of accuracy properties of faster, more operational algorithms.

References

Araujo, A. (1991), "The Once but not Twice Differentiability of the Policy Function," *Econometrica* 59, 1383-1391.

Araujo, A. and J.A. Scheinkman (1977), "Smoothness, Comparative Dynamics and the Turnpike Property," *Econometrica* 45, 601-620.

Blume, L., D. Easley and M. O'Hara (1982), "Characterization of Optimal Plans for Stochastic Dynamic Programs," *Journal of Economic Theory* 28, 221-234.

Bona, J.L. and M.S. Santos (1994), "On the Role of Computation in Economic Theory," mimeo, Centro de Investigación Económica, ITAM.

Brenner, S. and R. Scott (1994), *The Mathematical Theory of Finite Element Methods*, Springer-Verlag.

Chow, C.-S. and J.N. Tsitsiklis (1991), "An Optimal One-Way Multigrid Algorithm for Discrete-Time Stochastic Control," *IEEE Transactions on Automatic Control* 36, 898-914.

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

Christiano, L.J. and J.D.S. Fisher (1994), "Algorithms for Solving Dynamic Models with Occasionally Binding Constraints," mimeo, Department of Economics, University of Western Ontario.

Danthine, J.-P., J.B. Donaldson and R. Mehra (1989), "On some Computational Aspects of Equilibrium Business Cycle Theory," *Journal of Economic Dynamics and Control* 13, 449-470.

Dotsey, M. and C.S. Mao (1992), "How well do Linear Approximation Methods Work?" *Journal of Monetary Economics* 29, 25-58.

Falcone, M. (1987), "A Numerical Approach to the Infinite Horizon Problem of Deterministic Control Theory," *Applied Mathematics and Optimization* 15, 1-13.

Gallego, A.M. (1993), "On the Differentiability of the Value Function in Stochastic Growth Models," mimeo, Universidad de Alicante.

Gill, P.E., W. Murray and M.H. Wright (1981), *Practical Optimization*, Academic Press.

Johnson, C. (1987), *Numerical Solution of Partial Differential Equations by Finite Element Method*, Cambridge University Press.

Judd, K.L. (1991), *Numerical Methods in Economics*, mimeo, Hoover Institution.

Kahaner, D., C. Moler and S. Nash (1989), *Numerical Methods and Software*, PTR Prentice Hall.

Kehoe, T.J. (1991), "Computation and Multiplicity of Equilibrium," in W. Hildenbrand and H. Sonneschein, eds., *Handbook of Mathematical Economics*, vol. 4, North-Holland, Amsterdam.

Kehoe, T.J., D.K. Levine and P.M. Romer (1990), "Determinacy of Equilibrium in Dynamic Models with Finitely many Consumers," *Journal of Economic Theory* 50, 1-21.

IMSL Math/Library Users Manual (1984, IMSL Inc. 2500 City West Boulevard, Houston TX 77042).

Li, J.X. (1993), *Essays in Mathematical Economics and Economic Theory*, Ph.D. dissertation, Department of Economics, Cornell University.

Manuelli, R. and T. Sargent (1987), *Exercises in Dynamic Macroeconomic Theory*, Harvard University Press.

Marcet, A. (1994), "Simulation Analysis of Stochastic Dynamic Models: Applications to Theory and Estimation," in C. Sims, ed., *Advances in Econometrics Sixth World Congress*, vol. 2, Cambridge University Press.

Marcet, A. and D.A. Marshall (1994), "Solving Non-Linear Rational Expectations Models by Parameterized Expectations," mimeo, Universitat Pompeu Fabra, Barcelona.

Marsden, J.E. and A.J. Tromba (1981), *Vector Calculus*, W.H. Freeman and Company.

Montrucchio, L. (1987), "Lipschitz Continuous Policy Functions for Strongly Concave Optimization Problems," *Journal of Mathematical Economics* 16, 259-273.

Prescott, E.C. (1986), "Theory ahead of Business Cycle Measurement," *Quarterly Review*, Federal Reserve Bank of Minneapolis, 10 (4), 9-22.

Press, W.H., S.A. Teukolsky, W.T. Vetterling and B.P. Flannery (1992), *Numerical Recipes in FORTRAN. The Art of Scientific Computing*, Cambridge University Press.

Rust, J. (1994), "Numerical Dynamic Programming in Economics," mimeo, University of Wisconsin.

Santos, M.S. (1992), "Differentiability and Comparative Analysis in Discrete-Time Infinite-Horizon Optimization," *Journal of Economic Theory* 57, 222-229.

Santos, M.S. (1994), "Smooth Dynamics and Computation in Models of Economic Growth," *Journal of Economic Dynamics and Control* 18, 879-895.

Stoer, J. and R. Bulirsch (1993), *Introduction to Numerical Analysis*, Springer-Verlag.

Stokey, N.L. and R.E. Lucas (1989), *Recursive Methods in Economic Dynamics*, Harvard University Press.

Taylor, J.B., and H. Uhlig (1990), "Solving Non-Linear Stochastic Growth Models: A Comparison of Alternative Solution Methods," *Journal of Business and Economic Statistics* 8, 1-18.

6. Appendix

This appendix is made up of two parts. In Part I we indicate how the existing results for deterministic models can be extended to substantiate the second-order differentiability of the value function in our stochastic framework. In Part II we describe in some detail the algorithms used for numerical integration and maximization.

PART I

Second-Order Differentiability of the Value Function

Our purpose now is to provide proofs of Theorems 3.1 and 3.2. Our method of proof is based upon the differentiability analysis of Santos (1992, 1994). Since most of this analysis seems to carry over to the stochastic case, we shall only sketch key modifications to the previous method of proof in order to validate the present results.

The proofs of Theorems 3.1 and 3.2 follow from a series of lemmas that establish existence and continuity of the partial derivatives.

LEMMA A.1: *The second-order partial derivative $D_{11}W(k_0, z_0)$ exists and is jointly continuous on $int(K \times Z)$. Furthermore, this derivative can be characterized by the following quadratic optimization problem*

$$x_0 \cdot D_{11}W(k_0, z_0) \cdot x_0 = \sup_{\{\pi_t\}_{t\geq 0}} \sum_{t=0}^{\infty} \beta^t \int_{Z^t} \left[(\pi_t, \pi_{t+1}) \cdot D^2 v_{z_t}(k_t, k_{t+1}) \cdot (\pi_t, \pi_{t+1})\right] \mu^t(z_0, dz^t) \tag{6.1}$$

As in Section 3, here the maximization proceeds over all measurable functions $\{\pi_t\}_{t=0}^{\infty}$, $\pi_t : Z^{t-1} \to R^l$, for $t \geq 1$ with $\pi_0 = x_0$ fixed, the one-period objective $D^2 v_{z_t}(k_t, k_{t+1})$ is the Hessian matrix of the mapping $v(\cdot, \cdot, z_t)$ for z_t fixed, and $\{k_t, z_t\}_{t=0}^{\infty}$ is the optimal contingency plan to (2.1) for the initial value (k_0, z_0).

For this stochastic model, a formal proof of the existence of the derivative $D_{11}W(k_0, z_0)$ satisfying (6.1) is found in Gallego (1993). As discussed in Santos (1992) such derivative must be continuous in (k_0, z_0), since the value function corresponding to the optimization problem (6.1) varies continuously with the vector of initial conditions (k_0, z_0). These considerations suffice to establish the lemma.

From this characterization, one readily proves [cf. Santos (1994)] that the optimal plan $\{\widehat{\pi}_t\}_{t=0}^{\infty}$ to maximization problem (6.1) determines the derivative of the policy function with respect to k_0. That is, $\widehat{\pi}_t = D_1 g^t(k_0, z_0) \cdot \pi_0$ for $t \geq 1$, where $D_1 g^t(k_0, z_0)$ denotes the derivative of the function $g(g(...g(k_0, z_0), ...), z_{t-2}), z_{t-1})$ with respect to k_0 for every possible realization $(z_1, z_2, ..., z_{t-1})$. As illustrated in Section 3, these results imply that

$$\sum_{t=0}^{\infty} \beta^t \int_{Z^t} \left\| D_1 g^{t+1}(k_0, z_0) \right\|^2 \mu^t(z_0, dz^t) \leq \frac{L}{\alpha} \tag{6.2}$$

where $L = \sup_{(k_0, k_1, z_0) \in \Omega} \|D_{11} v(k_0, k_1, z_0)\|$.

LEMMA A.2: *The cross-partial derivatives $D_{12}W(k_0, z_0)$ and $D_{21}W(k_0, z_0)$ exist and are jointly continuous on int $(K \times Z)$. Furthermore, these derivatives can be computed by the expression*

$$\begin{aligned} D_{12}W(k_0, z_0)^T &= D_{21}W(k_0, z_0) \\ &= \sum_{t=0}^{\infty} \beta^t \int_{Z^t} [\left(\tfrac{\partial z_t}{\partial z_0}\right) \cdot (D_{31} v(k_t, k_{t+1}, z_t) \cdot D_1 g^t(k_0, z_0) + \\ &\quad D_{32} v(k_t, k_{t+1}, z_t) \cdot D_1 g^{t+1}(k_0, z_0))] \mu^t(z_0, dz^t) \end{aligned} \tag{6.3}$$

where $\dfrac{\partial z_t}{\partial z_0}$ denotes the derivative $\dfrac{\partial}{\partial z_0} \varphi\left(\varphi\left(\cdots\left(\varphi\left(z_0, \varepsilon_1\right), \varepsilon_2\right) \cdots\right), \varepsilon_t\right)$ for every possible realization $(\varepsilon_1, ..., \varepsilon_t)$ and $t > 0$, and the right-hand side of (6.3) is evaluated at the optimal contingency plan $\{k_t, z_t\}_{t=0}^{\infty}$.

PROOF: Let us consider the function

$$D_2 W^S(k_0, z_0) = \sum_{t=0}^{S} \beta^t \int_{Z^t} \left[D_3 v(k_t, k_{t+1}, z_t) \cdot \frac{\partial z_t}{\partial z_0} \right] \mu^t(z_0, dz^t) \tag{6.4}$$

Then by Assumptions 2 and 4 the sequence $\left\{D_2 W^S(k_0, z_0)\right\}_{S \geq 1}$ converges uniformly to $D_2 W(k_0, z_0)$ as defined in (2.5). Now, differentiating (6.4) with respect

to k_0 we obtain

$$D_{21}W^S(k_0,z_0)=\sum_{t=0}^{S}\beta^t\int_{Z^t}[\left(\tfrac{\partial z_t}{\partial z_0}\right)\cdot(D_{31}v(k_t,k_{t+1},z_t)\cdot \tag{6.5}$$

$$D_1g^t(k_0,z_0)+D_{32}v(k_t,k_{t+1},z_t)\cdot D_1g^{t+1}(k_0,z_0))]\mu^t(z_0,dz^t)$$

Observe that $D_{21}W^S(k_0,z_0)$ varies continuously with (k_0,z_0). Moreover, under the maintained assumptions and by virtue of (6.2) the sequence $\left\{D_{21}W^S(k_0,z_0)\right\}_{S\geq 1}$ converges uniformly to $D_{21}W(k_0,z_0)$ as defined in (6.3) [cf. (3.5)]. Hence, $D_{21}W(k_0,z_0)$ is well defined, jointly continuous, and can be computed by expression (6.3).

In order to prove the existence of $D_{12}W(k_0,z_0)$ at every interior point (k_0,z_0), the standard argumentation to establish the symmetry of the cross-partial derivatives [e.g., Marsden and Tromba (1981)] suffices to show that $D_{12}W(k_0,z_0)^T=D_{21}W(k_0,z_0)$ at every interior point (k_0,z_0). This completes the proof of the lemma.

PROOF OF THEOREM 3.2: Let $F(k_0,k_1,z_0)=D_2v(k_0,k_1,z_0)+\beta\int_Z D_1W(k_1,z_1)Q(z_0,dz_1)$. Observe that, under the maintained assumptions and the previous results, functions $D_2v(k_0,k_1,z_0)$ and $D_1W(k_1,z_1)$ are of class $\mathcal{C}^1$ over the interior of their domains. Moreover, from a simple application of the bounded convergence theorem [e.g., Stokey and Lucas (1989, Ch. 7)] one can show that $\int_Z D_1W(k_1,z_1)Q(z_0,dz_1)$ is $\mathcal{C}^1$ as a function of k_1. Therefore, function F must also be of class $\mathcal{C}^1$. Then, by the asserted concavity of v and the concavity of W on k_1, the implicit function theorem implies that the optimal policy $k_1=g(k_0,z_0)$ is a $\mathcal{C}^1$ mapping.

To complete the proof of Theorem 3.2, we need to show that $\sum_{t=1}^{\infty}\beta^t\int_{Z^t}\left\|D_1g^t(k_0,z_0)\right\|^2\mu^t(z_0,dz^t)\leq\frac{L}{\alpha}$, for some constant $L>0$. This is, however, expression (3.4). The proof is complete.

LEMMA A.3: *The second-order partial derivative $D_{22}W(k_0,z_0)$ exists and is jointly continuous on int$(K\times Z)$. Furthermore, this derivative can be computed*

by the expression

$$D_{22}W(k_0,z_0)=\sum_{t=0}^{\infty}\beta^t\int_{Z^t}[D_3v(k_t,k_{t+1},z_t)\cdot\tfrac{\partial^2 z_t}{\partial z_0^2}+\left(\tfrac{\partial z_t}{\partial z_0}\right)\cdot(D_{31}v(k_t,k_{t+1},z_t)\cdot$$

$$D_2g^t(k_0,z_0)+D_{32}v(k_t,k_{t+1},z_t)\cdot D_2g^{t+1}(k_0,z_0)+D_{33}v(k_t,k_{t+1},z_t)\cdot\left(\tfrac{\partial z_t}{\partial z_0}\right)]\mu^t(z_0,dz^t) \tag{6.6}$$

where $\dfrac{\partial z_t}{\partial z_0}$ *and* $\dfrac{\partial^2 z_t}{\partial z_0^2}$ *are the previously defined derivatives, and the right-hand side of (6.6) is evaluated at the optimal contingency plan* $\{k_t,z_t\}_{t=0}^{\infty}$.

PROOF: The proof follows the same lines as that of the preceding lemma. After differentiating (6.4) with respect to z_0 and taking the limit we obtain expression (6.6). We then have to prove that this latter expression is uniformly bounded. The boundedness of the first term in (6.6) follows from Assumptions 2 and 4. Moreover, as discussed in Santos (1992), an upper estimate for the second term may be obtained from the fact that

$$D_2g^t(k_0,z_0)=\sum_{s=1}^{t}D_1g^{s-1}(k_{t-s+1},z_{t-s+1})\cdot D_2g(k_{t-s},z_{t-s})\cdot\frac{\partial z_{t-s}}{\partial z_0}$$

for $t=1,2,...,$ and every optimal realization $\{k_t,z_t\}_{t=0}^{\infty}$. The proof is complete.

Finally, we note that Theorem 3.1 is a simple consequence of Lemmas (A.1)-(A.3).

PART II

Numerical Integration and Maximization

In this last part, we discuss the algorithms employed for the integration and maximization of our functions. For stochastic objectives, we have first integrated out over the random variable, and then proceeded to the numerical maximization. All optimization problems considered so far involve a univariate stochastic process, and in consequence the same numerical scheme has been followed in all our integrations. Regarding maximizations, algorithms for functions in several variables are usually inefficient for one-dimensional problems. Accordingly, a distinction has been made between univariate maximization (Examples 1 and 2) and

multivariate maximization (Example 3). As already emphasized, our integration and maximization procedures follow standard practice in numerical analysis. (All subroutines may be found in commonly used libraries such as IMSL and NAG.) For a theoretical discussion of these methods, the reader is referred to Gill, Murray and Wright (1981), Kahaner, Moler and Nash (1989), and Press *et al.* (1992).

Integration

To avoid issues related to the evaluation of indefinite integrals, the i.i.d. process $\{\varepsilon_t\}$ has been restricted to a compact domain $[a, b]$. It should be pointed out that in our examples mean-preserving changes of these interval values will alter the domain of definition of the value and policy functions but not their functional forms. In other contexts, this boundedness assumption may be viewed as an approximation of an underlying random process defined over an unbounded domain. In most circumstances, it is then plausible to obtain a good estimate of the error stemming from such an approximation.

In all our computations, we have posited the same density function for the stochastic process $\{\varepsilon_t\}$. The density function has been defined as follows

$$g(x) = \begin{cases} \dfrac{\int_a^x f(s)ds}{\int_a^b f(s)ds} & \text{for } x \in [a, b] \\ \\ 0 & \text{otherwise} \end{cases}$$

where $f(x)$ is the density function of a normal distribution with zero mean and standard deviation equal to 0.008. Such a small standard deviation concentrates most of the mass around the mean, and complicates the process of numerical integration. The boundary values of the interval $[a, b]$ have been set to $a = -0.032$ and $b = 0.032$ so as to encompass in each side four standard deviations from the mean. Hence, for parameter value $\rho = 0.9$, the domain of z is restricted such that *logz* lies in the interval of values $[-0.32, 0.32]$.

All our integrations have been performed under the subroutines *qsimp* and *qtrap* as specified in Press *et al.* (1992, Sec. 4.2). As argued in this treatise, these subroutines are fairly efficient for the integration of relatively simple problems. For each given z, our task is to compute the conditional expectation over the stochastic innovation. This calculation is implemented by the corresponding routines via an *N-stage refinement* of an *extended trapezoidal rule* [*op. cit.*, equation

(4.1.11)]. This rule allows for successive subdivisions of existing panels in order to benefit from previous functional evaluations. Thus, if S_N is the integral evaluation obtained from N equally spaced points, then these values can also be used to compute S_{2N}. Moreover, as illustrated in Press *et al.* (1992, p. 133) *Simpson's rule* can be defined as $S^N = \frac{4}{3}S_{2N} - \frac{1}{3}S_N$, and for sufficiently smooth functions this substraction will cancel out the leading error term.

These routines have embedded an *automatic quadrature algorithm*, which computes a numerical estimate of the approximation error for two successive evaluations S^N and S^{2N}. In all our numerical exercises the algorithm is supposed to stop once such an approximation error is less than or equal to 10^{-8}. This is the value assigned to $TOLI$.

As discussed in Kahaner, Moler and Nash (1989, p. 155), it is usually more costly to estimate the integration error than to perform integral evaluations. Hence, this type of algorithm is sometimes implemented at the initial stages of a project. In our exercises, the required level of accuracy guarantees that the integration error is negligible as compared to the approximation error stemming from the piecewise linear interpolation of our functions over all considered grids of mesh size $h \geq 10^{-3}$.

Maximization

Techniques for solving maximization problems are generally similar to those of finding zeroes of non-linear equations. In our case, all our functional interpolations generate kinks at the nodal values, and hence our analysis has been restricted to algorithms suitable for non-smooth optimization. Also, by a straightforward substitution of the constraints into the objectives, our problems fall into the class of unconstrained maximization. As our objectives are strictly concave over the decision variables, these functions contain a unique maximum.

(1) *Univariate maximization.* For the process of univariate maximization, we have initially *bracketed the maximum.* That is, in some very simple way we select a triplet of points $a < b < c$ such that $f(b)$ is greater than both $f(a)$ and $f(c)$. (This guarantees the existence of a maximum inside the chosen interval. Moreover, by the strict concavity of the objective such solution is the desired global maximum.)

Once the maximum has been bracketed, we have followed a searching process that exploits regularity properties of our functions. As already pointed out, the

derivatives are not well defined at vertex points; accordingly, it seems safest to consider routines that are safeguarded by functional evaluations (instead of those just based on first-order derivatives). We should remark, however, that these kinks cannot be arbitrarily large since following the line of argumentation in Montrucchio (1987) one can show that exact maximization for all initial conditions of the domain should yield at each stage value functions with bounded curvature. Consequently, as a first approach to the problem it may be reasonable to assume that the objectives are smooth.

This is the strategy followed by Brent's method [cf., Press *et al.* (1992, Sec. 10.2)]. The method proceeds along the following steps.

(a) *Smooth approximation.* The routine selects three given function values, and constructs a parabolic approximation. Then it quickly determines the maximum of the parabola. If this maximum point falls within certain limits (i.e., the maximum is *cooperative*), then this value is added in the next iteration for a subsequent parabolic approximation until a desired level of accuracy is achieved. Convergence to the true maximum is of order 1.324.

(b) *Golden-section search.* If the parabolic approximation is not a reasonable one, then the routine switches to a more reliable but slower method called golden-section search. This procedure is analogous to the familiar method of *bisection* for finding the zeroes of a univariate function. Given at each stage a bracketing triplet of points, golden-section search tries a point that is a fraction 0.38197 into the largest of the two intervals from the central point of the triplet. With the four points now available, the procedure then selects a new bracketing triplet. Following this iterative process, the interval of search is eventually reduced at each stage by $1 - 0.38197 = 0.61803$, which corresponds to the rate of convergence of this method.

We have implemented this routine as specified in Press *et al.* (1992, Sec. 10.2), where this maximization procedure is discussed. In all our examples $TOLM$ has been set to 10^{-8}, which means that the computed maximum should be within such distance from the true one. This level of accuracy then implies that the computed value for the objective must be (up to a constant) within a 10^{-16} distance from the true one.

The above method falls into the class of the so-called *safeguarded* procedures, which combine fast algorithms with slower, more reliable ones. Press *et al.* (1992, Sec. 10.3) discuss another method of this nature which seems appropriate for univariate concave optimization. The method proceeds as follows. Given a brack-

eting triplet of points $a < b < c$, one determines the direction of the derivative at the intermediate point, b. This information then defines the next interval of search, which would be either $[a, b]$ or $[b, c]$. The value of the derivatives of the two chosen points can then be used to produce another intermediate point by some root finding procedure such as the *secant* method. If this method yields values beyond certain limits, then one bisects the interval under consideration.

Of course, in order to implement this latter safeguarded procedure, concavity and smoothness properties of the univariate problem are essential. In the unidimensional case, concavity is always preserved by piecewise linear interpolations. Regarding differentiability, we could compute for instance one-side derivatives, or else resort to higher-order interpolations preserving first-order differentiability. Since Brent's method seems efficient for all cases studied, we have refrained from consideration of this latter alternative.

(2) *Multivariate maximization.* There are also here a host of algorithms, the usefulness of which depends on the dimensionality, concavity and smoothness properties of the optimization problem. Given our interpolation procedure, the derivatives are not well defined at the edges of the subdivisions. For reasons pointed out above, these kinks cannot be arbitrarily large. Hence, we should expect that our functions are approximately smooth. Also, since there is only one endogenous state variable, all our objectives are concave over the decision variables.

Two popular methods in multivariate (non-smooth) optimization are *conjugate-gradient* and *quasi-Newton* methods. Their convergence properties are somewhat involved.

Conjugate-gradient methods construct a sequence of searching directions which satisfy certain orthogonality and conjugacy conditions so as to improve at each stage the search for a maximum. Conjugate-gradient methods do not require knowledge of the Hessian matrix, and hence their applicability extends to large-scale problems.

Quasi-Newton methods exploit information on the slope and curvature of the objective function. This is the methodology that we have followed, and it seems appropriate for the problem at hand. Of course, a general method for non-smooth functions may be highly inefficient when applied to smooth optimization problems [Gill, Murray and Wright (1981, p. 93)].

As suggested in this latter monograph, we have computed the gradient using *forward-differences*, and when this formulation has not been sufficiently accurate

we have employed *central differences.* Double precision has been implemented in these calculations, since inaccuracies in the computation of the gradient may stop the algorithm at an undesired solution.

Regarding second-order differentiation, quasi-Newton methods try to estimate the curvature of the function without explicitly computing the Hessian matrix. Thus, one starts each iteration with a matrix B_k which reflects known second-order information, and which is supposed to be an approximation of the true Hessian if the function is sufficiently smooth. (At the initial stage one usually starts with B_0 equal to the identity matrix, and in such case the algorithm reduces to the *steepest-descent* method.)

Given a matrix B_k, the search direction, p_k, is the solution to

$$B_k \cdot p_k = -g_k$$

where g_k is the gradient vector. Then the new Hessian estimate B_{k+1} is updated by the *Davidon-Fletcher-Powell* formula. For this implementation, we have used the ZXMWD routine (IMSL Technical Manual). Again, in all our computations the accuracy parameter value, $TOLM$, has been set to 10^{-8}.

No. of vertex points	Mesh size	No. of iterations	CPU time	Max. observed error in g	Max. observed error in W
100	10^{-1}	99	3	4.3×10^{-2}	1.8×10^{-1}
1000	10^{-2}	189	65	3.7×10^{-2}	1.7×10^{-2}
10000	10^{-3}	278	893	3.6×10^{-4}	1.8×10^{-5}

Table 1(a): Example 1. $A = 5, \beta = 0.95, \alpha = 0.34$; without continuation.

No. of vertex points	Mesh size	No. of iterations	CPU time	Max. observed error in g	Max. observed error in W
100	10^{-1}	500	16	4.1×10^{-2}	9.6×10^{-1}
1000	10^{-2}	958	324	3.6×10^{-2}	9.6×10^{-2}
10000	10^{-3}	1416	4471	3.6×10^{-4}	9.6×10^{-5}

Table 1(b): Example 1. $A = 5, \beta = 0.99, \alpha = 0.34$; without continuation.

No. of vertex points	Mesh size	No. of iterations	CPU time
1000	10^{-2}	460	171
10000	10^{-3}	919	3251

Table 1(c): Example 1. $A = 5, \beta = 0.99, \alpha = 0.34$; with the continuation method.

No. of vertex points	Mesh size	No. of iterations	CPU time	Max. observed error in g	Max. observed error in W
3×43	0.3872	45	3	9.91×10^{-2}	2.88
9×143	10^{-1}	96	77	3.03×10^{-2}	2.1×10^{-1}
33×500	0.0282	148	1961	8.9×10^{-3}	1.48×10^{-2}

Table 2(a): Example 2. $A = 5, \beta = 0.95, \alpha = 0.34, \rho = 0.90$; without continuation.

No. of vertex points	Mesh size	No. of iterations	CPU time	Max. observed error in g	Max. observed error in W
3×43	0.3872	16	223	10^{-1}	15.42
9×143	10^{-1}	485	487	3.05×10^{-2}	1.11
33×500	0.0282	749	9964	8.7×10^{-3}	7.6×10^{-2}

Table 2(b): Example 2. $A = 5, \beta = 0.99, \alpha = 0.34, \rho = 0.90$; without continuation.

No. of vertex points	Mesh size	No. of iterations	CPU time
9×143	10^{-1}	52	46
33×500	0.0282	54	710

Table 2(c): Example 2. $A = 5, \beta = 0.95, \alpha = 0.34, \rho = 0.90$; with the continuation method.

No. of vertex points	Mesh size	No. of iterations	CPU time	Max. observed error in g	Max. observed error in W
3×43	0.3872	11	21	1.2×10^{-1}	2.42
9×143	10^{-1}	57	1041	3.6×10^{-2}	0.209
33×500	0.0282	109	25206	1.06×10^{-3}	1.46×10^{-2}

Table 3: Example 3. $A = 10, \beta = 0.95, \alpha = 0.34, \rho = 0.90$; without continuation.

	$\sigma(y)$	$\frac{\sigma(k)}{\sigma(y)}$	$\frac{\sigma(c)}{\sigma(y)}$	$\frac{\sigma(u)}{\sigma(y)}$	$corr(k,y)$	$corr(c,y)$
Exact solution	1.3535	0.9974	1.1086	0	0.8389	0.9723
Quadratic approximation	1.3537	0.9970	1.1084	0	0.8394	0.9724

Table 4: Standard deviations, $\sigma(i)$, and correlation coefficients, $corr(i,j)$, for $i, j = k, c, u, y$.

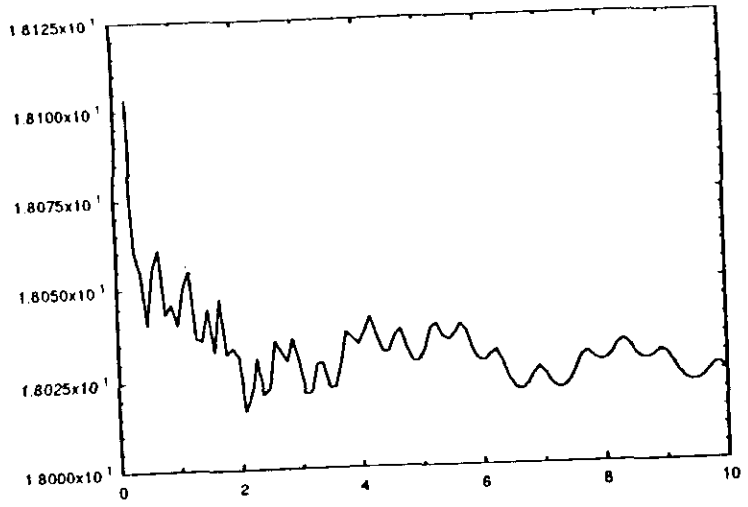

FIGURE 1(a).- Observed error $e^h_{\hat{n}}(k) = \left|W(k) - W^h_{\hat{n}}(k)\right|$, $h = 10^{-1}$, $\hat{n} = 99$, $\beta = 0.95$, for the deterministic growth model of Example 1.

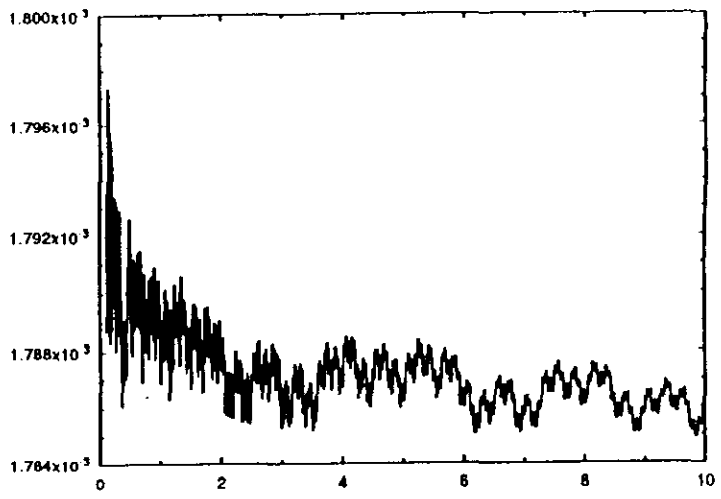

FIGURE 1(b).- Observed error $e^h_{\hat{n}}(k) = \left|W(k) - W^h_{\hat{n}}(k)\right|$, $h = 10^{-2}, \hat{n} = 189$, $\beta = 0.95$, for the deterministic growth model of Example 1.

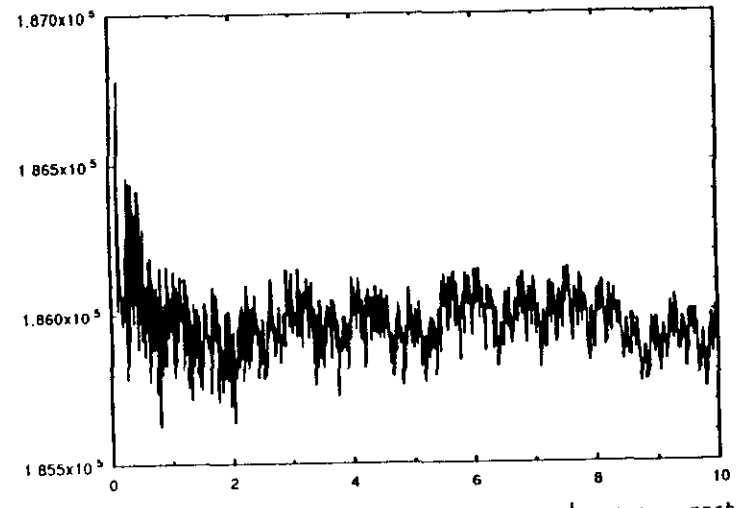

FIGURE 1(c).- Observed error $e^h_{\hat{n}}(k) = \left|W(k) - W^h_{\hat{n}}(k)\right|$; $h = 10^{-3}, \hat{n} = 278$, $\beta = 0.95$, for the deterministic growth model of Example 1.

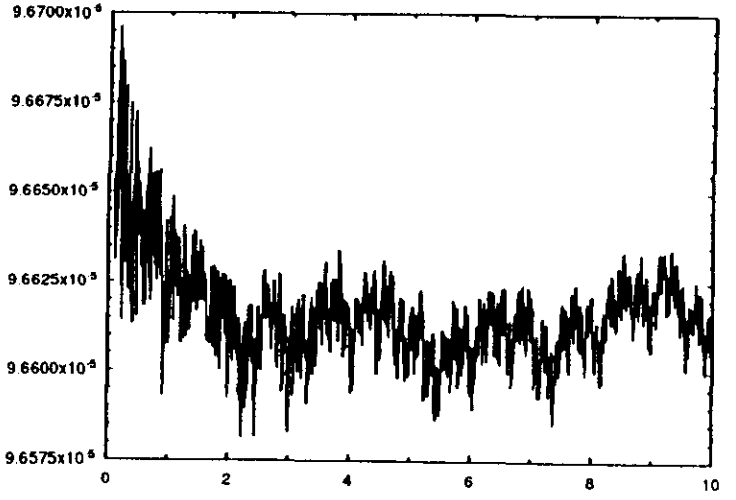

FIGURE 1(d).- Observed error $e^h_{\hat{n}}(k) = \left|W(k) - W^h_{\hat{n}}(k)\right|$, $h = 10^{-6}, \hat{n} = 1416$, $\beta = 0.99$, for the deterministic growth model of Example 1.

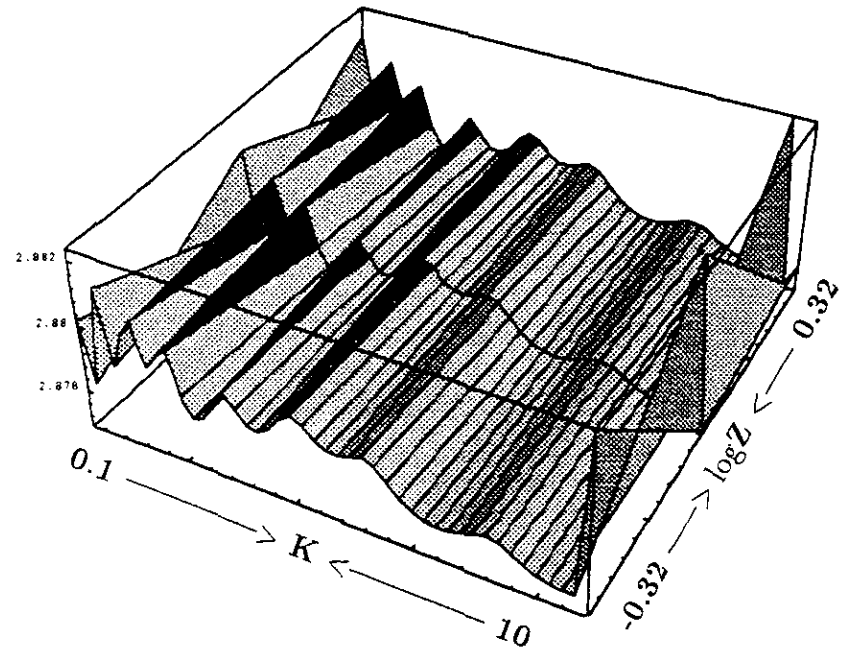

FIGURE 2(a).- Observed error $e_{\hat{n}}^h(k) = \left|W(k) - W_{\hat{n}}^h(k)\right|$, $h = 0.3872$, $\hat{n} = 45$, for the deterministic growth model of Example 2.

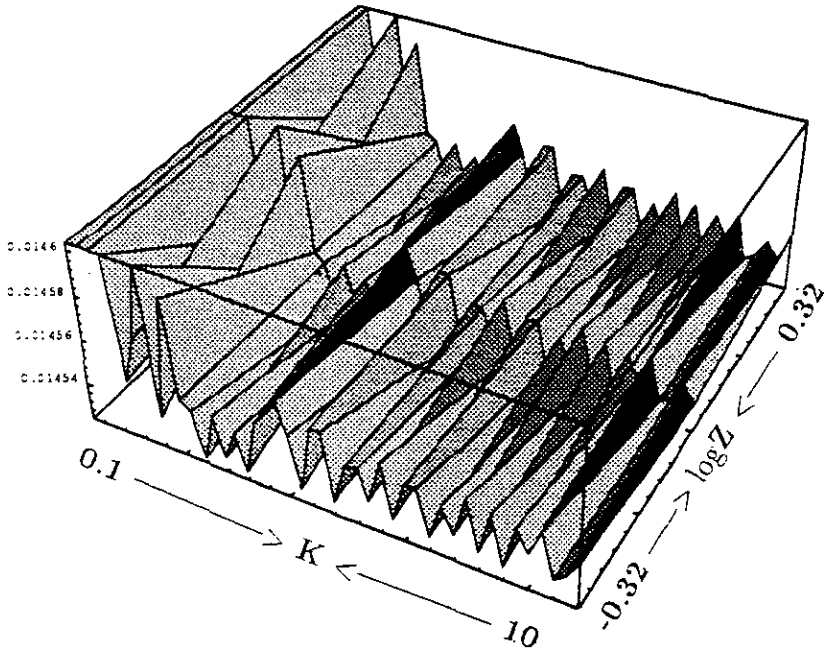

FIGURE 2(b).- Observed error $e_{\hat{n}}^h(k) = \left|W(k) - W_{\hat{n}}^h(k)\right|$, $h = 10^{-1}$, $\hat{n} = 96$, for the deterministic growth model of Example 2.

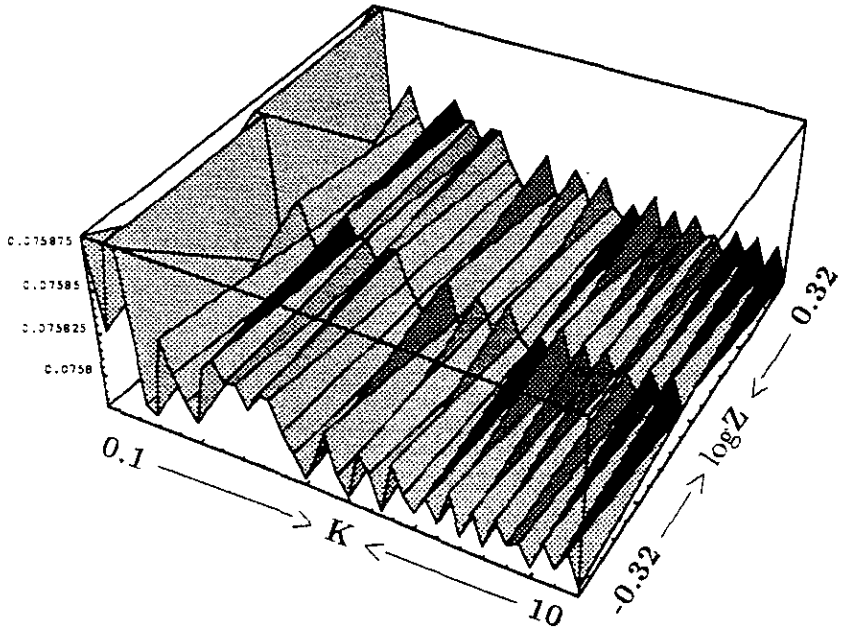

FIGURE 2(c).- Observed error $e_{\hat{n}}^h(k) = \left|W(k) - W_{\hat{n}}^h(k)\right|$, $h = 0.0282$, $\hat{n} = 148$, for the deterministic growth model of Example 2.

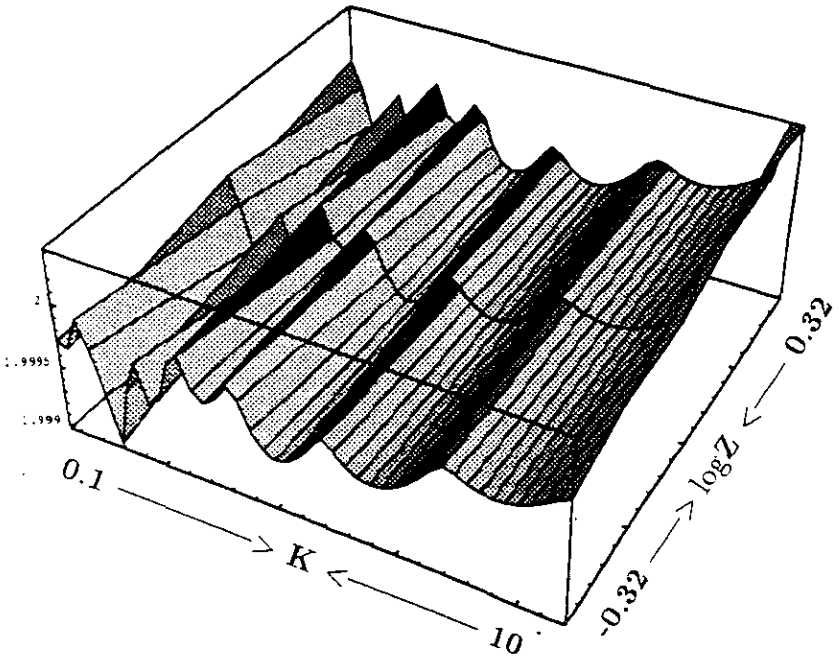

FIGURE 3(a).- Observed error $e_{\hat{n}}^h(k) = \left|W(k) - W_{\hat{n}}^h(k)\right|$, $h = 0.3872$, $\hat{n} = 11$, for the deterministic growth model of Example 3.

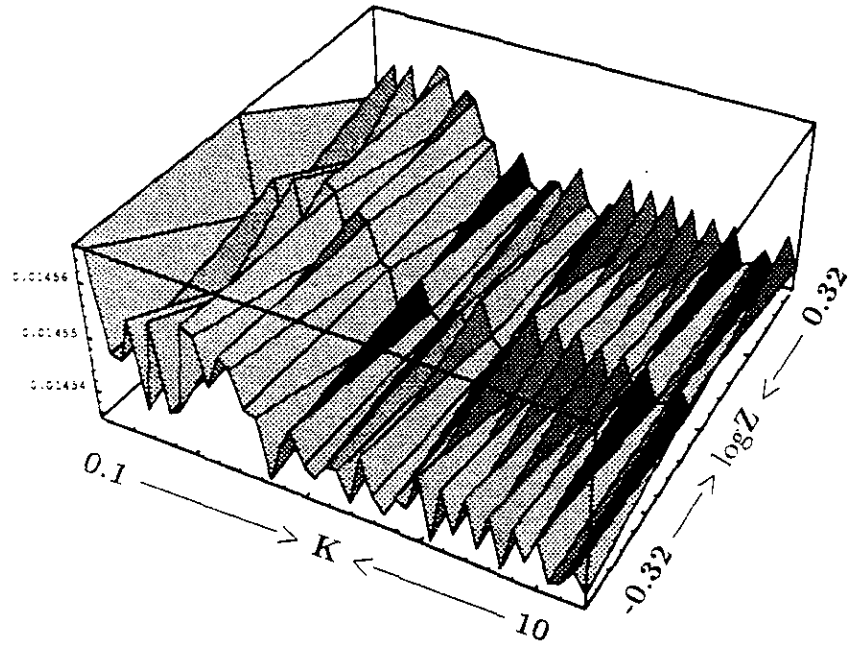

FIGURE 3(b).- Observed error $e_{\hat{n}}^h(k) = \left|W(k) - W_{\hat{n}}^h(k)\right|$, $h = 10^{-1}$, $\hat{n} = 57$, for the deterministic growth model of Example 3.

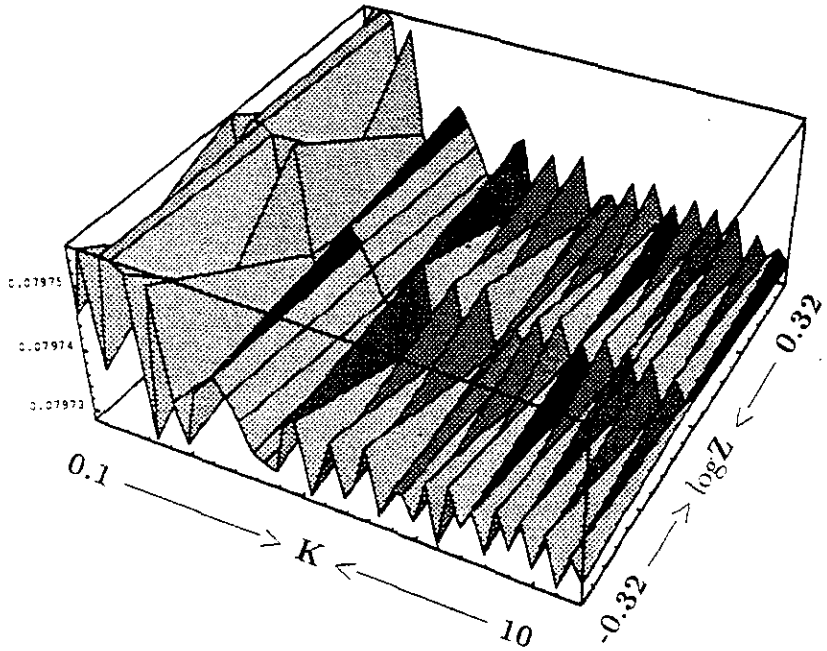

FIGURE 3(c).- Observed error $e_{\hat{n}}^h(k) = \left|W(k) - W_{\hat{n}}^h(k)\right|$, $h = 0.0282$, $\hat{n} = 109$, for the deterministic growth model of Example 3.