Title
Solving Infinite Horizon Growth Models with an Environmental Sector

Authors
Kelly, David L.
Kolstad, Charles D.

Publication Date
1999-07-25

Solving Infinite Horizon Growth Models with an Environmental Sector[*]

David L. Kelly[**]

Charles D. Kolstad[***]

Current Version: July 25, 1999

ABSTRACT

This paper concerns computational models in environmental economics and policy, particularly so-called integrated assessment models. For the most part, such models are simply extensions of standard neoclassical growth models, extended by including the environment and pollution generation. We review the structure of integrated assessment models, distinguishing between finite horizon and infinite horizon models, both deterministic and stochastic. We present a new solution algorithm for infinite horizon integrated assessment models, relying on a neural net approximation of the value function within an iterative version of the Bellman equation.

---

[*] Research supported by US Department of Energy grant number DE-FG03-96ER62277 to the University of California, Santa Barbara.

[**]Department of Economics, University of Miami, Box 248126, Coral Gables, FL 33124.

[***]Department of Economics, Bren School of Environmental Science and Management, and Environmental Studies Program, University of California, Santa Barbara, CA 93106-9210.

I. Introduction

Authors in many areas of economics now use numerical methods to solve analytically intractable problems. This approach to economics, known broadly as computational economics, offers attractive advantages such as the ability to solve complex problems, to provide policy results for realistically parameterized models, and to simulate economic models. Computational economics has long been a part of environmental economics as well, especially in relation to energy and climate change.

An important area of environmental policy is integrated assessment, centered around the integrated assessment model. An integrated assessment model combines the scientific and economic aspects of an environmental problem in order to assess policy options (as opposed to advancing knowledge for its own sake). Use of integrated assessment models today is in the context of climate change policy, but the models also have found applications in areas such as energy modeling and ozone depletion. See Kelly and Kolstad (1999b) or Weyant, et. al. (1996) for surveys of integrated assessment models in climate change. This paper focuses on the computational issues surrounding integrated assessment models. The primary application is climate change; however, most of the work applies equally well to other types of integrated assessment involving significant dynamic dimensions

The optimal growth model is the basis for many integrated assessment models of stock pollutants. Typically, an environmental externality of some kind is embedded in the optimal growth model of Ramsey (1928), Koopmans (1965), or Cass (1965). In some cases, this framework is further embedded in a general equilibrium trade model (eg, Manne, et. al. 1995). Hence one might expect that the burgeoning literature on numerical solutions to optimal growth models would apply to integrated assessment models. Although a few standard numerical solutions apply in general, integrated assessment models possess several unique features which make many solution methods difficult or impossible to apply.

The purposes of this paper are (a) to discuss methods which solve computational integrated assessment models and (b) to present a particular algorithm for solving a class of infinite horizon economic growth models, possibly including an environmental secton (integrated assessment models). In the next section of the paper, we provide background information on solution methods for integrated assessment models, first distinguishing between single region and multi-region methods and then finite horizon (FH) and infinite horizon methods (IH). In the subsequent section we present a stylized integrated assessment model and define FH and IH solution approaches. We then present a new solution method for solving IH problems (growth models with or without an environmental sector). The solution method does not rely on an approximation around the steady state, and does not directly approximate the policy function. Hence the solution method overcomes many of the difficulties associated with using infinite horizon methods for integrated assessment. These features make our method ideal for integrated assessment models. However, the procedure does have a few drawbacks, including relatively slow computation time and perhaps relative difficulty of use and programming.

II. Background

Two of the most well-known integrated assessment models are the DICE model of William Nordhaus (1994) and the Global 2100 model of Alan Manne and Richard Richels (1992). Both of these models involved modified Ramsey growth models and have been the basis for a number of other integrated assessment models (see Weyant et al, 1996). The DICE model is a single region model with a climate sector, endogenous emission control and climate damage. The Global 2100 model is a regional model but with no trade among regions and no endogenous emission control. Thus each region is treated as an independent Ramsey growth model. Over the past several years, both of these research groups have developed regional variants of their models with explicit though simple trade (Manne et al, 1995). In both cases, capital is allowed to freely

flow between regions which means that the marginal productivity of capital must be the same in every region, an additional constraint that considerably complicates the solution.[1]

Most integrated assessment models are deterministic, a reflection of the computational complexity and long run nature of climate change. Still, many aspects of climate change, such as the assumed temperature change from a doubling of greenhouse gasses or the damage from a three degree Celsius temperature change, are notoriously uncertain. Further, many of the state variables in climate change, such as the temperature or rainfall, are quite stochastic. Many authors address uncertainty by assuming certainty equivalence. Under certainty equivalence, the optimal policy under uncertainty equals the mean of all policies derived from assuming each possible realization of the uncertain parameter is known with certainty. However, certainty equivalence holds in few if any climate change models. Thus, certainty equivalence is at best an approximation which ignores the risk aversion inherent in the policy debate over climate change. Certainty equivalence is also not easy to apply to models with stochastic state variables. Several models are solved under uncertainty without assuming certainty equivalence. For example, Peck and Teisberg (1989) and Kolstad (1996) allow for simple randomness over the most uncertain parameters, which is resolved at a fixed time in the future. Kelly and Kolstad (1999a) allows for stochastic temperature and randomness over parameters, which is resolved period by period through Bayesian Learning. As with trade, the addition of uncertain parameters without certainty equivalence or stochastic state variables considerably complicates the solution.

In the context of the infinite horizon Ramsey optimal growth model, perhaps the most popular solution technique is the linear-quadratic (LQ) method. First proposed by Kydland and Prescott (1982), papers which use or test the LQ method include Christiano (1987b, 1988), Hansen (1985), Christiano and Eichenbaum (1988), Cooley and Hansen (1989), and Hansen and Sargent (1988). The LQ method makes a quadratic approximation of the utility function around

[1] The basic solution technique of these models is to maximize a weighted sum of utilities (one "utility" function for each region) with the weights endogenously adjusted so that either the marginal productivity of capital is the same for every region (Manne et al, 1995) or so that there

the steady state and a linear approximation of the transition equations. The resulting Bellman equation has a quadratic value function as a solution and linear first order conditions. Hence the problem can be solved by repeatedly solving a linear set of equations until the solution converges. The LQ method offers several attractive features: the LQ method is one of the fastest solution methods, the LQ method generalizes well to complex problems with no curse-of-dimensionality, and the LQ method is intuitive, easy to program, and use. Although the LQ method is quite robust, it does not generalize well to climate change. For many environmental problems, for climate change, the starting condition is not close to the steady state, hence methods which use local approximations around the steady state (such as LQ approximations) are particularly inaccurate, especially since the main interest is in the numerical value of the current policy.

Other methods are also difficulty to apply. Policy variables are not necessarily monotonic in the state variables,[2] hence fast value approximation methods are not very useful. Most integrated assessment models have a large number of states and controls and thus suffer from a curse of dimensionality. In such cases, most discrete pure grid methods do not work well (a minimal problem, such as Kelly and Kolstad (1999a), 1998, has 7 states, 2 controls, and 2 random variables, hence even a modest grid of 100 points per variable results in 100^{11} grid points for a pure grid method). Solution methods based on flexible functional forms typically use a smaller grid. But the Euler equations are typically difficult to specify, because there are multiple control variables and many lagged control terms. Hence, approximations of the conditional expectation of the first order condition (as in parameterized expectations of Den-Haan and Marcet, 1990) or the policy function (as in the minimum weighted residuals method of Judd, 1991) are possible but not appealing.

---

are no net inflows or outflows of capital into any region over the horizon of the model (Nordhaus and Yang, 1996).

[2] In fact there is often a 'U' shaped relationship between growth and environmental policy. See Grossman and Krueger (1995) or Kelly (1998a).

Given these difficulties, it is perhaps not surprising that almost all integrated assessment models use the simplest possible solution method: truncation of the model to a finite horizon so that standard non-linear programming methods may be applied. We denote this techique the finite horizon (FH) method. Nordhaus (1994), Peck and Teisberg (1993) and Manne and Richels (1992) are just a few examples. In the FH method, the utility from consumption and the environment is fixed after a finite terminal year. Furthermore, the social planner or agent must have a specified amount of assets at the terminal date. Given these assumptions, the researcher need only solve a simple constrained optimization problem with standard software.

Although FH is easy to use, robust to almost all integrated assessment models, and has relatively low computation time, FH does have limitations. The FH method does not work well on stochastic models for which certainty equivalence does not hold. The FH method is not recursive and thus requires specification of all possible realizations of the random variables over time. Hence for a stochastic problem spanning hundreds of years such as climate change, only the simplest possible random variables are possible (for example Peck and Teisburg, 1989 allow only a single two-state random variable realized in a single future period). Second, inference is difficult because the solution produces only an optimal solution path, and says nothing about the relationship between the solution and the state variables and parameters (such as the LQ or flexible functional form methods which give policy functions). Because FH does not compute the policy function, the model cannot be simulated for alternative starting conditions or realizations of the random variables without resolving the entire model. Finally, the solution method can be sensitive to the specification of the terminal conditions, which makes sensitivity analysis difficult.

III. The Structure of Integrated Assessment Models

Consider a discrete time model (many results here also extend to the continuous time case), where t indexes time. In keeping with the concept of integrated assessment, we divide the model

into three sectors: a climate (or scientific) sector, an economic sector, and an emissions sector. The climate sector consists of equations that determine the movement of climate variables such as temperature, ocean temperature, and rainfall over time. In other models, the climate sector is the equations that determine changes in the atmosphere, or other environmental system affected by pollution. The economic sector consists of equations which determine the evolution of the economy, damage from climate change, as well as the costs of climate change control and resource constraints. The emissions sector consists of equations which govern the generation of greenhouse gasses and related pollutants. We can write the three sectors as:

$$\begin{aligned} S_{R,t+1} &= f_R\left(S_{R,t}, S_{M,t}, \varepsilon_{R,t+1}\right) \\ S_{E,t+1} &= f_E\left(S_{E,t}, S_{R,t}, C_t, \varepsilon_{E,t+1}\right) \\ S_{M,t+1} &= f_M\left(S_{E,t}, C_t, \varepsilon_{M,t+1}\right) \end{aligned} \tag{1}$$

Here S_i is a vector of states for the climate, economy, and emissions, respectively and f_i are vectors of transition equations, C_t is a vector of control variables, and ε_i are random variables. Note the linkage between the sectors. Economic activity affects emissions, thus linking the economic and emissions sectors, emissions affects climate, linking the climate and emissions sectors, and climate change affects the economy through damage, thus linking the climate and economic sectors. However, not all sectors are linked in the same way. For example, control variables can reduce emissions, but do not change the climate directly. Let $S = \left[S_{R,t} \quad S_{E,t} \quad S_{M,t}\right]$ and similarly for f; then combining the transition equations together gives:

$$S_{t+1} = f\left(S_t, C_t, \varepsilon_{t+1}\right) \tag{2}$$

Our model includes random shocks to the economic and climate variables, since many state variables, such as temperature and productivity, have a stochastic component. Epsilon can also represent uncertain scientific parameters such as the damage from a three degree temperature

change which are random from the point of view of the decision maker (our specification also allows for learning to reduce the variance of the uncertainty over time).

Next we describe the behavior of the control variables. Traditionally, there are two possible models. The first uses existing forecasts or proposed policies to specify the control variables. Following Weyant, et. al. (1996), we call such models *policy evaluation models*. The second chooses an objective function, and determines the optimal control using optimization. Following Weyant, et. al. (1996), we call such models *policy optimization models* (see Kelly and Kolstad, 1999b, for a further discussion). Consider first a policy optimization problem. Suppose there exists a time separable objective function U and suppose that there is a set of constraints on the choice of controls, Γ. It is natural to think of U as a representative consumer's utility. Then we may write the social planning problem as:

$$W = \max_{C_t \in \Gamma(S_t)} \left\{ \mathrm{E}\left[\sum_{t=0}^{\infty} \beta^t \, \mathrm{U}(C_t, S_t) \right] \right\} \qquad \text{Subject to:} \tag{3}$$

$$S_{t+1} = \mathrm{f}(C_t, S_t, \varepsilon_{t+1})$$

$$\varepsilon_{t+1} \sim \Omega(S_t)$$

$$\lim_{t \to \infty} \beta^t \, \mathrm{U}_s(C_t, S_t) \cdot S_t = 0$$

That is, maximize expected present discounted utility subject to the transition equations and a transversality condition. Note that policy evaluation models and deterministic are a sub case of the above framework for a well specified choice for Γ and Ω, respectively

There is a well-known recursive representation of the above problems which is especially useful for computational analysis. Let $v(S)$ denote the value function, or the value of W in Eqn. (3). Then problem (3) satisfies a dynamic consistency condition. We can define a sequence of functions, $\{v_i(S)\}$ by using Bellman's principle of optimality:

$$v_i(S) = \max_{C \in \Gamma(S)} \{ \mathrm{U}(S,C) + \beta \mathrm{E}[v_{i-1}(\mathrm{f}(S,C,\varepsilon))] \} \quad (4)$$

Under minimal assumptions, the sequence of functions defined by equation (4) converges to a fixed point), $v(S)$, which also satisfies (3):

$$v(S) = \max_{C \in \Gamma(S)} \{ \mathrm{U}(S,C) + \beta \mathrm{E}[v(\mathrm{f}(S,C,\varepsilon))] \} \quad (5)$$

Of course there is no transversality condition as the conditions for existence of a value function effectively rule out the same exploding solutions which the transversality condition rules out. Because equation (4) is recursive, it provides the basis for a computational solution to to the value function given in (5). We refer to equation (4) or equation (5) as the Bellman equation.

IV. Optimization Over a Finite Horizon

Consider the problem of solving the infinite dimensional problem given in (3) numerically. Perhaps the most common solution method in integrated assessment is the finite horizon method (FH). Suppose we truncate the model (3) by assuming utility is constant after T periods. Then:

$$W = \max_{C_t \in \Gamma(S_t)} \sum_{t=0}^{T} \beta^t \, \mathrm{U}(C_t, S_t) + g(C_T, S_T, T) \quad \text{Subject to:} \quad (6)$$

$$S_{t+1} = \mathrm{f}(C_t, S_t, \varepsilon_{t+1})$$

$$\varepsilon_{t+1} \sim \Omega(S_t)$$

This problem has a finite horizon because controls are chosen only for time periods up to T. What happens after T is relevant only insofar as it affects actions in the present; ie, well before T. Unfortunately, if the world is assumed to end at time period T, then it will be optimal to

slowly consume capital as T approaches, distorting decision well prior to T. There are a number of approaches to dealing with the "terminal condition problem" in FH models (see Blitzer et al, 1975). The simplest approach is to assume the world ends at time period T, relying on the discount rate and a large value of T to minimize effects on the present. Alternatively, one may assume that the economy reaches a steady state in time period T. In such a case, the model can be rewritten as:

$$W = \max_{C_t \in \Gamma(S_t)} \sum_{t=0}^{T-1} \beta^t \, \mathrm{U}(C_t, S_t) + \frac{\beta^T}{1-\beta} \mathrm{U}(C_T, S_T) \qquad \text{Subject to:} \qquad (7)$$

$$S_{t+1} = \mathrm{f}(C_t, S_t, \varepsilon_{t+1})$$

$$\varepsilon_{t+1} \sim \Omega(S_t)$$

$$S_{T+1} = \bar{S}$$

Assuming T is large enough, the economy and climate can be expected to be at a steady state well before the terminal condition nears and the discount rate effectively reduces concern about far future utilities to zero regardless. Hence truncation should have no effect on the optimal decisions (but, to save computation time, set no higher than that). However, if the terminal period is set too low, the economy may be forced to converge to the steady state much more quickly than would otherwise be the case. For example, suppose the steady state temperature is 4 degrees above current levels. The temperature changes very slowly over time; large changes requires very large emissions. Hence convergence to the steady state is quite slow (it takes centuries to millenia to stabilize sea levels). If T is set too low, the optimization routine may be forced to increase emissions drastically in all periods in order reach the steady state by period T.

Another approach is to specify that g in Eqn. (6) is a discounted sum of a stream of future utilities, but assume future utility is simply utility in time period T, growing at a fixed rate, r:

$$W = \max_{C_t \in \Gamma(S_t)} \sum_{t=0}^{T} \beta^t \, \mathrm{U}(C_t, S_t) + \sum_{t=T+1}^{\infty} \beta^t (1+r)^{t-T} U(C_T, S_T) \qquad \text{Subject to:} \qquad (8)$$

$$S_{t+1} = \mathrm{f}(C_t, S_t, \varepsilon_{t+1})$$

$$\varepsilon_{t+1} \sim \Omega(S_t)$$

A similar approach involves assuming the capital stock grows at some exogenous rate.

Regardless of how the terminal condition is set, the appropriate terminal period is quite sensitive to parameters such as the discount rate and the rate of growth of population. Hence the terminal period must often be reset during sensitivity analysis. Indeed, setting the terminal period is perhaps the most difficult part of the problem. Once the terminal condition is set appropriately, the problem is equivalent to a standard maximization problem. Let η_c be the dimension of the static decision vector, then simply choose $(T+1)\eta_c$ decisions to maximize a somewhat complex, but finite objective function. However, there is one complication. The problem is not recursive and so all decision variables are evaluated simultaneously. Modern optimization software is quite efficient and can handle problems with many states and controls.

Stochastic problems where certainty equivalence does not hold are particularly difficult to solve in a finite horizon context. Typically stochastic models involve some random variable whose realization influences how the economic and climate system evolves. Therefore, the entire tree of random variables (and the expectations of the random variables) over time must be computed. In a recursive structure, the programmer need only look one step forward and hence consider only the random variables associated with the next time period. Hence for all practical purposes, only the most simple distributions are considered with the FH method. For example, in Peck and Tiesburg (1989) there are only two states and a single random variable which is resolved at a fixed point in the future. Kolstad (1996) considers a modestly more detailed stochastic structure which generates a significantly more complex FH model. Infinite horizon methods where the state of a random variable or the state of information is represented explicitly is a much more natural way of representing stochastic problems.

V. Solving Infinite Horizon Models

Consider the problem of finding the value function v numerically as the fixed point of equation (4). Note that if one knows v, then it is easy to calculate the optimal action to take at any point in time; simply use standard optimization to solve the right-hand-side of equation (5) for C^* (the policy variable which is of primary interest). But the real problem is finding a v which is the fixed point of equation (4). Equation (5) is really a functional equation with, as an unknown, the function v.

Our idea is to define a set of functions which is dense in a set of functions of which v is a member. It is straightforward to show that v is a member of the set of C^1 functions over a compact set $A \subset \Re^n$, where A is the subset bounded by the minimum and maximum sustainable values of the state variables.[3] We next define a set of functions $\Phi(S;\chi): \Re^n \times \Re^m \to \Re$, where elements of the set differ according to values of the parameter vector, χ. The function Φ is a flexible functional form, or a set of functions which is dense in the space of continuous functions over a compact set. Thus our method is similar to other methods which use flexible functional forms, except that we approximate the value function rather than the policy function or the right hand side of the Euler equations. Given that Φ is dense in the space of functions for which v is a member, then by definition for any $\lambda > 0$, there exists an m^* such that for $m > m^*$ there exists a $\chi \in \Re^m$ such that $\|\Phi(S;\chi) - v(S)\|_S < \lambda$.

The algorithm uses the Bellman equation (4) to find the value function, except that we iterate using the approximation of v, Φ, rather than the unknown v. For simplicity, let $v_0(S) = \Phi(S;\chi_0)$ be concave (but not necessarily strictly so). Since we know v_0, for any value of the state variable S_i we may compute $v_1(S_i)$ via the equation (4):

---

[3] If part of the state space takes on arbitrarily large or small values, the state space can be mapped into a compact space by taking a bounded, monotonic transformation of the state variables.

$$v_1(S_i) = \max_{C \in \Gamma} \{ \mathrm{U}(S_i, C) + \beta \mathrm{E}[\Phi[g(S_i, C, \varepsilon); \chi_0]] \} \tag{9}$$

We use a sequential quadratic programming approach (with analytic first derivatives) for the constrained non-linear maximization in equation (9). We evaluate the expectation operator using a standard numerical quadrature routine (typically 12-24 points).

We use the resulting set of points $v_1(S_1) \ldots v_1(S_p)$ to approximate v_1 with the flexible functional form $\Phi(S; \chi_1)$. The most difficult part of the algorithm is obtaining a good approximation of v_1. Here we must define the two main features of the approximation part of the algorithm: (1) how to choose the data set $S_1 \ldots S_p$ and (2) how to approximate v_1 with $\Phi(S; \chi_1)$. For (1), we follow discrete grid methods and evenly space grid points between the minimum and maximum sustainable values of each state variable. For each state variable, assign p/n grid points, evenly space between the minimum and maximum sustainable values of the state. The parameter p is set as large as is computationally possible to minimize error, although p is much smaller than in a typical pure grid based approach. Note that, like discrete grid methods, the approximation improves if the algorithm places more points where the value function has significant curvature (see Trick and Zin, 1993 and 1995). This is often important in integrated assessment models, where the value function is nearly constant in some state variables (see Kelly and Kolstad, 1999a). Next, we approximate v_1 by choosing χ_1 such that:

$$\chi_k = \arg\min_{\chi} \left\| \Phi(S; \chi) - v_k(S) \right\| \tag{10}$$

with k=1. An obvious norm is L^2; that is, find the χ which minimizes the sum of squared residuals over the set of values $v_1(S_1) \ldots v_1(S_p)$. We solve the above minimization with an unconstrained Gauss-Newton algorithm, using analytical first derivatives and second derivative

information (alternatively, one could use algorithms which exploit features of the Hessian matrix common to least squares problems, see Dennis, 1977 for a survey of methods).

We use the approximation for v_1 to find a v_2 using the Bellman equation. Or more generally, we use the approximation for v_k to find v_{k+1}, using eqn. (4):

$$v_{k+1}(S_i) = \max_{C\in\Gamma}\{U(S_i, C) + \beta E[\Phi[g(S_i, C, \varepsilon); \chi_k]]\} \tag{11}$$

The idea is that since v_{k+1} is closer to v than v_k, then if $\Phi(\chi_{k+1})$ is sufficiently close to v_{k+1}, then $\Phi(\chi_{k+1})$ is closer to v than v_k. Hence, using $\Phi(\chi_k)$ on the right hand side of the recursive version of Bellman's equation results in a v_{k+1} which is closer to v than v_k. We then approximate v_k using the flexible functional form and repeat until:

$$\|v_k(S) - v_{k-1}(S)\|_S < \eta^* \tag{12}$$

Here the norm is the sup norm and η^* is a small number (typically 10^{-4} or less). The resulting flexible functional form approximates the value function: the parameters of Φ form a sequence $\{\chi_0, \chi_1, ...\}$ which converge to a χ^* which defines the approximate solution, $\Phi(S;\chi^*)$, to the value function v.

Finally, we must specify the flexible functional form. There are a number of alternative parametric families available for use as Φ. The primary requirements are that the flexible functional form be dense in the space of continuous functions over a compact subset of $\Re^n$ and that the parameterization be computationally efficient. Each extra parameter in χ adds computation time, so ideally we require a flexible functional form which approximates well with relatively few parameters. Unfortunately there is little guidance from the literature here. Judd (1991) argues for the use of a series of Chebycheff polynomials as more computationally efficient. Neural networks have a particularly compact representation, for which the number of

parameters does not increase exponentially with the number of state variables (as in say a polynomial approximation).

In our work, we use a one layer, feed forward neural network as our flexible functional form. Specifically, let:

$$\Phi(S;\chi) = \chi_{1l}^{'} \tanh\left(\chi_{2\ell}^{'} S + \chi_{3\ell}\right) + \chi_4 \qquad (13)$$

where χ_{2l} is a matrix, χ_{1l} and χ_{3l} are vectors, and χ_4 is a scaler component of the parameter vector χ and S is the state vector. The hyperbolic tangent function in (13) is known as the "squashing function," and can be any strictly increasing function $\sigma : \Re \to [a, b]$, where *a* and *b* are finite. There is little guidance in the literature as to which squashing function to use, and our experience is that the more common ones (tanh and $1/\left(1 + e^{-x}\right)$) work similarly.

Other versions of neural nets add additional sums of squashing functions ("layers") or additional weights ("connections"). For example, one could replace *S* with a weighted sum of tanh functions which are in turn functions of *S* or replace χ_4 with a linear function of χ_{2l}S. However, additional layers or connections are not needed here since it is well known that the neural network in (13) has the denseness property described above (see for example, White 1992).

Although we refer to our neural network in terms of statistics and approximation theory, one could also refer to the neural net in its original terms: *S* is the "input," phi is the "output," X are the "weights" which are "trained" on data *S;v* by minimizing sum of squared errors (described below). For an excellent discussion of the properties of neural networks, see White (1992).

An important decision is the number of approximating functions to use (*l*), which determines the number of parameters, since $m = n(l + 2) + 1$. When observations are subject to independent and identically distributed noise, the statistics literature often suggests setting m so that the number of parameters equals the square root of the number of observations or less (see,

for example White, 1992). Since the value function is deterministic we can use at least the square root of the number of observations without concern with over-fitting. To be safe, we set the maximum number of parameters equal to the square root of the number of observations; hence the maximum number of approximating functions is: $l = \frac{\sqrt{p}-1}{n} - 2$. In theory, however, the only requirement for convergence of the algorithm is that l goes to infinity as p goes to infinity.[4]

Accuracy of the approximation depends primarily on the accuracy of the neural net in approximating the value function. Accuracy of the neural net in turn depends on two factors: the number of approximating functions (l) and the size of the data set p relative to the size of the compact set A. As l increases, the neural net better fits the data (λ shrinks). However, more approximating functions increases computation time both because there are more parameters to estimate and because each approximating function is evaluated repeatedly in the constrained optimization part of the Bellman's equation (9). As the size of the data set p increases, the neural network has more observations to fit the underlying function, thus increasing accuracy. Without such data, the neural net may under or overestimate curvature in the value function. Of course, as the size of the data set increases, so does computation time both because estimating the parameters over a larger data set requires more computation and because generating a larger data set requires more optimization in the Bellman equation. In practice, most of the error tends to come from a small data set; the neural network easily approximates the typically smooth, concave, deterministic value function without a lot of parameters. Hence we typically set p as large as possible while keeping l well below the value which equates the number of parameters to the square root of the number of observations.

Although p is large relative to l, the accuracy of the neural net must be sufficient to insure convergence. If the neural net is not accurate enough, the data $v_1(S_1)\ldots v_1(S_p)$, generated from

[4] This result plus general convergence results of the algorithm may be found in Kelly (1998b)

the neural network might result in v_1 which is farther from v than v_0, and the algorithm would cease to contract. Furthermore, since there are two sources of error, one from iterating on the value function and the other from the neural network approximation, the accuracy of the approximation in general larger than $\eta*$. Kelly (1998b) shows theoretically that if $\eta**$ accuracy is desired then set:

$$\eta^* = \frac{2}{1+\beta}\eta^{**}$$

$$\lambda = \frac{1-\beta}{1+\beta}\eta^{**}$$

Here λ is with respect to the sup norm. If the above settings are used, the algorithm converges to a value function which differs from v by at most $\eta**$.

Beyond the above considerations, computation time generally depends on the number of states and random variables in the model. The algorithm solves a simple model with one state and one random variable in minutes. A particularly large problem (Kelly and Kolstad, 1999a) of seven states and two random variables with $p = 57{,}600$ and $l = 16$ took approximately 24 hours when programmed in C on a 75mHz Sun Sparc 20 workstation, which is already slow by today's standards. However, the time step in the model was a decade (leading to a low discount rate) and the initial value function was a concave function which included the steady state.

This approach is essentially the same as the contraction mapping constructive proofs of existence of solution to dynamic program (see, for example, Stokey and Lucas, 1989). These proofs define an operator (T) from the space of continuous functions to the space of continuous functions. Choosing an arbitrary continuous function (v) on the value function on the right-hand side of equation (12), the left-hand sets of the equation defines T(v). If T is a contraction mapping, then multiple applications of T to v eventually converge to a fixed point, a solution to equation (12). In our case, we are dealing with a restricted set of functions, $\Phi(S,\chi)$. Iteration on χ as described above is essentially the same as multiple applications of T.

The algorithm for finding the value function is shown below.

1. Start with an initial χ_0 and resulting flexible functional form Φ. Let k = 0.
2. Using (11), compute $v_{k+1}(S_1)\ldots v_{k+1}(S_p)$ by computing the optimal controls given the approximate value function $\Phi(\chi_k)$
3. Using (10) compute χ_{k+1}, the parameter vector which minimizes the sum of squared differences between v_k, and $\Phi(\chi_{k+1})$
4. Repeat steps (2-3) until $\|v_k - v_{k-1}\| < \eta^*$.

This defines the algorithm which solves the model (5). Kelly (1998b) shows that the algorithm converges to the true v as k and p go to infinity:

$$\lim_{k,p\to\infty} \left\| \Phi\left(\chi_k^{m(p)}\right) - v \right\| = 0$$

While the algorithm is much more computationally intensive than the finite horizon approach in section (IV) using standard optimization software, two clear advantages of this approach are: a) stochasticity is easily represented; and b) once solved, the solution (optimal actions) for all values of the state vector is also known -- no further computations are necessary for other values of the state vector. This is a very important advantage of dynamic programming. It allows straightforward comparative statics analysis as well as great flexibility in policy analysis and simulation. Furthermore, since the algorithm approximates the value function with a flexible functional form across the entire state space, the model is approximated well even away from the steady state. Furthermore, the algorithm is capable of becoming arbitrarily accurate, unlike the LQ method where the accuracy is bounded by the accuracy of the approximation of the return function. Furthermore, since the value function is in a single dimension, only one value function needs to be computed instead of multiple policy functions or first order conditions.

VI. Conclusions

Integrated assessment has become an important tool of applied environmental economics analysis, particularly in the climate change area. A number of numerical models of the environment-economy have been developed for doing integrated assessment. Many of these models are based on the classic Ramsey optimal growth framework, though considerably embellished to deal with the complexities of the economy and climate. Computational issues have become important in solving these empirical optimal growth models.

We shown how the finite horizon and infinite horizon optimal growth models are of similar structure. We have also examined solution techniques for both classes of models. Most integrated assessment models are finite horizon deterministic models, solved as nonlinear optimization problems. With stochastic elements, the infinite horizon models tend to be easier to solve than the finite horizon models, yet solution methods are not widely available.

In this paper we have presented an algorithm for solving the infinite horizon optimal growth model, based on the recursive version of Bellman's equation of optimality. Future research would likely entail comparison of the efficiency of the several approaches to solving infinite horizon models. The state-of-the-art is still not entirely satisfactory, with all methods being highly computationally intensive and hampered by the curse of dimensionality.

compecon99.doc

References

Blitzer, Charles R., Peter B. Clark, and Lance Taylor, *Economy-Wide Models and Development Planning* (World Bank, Washington, DC, 1975).

Cass, D., "Optimum Growth in Aggregate Model of Capital Accumulation", *Review of Economic Studies*, 32:233-46 (1965).

Christiano, L., "Is Consumption Insufficiently Sensitive to Innovations in Income?", *American Economic Review Papers and Proceedings*, 77:337-341 (1987b).

Christiano, L., "Why Does Inventory Investment Fluctuate So Much?", *Journal of Monetary Economics*, 21:247-280 (1988).

Christiano, L. and Eichenbaum, M., "Is Theory Really Ahead of Measurement? Current Real Business Cycle Theories and Aggregate Labor Market Fluctuations," Technical Report 2700, National Bureau of Economic Research Working Paper.

Cooley, T. and Hansen, G., "The Inflation Tax in a Real Business Cycle Model", *American Economic Review*, 79:733-48 (1989).

Grossman, G. and Krueger, P., "Economic Growth and the Environment", *Quarterly Journal of Economics*, 112 (1995).

Dennis, J., "Nonlinear Least Squares," pp. 269-312 in D. Jacobs (ed.), *State of the Art in Numerical Analysis*, Academic Press (1977).

Den Haan, W. J. and Marcett, A., “Solving the Stochastic Growth Model by Parameterizing Expectations”, *Journal of Business Economics and Statistics*, 8(1):99-113 (1990).

Hansen, G., “Indivisible Labor and the Business Cycle”, *Journal of Monetary Economics*, 16:309-327 (1985).

Hansen, G. and Sargent, T., “Straight Time and Overtime in Equilibrium”, *Journal of Monetary Economics*, 21:281-308 (1988).

Kenneth L. Judd, 1991, Minimum Weighted Residual Methods for Solving Dynamic Economic Models, unpublished manuscript, Hoover Institution, Stanford University.

David L. Kelly, “On Kuznets Curves Arising From Stock Externalities,” mimeo, University of Miami Economics Department (1998a).

David L. Kelly, "Convergence results for Dynamic Programming Using Neural Networks" mimeo, Department of Economics, University of Miami (1998b).

David L. Kelly and Charles D. Kolstad, “Malthus and Climate: Betting on a Stable Population,” mimeo, UCSB Economics Department, February 1996.

David L. Kelly and Charles D. Kolstad, “Bayesian Learning, Growth, and Pollution,” *J. Econ. Dynamics and Control*, 23(4):491-518 (1999a)

David L. Kelly and Charles D. Kolstad, "Integrated Assessment Models for Climate Change Control," In: Henk Folmer and Tom Tietenberg, (eds), *International Yearbook of*

Environmental and Resource Economics 1999/2000: A Survey of Current Issues, Cheltenham, UK: Edward Elgar, (1999b).

David L. Kelly, Charles D. Kolstad, Michael E. Schlesinger, and Natalia G. Andronova, "Learning About Climate Sensitivity from the Instrumental Temperature Record" mimeo, UCSB Economics Department, January, 1998.

Charles D. Kolstad, "Looking vs. Leaping: The Timing of CO_2 Control in the Face of Uncertainty and Learning," pp. 63-82 in Y. Kaya, N. Nakicenovic, W. D. Nordhaus and F. L. Toth (eds), *Costs, Impacts and Benefits of CO_2 Mitigation*, IIASA, Austria (June 1993).

Charles D. Kolstad, "The Timing of CO_2 Control in the Face of Uncertainty and Learning," Chapter 4 in E. van Ireland (ed), *International Environmental Economics*, Elsevier, Amsterdam (1994).

Charles D. Kolstad, "Learning and Stock Effects in Environmental Regulation: The Case of Greenhouse Gas Emissions," *Journal of Environmental Economics and Management*, 31:1-18 (1996).

Koopmans, "On the Concept of Optimal Economic Growth," in *An Economic Approach to Development Planning*, North Holland, 1965.

Kydland, F. and Prescott, E., "Time to Build and Aggregate Fluctuations," *Econometrica*, 50:1345-70 (1982).

A.S. Manne and R.G. Richels, *Buying Greenhouse Insurance: The Economic Costs of CO2 Emission Limits* (MIT Press, Cambridge, 1992).

A.S. Manne, R. Mendelsohn and R. Richels, "A Model for Evaluating Regional and Global Effects of GHG reduction policies," *Energy Policy*, **23**:17-34 (1995).

D. Meadows, et. al., *The Limits to Growth* (Universe Books, New York, 1972).

W. D. Nordhaus, *Managing the Global Commons: The Economics of Climate Change* (MIT Press, Cambridge, Mass., 1994).

W. D. Nordhaus and Z. Yang, "A Regional Dynamic General-Equilibrium Model of Alternative Climate-Change Strategies," *American Economic Review*, 86:741-65 (1996).

S. C. Peck, H-P Chao, and T.J. Teisberg, "Optimal Control and Learning Strategies when Environmental Costs are Cumulative," *Proc., IFAC/IFORS/IAEE International Symposium on Energy Systems, Management and Economics*, Oct. 25-7, 1989, Tokyo.

S.C. Peck and T.J. Teisberg, "CETA: A Model for Carbon Emissions Trajectory Assessment," *Energy Journal*, 13(1):55-77 (1992).

S.C. Peck and T.J. Teisberg, "Global Warming Uncertainties and the Value of Information: An Analysis Using CETA," *Resource and Energy Economics*, **15**:71-98 (1993).

Frank Ramsey, 1928, A Mathematical Theory of Saving, Economic Journal.

Nancy L. Stokey and Robert E. Lucas Jr., 1989, Recursive Methods in Economic Dynamics (Harvard, Cambridge, Mass.)

Trick, M. and Zin, S., "A Linear Programming Approach to Solving Stochastic Dynamic Programs," mimeo, Carnegie Mellon University, 1993.

Trick, M. and Zin, S., "Adaptive Spline Generation: A New Algorithm for Solving Stochastic Dynamic Programs," mimeo, Carnegie Mellon University, 1995.

Weyant, et. al., "Integrated Assessment of Climate Change: An Overview and Comparison of Approaches and Results," pp. 367-439 in J. P. Bruce, et. al. (eds), *Climate Change 1995: Economic and Social Dimensions of Climate Change*, Cambridge University Press, Cambridge (1996).

H. White, *Artificial Neural Networks* (Blackwell Publishers, Cambridge, Mass., 1992).