SOLVING NONLINEAR STOCHASTIC EQUILIBRIUM MODELS "BACKWARDS"

by

Christopher A. Sims

Discussion Paper No. 206, January 1985

REVISED VERSION

Center for Economic Research
Department of Economics
University of Minnesota
Minneapolis, Minnesota 55455

Abstract

A method is presented for cheaply generating simulated solution paths for dynamic stochastic optimization models. There is no linearity restriction on the models, and the method involves no approximation. It does, however, abandon the usual convention that the process generating the exogenous disturbances is "given", instead generating mutually consistent processes for exogenous disturbances and choice variables.

SOLVING NONLINEAR STOCHASTIC EQUILIBRIUM MODELS "BACKWARDS"

by Christopher A. Sims

November 1984*

---

* The October version of this paper contained an important error in the second example. This version corrects that error, explains how to make the search for stable backwards solutions more systematic, and adds some comments about use of the methods for estimation.

Economists have in recent years been applying the assumption that economic agents dynamically optimize under uncertainty. Models embodying this assumption can be very difficult to solve, however, and those which have been used have therefore been highly simplified or approximate. We have solution methods for "quadratic-linear" models, in which agents have linear decision rules, and for general nonlinear models with small discrete state spaces. Kydland and Prescott have shown the way toward approximating non-quadratic-linear models with quadratic linear ones, thereby making them soluble, but their method is essentially equivalent to assuming that agents apply certainty equivalence to their problems. For some issues the effects of the failure of certainty equivalence are of central interest. This paper suggests a class of solution methods which allow analysis of a broader range of stochastic equilibrium models than has heretofore been practical.

Most of the methodology economists have applied in this area is borrowed from engineering. For an engineer, just as for an economic agent, solving the stochastic optimization problem requires a method for finding the appropriate decision rule by which to make controlled variables respond to a given forcing stochastic process. An economist, however, can take a more symmetric view of control variables and forcing variables. In modeling behavior, he is interested in the mapping between the stochastic processes for controlled and uncontrolled variables. Since he may observe data on both processes, both are equally "given" to him. In fact the variables taken as given by one group of agents in his model may well be taken as choice variables by other agents. For an economist, therefore, a method of taking an arbitrary stochastic process for controlled variables and determining the corresponding stochastic process for the uncontrolled variables may be as useful as the standard engineering solution methods which work in the reverse direction. It turns out that in many cases of interest finding this "backwards" mapping from controlled processes to forcing processes is much easier than the reverse.

1. The method

Let C be a kx1 vector of control variables, S be an mx1 vector of state variables, v be an nx1 vector of disturbances, and M be an mx1 vector of stochastic Lagrange multipliers. We suppose that the problem is

$$\max_{C,S} E \sum_{t=1}^{\infty} U(C_t)R^t$$

subject to

1) $$S_t = g(C_t)+h(S_{t-1})+v_t \quad ,$$

where variables dated t and earlier are known when C_t is chosen.

The first order conditions for this problem are

2) $$DU_t = M_t Dg_t$$

3) $$M_t = RE_t[M_{t+1}]Dh_t,$$

where M_t is a stochastic Lagrange multiplier. The problem is hard in general mainly because the appearance of the expectation in (3) introduces an integral into the set of equations which must be solved; knowing the conditional distribution of v_{t+1} at t does not allow us to compute the integral until we also know the mapping from v to M. But suppose that instead we introduce an arbitrary form for the stochastic process followed by M. If we have chosen the form to be convenient, say a linear autoregression or a linear autoregression in logarithms, then the expectation on the right of (3) will be easily found as a function of information dated t and earlier. Then (2) and (3) constitute m+k equations in the m+k variables S_t, and C_t, together with M's dated t and earlier. If they have full rank Jacobian, we can solve the two sets of equations to express S_t and C_t in terms of terms of M dated t and earlier. Substituting these solutions into the constraint (1), we obtain m equations in v_t and M's dated t and earlier. Thus, if we generate a simulated path for the M_t process, we can use the first order conditions and the budget constraint to solve for C, S, and v, obtaining a realization of C, S and v as a joint stochastic process which is consistent with optimal choice of C and S taking the process v as given. If we began with a stationary process for M, the processes for C, S and v will all be stationary as well.

Consider a simple example, a stochastic growth model. Individuals maximize

4) $$E[\sum_{t=1}^{\infty} .95^t C_t^{5}]$$

subject to

5) $$C_t^2 + K_t = \log K_{t-1} + v_t \quad .$$

The first order conditions are

6) $$.5/C_t^{5} = 2C_t M_t$$

7) $$M_t = .95E_t[M_{t+1}]/K_t \ .$$

Let M_t be generated by

8) $\quad \log M_{t+1} = .8 \log M_t + u_{t+1}$,

where u is serially independent and Normal with mean zero, standard deviation .05. Then if information at t consists of M_s for $s<t+1$,

9) $\quad E_t[M_{t+1}] = M_t^{.8} e^{.00125}$

Rearranging (6) and using (9) in (7) gives us

10) $\quad C_t = .39685\ M_t^{-.66667}$

11) $\quad K_t = .9512\ M_t^{-.2}$.

Substituting (10) and (11) in (5) produces an expression for v_t in terms of M_t. In this model the natural definition of the gross interest rate is

12) $$r_t = M_t C_t / (R E_t[M_{t+1} C_{t+1}])$$

$$= 1.0525\ M_t^{.06667}\ .$$

From (10)-(12) it is obvious that K and C move together, with C the more volatile of the two, and that r moves in the opposite direction in this solution. To obtain more interesting and realistic dynamics, we could make the M process second or higher order, add a constant to it, or the like. We will discuss such possibilities further below, where we consider cases where the choice of the M process cannot be made so freely.

2. Generalizing the method: stability problems.

The general form of the problem considered so far is restricted in that the state variable[1]

---

[1] Though we call S a "state variable", it does not fit the usual definition of one in dynamic programming. Because we have not restricted the nature of the dependence of v on past information, S does not in general contain all information relevant for predicting the future of the system.

---

is excluded from the utility function, the disturbance term in the constraint (1) enters additively, and the constraint is separable. When these restrictions are relaxed, the method does not work so simply. Relaxing these restrictions, we have the problem

$$13) \qquad \max E\left[\sum_{t=1}^{\infty} U(C_t, S_t) R^t\right]$$

subject to

$$14) \qquad S_t = g(C_t, S_{t-1}, v_t) \ .$$

Now the first order conditions are

$$15) \qquad D_1U_t = RM_tD_1g_t \quad \text{and}$$

$$16) \qquad D_2U_t = RE_t[M_{t+1}D_2g_{t+1}] - M_t \ .$$

To implement a backwards solution, we need to avoid calculating the expectation in (16). We can always do this by simply introducing an "innovation" into (16) while eliminating the expectation operator, to arrive at

$$17) \qquad D_2U_t + w_{t+1} = RM_{t+1}D_2g_{t+1} - M_t \ .$$

If $E_t[w_{t+1}]=0$, (17) expresses the same condition as (16). Taking w as given, (14), (15) and (17) constitute a system of 2m+k difference equations in the 2m+k variables C, S, and M, plus the n variables v. To get a determinate solution, we must impose n additional conditions on the system. The usual practice would be to make these n additional equations ones in which v_{t+1} is generated from its own past plus serially independent shocks. However with such a set of auxiliary conditions, the system is nearly always unstable. Borrowing deterministic methods, we could search over initial choices of C to find one yielding a stable path, with a given generated path for the disturbances w in (17) and for those in the auxiliary v equations. However in doing so we would be generating dependence between current C and the future disturbances, violating the assumption required to justify (17) as a replacement for (16).

If we are not tied to a precise form for the v process, we need not make our n auxiliary equations ones which generate v from its own past. Any relation among the variables in the system, including the v's and w's, which leads to a nonsingular Jacobian for the system, will suffice to allow the system to determine C, S, M and v with a given path for w. If the problem has any stationary solution, then there must be some way to choose the n additional equations in such a way that the solution is stable.

Often the quickest way to do this is to make a shrewd guess, as in our first example above. In choosing a process for M_t in that example we in effect added an error

$$w_{t+1}=.95(M_{t+1}-E_t[M_{t+1}])/K_t \tag{18}$$

to (7), and added (8) as an auxiliary equation. Note that though (18) and (8) seem to involve two disturbances u and w, the two are related by an analytic expression, so there is really only one disturbance process.

However shrewd guesses do not not always work, and when they don't a more systematic approach is possible! One can linearize the system (14), (15), (17) about the steady state solution, then choose the auxiliary equations to make all the roots of the completed linearized system fall in the stable region. This method can become quite burdensome for large systems because of the analytic differentiation required, but it is nice to know it is available as a last resort when a reasonable number of shrewd guesses have failed to work. It should also be noted that often one is interested in exploring the range of solutions the model can generate as the nature of the exogenous disturbances varies. An explicit local stability analysis can be helpful, or even essential, in doing a thorough job of this.

The shrewd guess method is usually so much easier than explicit local stability analysis of the system, that when shrewd guessing fails, one should usually consider the possibility of modifying the model's stochastic specification. For example, a model with a single disturbance process v leaves room for only one auxiliary equation, and that one equation must stabilize the system. Often there is no strong reason to suppose a single disturbance, and adding an additional one, by allowing two dimensions for the shrewd guess approach to work in, may make an easy problem of a hard one.

It is important in practice to examine systematically the structure of the difference equations one is generating with this method. A commonly encountered pitfall is the situation where the Jacobian of the system with respect to current values of C, S, M and v is singular, yet a solution appears possible. For example, S_t might not enter the system at all, while S_{t-1} does. If one redates S, letting $S_t^*=S_{t-1}$, one might then find a system that generates a stable solution for all variables. This procedure is not legitimate, because it makes S_t^*, along with C_t, M_t, and v_t, depend on the shock w_t, violating the model's information assumptions; in effect it allows the model to see the exact value of w_t when choosing S_{t-1}. This pitfall is really the same as the problem of unstable solutions. If S_t did enter the system but very weakly (so the Jacobian was near singularity), then the likely result would be a highly unstable system of difference equations.

We illustrate the foregoing pitfalls and possibilities with a simple growth model.

The objective is

$$\max E\sum_{t=1}^{\infty} \log(C_t)R^t$$

subject to

19) $$C_t + K_t = v_t K_{t-1}^{.5} \quad .$$

This has the FOC's

20) $$1/C_t = M_t$$

21) $$M_t = .5RE_t[M_{t+1}v_{t+1}]/K_t^{.5} \quad .$$

A natural approach might be to take

22) $$Z_t = M_t v_t / K_{t-1}^{.5} \quad .$$

This has the appeal that it converts (21) to

23) $$M_t = .5RE_t[Z_{t+1}] \quad .$$

Generating Z and its predictor as a linear autoregression in logs, we can solve for the C path from (20) and (23). Then (22) gives us

24) $v_t = C_t Z_t K_{t-1}^{.5}$.

Substituting (24) in the budget constraint (19) gives us

25) $C_t + K_t = C_t Z_t K_{t-1}$.

It is easy to see that (25) is not a promising difference equation from which to generate K. For example, if Z is nearly constant, then ZC is nearly 1/R (from (22)), implying that (25) yields explosive solutions for K.

Guessing at a choice for Z which will work in this problem is quite unlikely to succeed. The lazy man's approach is then to wonder whether we need to insist that there is only one kind of disturbance in the technology (19). Replacing (19) with

26) $C_t + K_t = v_t K_t^{.5} + u_t$

may seem innocent enough from the point of view of the economics of the problem, and it leaves the FOC's unaffected. But with two error terms we can have two auxiliary equations. We can, for example, generate K directly as well as our Z from (22). With stable paths for both C (from (20) and (23)) and K, the analogue of (25) is

27) $C_t + K_t = C_t Z_t K_{t-1} + u_t$,

which determines a stable path for u_t. Even if one wished to analyze the case where shocks to the level of production were highly correlated with shocks to the rate of return, i.e. where v were the main source of disturbance, it might be best to follow this two-shock strategy and experiment with the specification for the equations generating K and Z to see what is needed to make the variation in u much less important than that in v.

However, a direct approach to analyzing the instability problem, preserving the single disturbance v, is also possible. Substituting (20) in (21) and eliminating the E_t by adding a shock, we obtain

$$28)\qquad 1/C_{t-1} + w_t = .5Rv_t/(K_{t-1}^{.5}C_t)\ .$$

Treating (19) and (28) as a pair of difference equations in the three variables C, K, and v, we linearize them about steady state values, to obtain as the matrix of polynomials in the lag operator in the linearization

$$\begin{bmatrix} 1 & 1 - .5vK^{-.5}L & -K^{.5} \\ .5Rv/(K^{.5}C^2)-(1/C^2)L & [.25Rv/(K^{1.5}C)]L & -.5R/(K^{.5}C) \end{bmatrix}.$$

Here we are using unsubscripted variables to designate steady state values. We can simplify the analysis by creating a zero in the upper right of this matrix, which can be done by multiplying the second row of the matrix by 2KC/R and subtracting the result from the first row. It also helps to multiply the second row by C^2 and use the fact that in steady state $.5RvK^{-.5}=1$. None of these transformations affect the roots of the determinant of the matrix which results when we add an equation to the system. The simplified matrix is

$$\begin{bmatrix} 1 - 2K/(CR)(1-L) & 1 - (2/R)L & 0 \\ 1-L & (.5C/K)L & -C/v \end{bmatrix}$$

Now we would like to add an equation to this system in such a way that all the roots of the determinant lie outside the unit circle. Note, though, that in steady state $K/C=R/(2-R)$. Using this fact makes the term in the upper left of the matrix $-R+2L$. This is, however, just a scalar multiple of the $1-(2/R)L$ term in the second position of the first row. Thus no matter how we complete this matrix, its determinant will have one root equal to $R/2<1$. This reflects the fact that there is no stationary solution to this system in which the Euler equation (28) has a non-zero disturbance. We have here a model in which optimal behavior is myopic.

To see this, note that by combining (19) and (28) we can obtain an equation in K/C alone,

$$29) \qquad 1 + K_t/C_t = (2/R)(K_{t-1}/C_{t-1}) + \eta_t,$$

where $E_t[\eta_{t+1}]=0$, all t. If η_t is stationary and with non-zero variance, any path on which K_t/C_t is positive makes K_t/C_t explode. Therefore stationary solutions to the model must make the variance of η_t zero. That is, the ratio K_t/C_t is kept fixed at the value of $R/(2-R)$, which will keep (29) true identically with $\eta_t=0$. Optimal behavior is very simple -- keep consumption at a fixed fraction of the capital stock. There is no dependence of this rule on the stochastic behavior of disturbances. Of course once we recognize this characteristic of the solution, generating simulated solution paths is trivially easy. But before we recognize it, the search for shrewd guesses at equations which will complete and stabilize the model may absorb a lot of wasted energy.

A less drastic and possibly more instructive example of a model for which it is difficult to find a stable solution arises when we add to the right-hand side of (19) a term to make the rate of depreciation less than one, giving us

$$30)\qquad C_t + K_t = v_t K_{t-1}^{.5} + .9K_{t-1} \quad .$$

The solution to this model is not myopic, yet it is not easy to guess at a stabilizing completion of the model. Here the Euler equation becomes

$$31)\qquad w_t + 1/C_{t-1} = R(.5v_t K_{t-1}^{-.5} + .9)/C_t$$

and the matrix polynomial operator for the linearized system (30)-(31) is

$$\begin{bmatrix} 1 & 1-(.5vK^{-.5}+.9)L & -K^{.5} \\ \{R[.5vK^{-.5}+.9]-L\}/C^2 & (.25RvK^{-1.5}/C)L & -.5RK^{-.5}/C \end{bmatrix}$$

Here in steady state $.5vK^{-.5}=(1-.9R)/R$ and $K/C=R/(2-1.9R)$. With R=.95, after multiplying the second row by C^2, we can rewrite the matrix as

$$\begin{bmatrix} 1 & 1-1.053L & -K^{.5} \\ 1-L & .01488L & -.0975K^{.5} \end{bmatrix} .$$

Multiplying the first row by .0975 and subtracting it from the second produces

$$\begin{bmatrix} 1 & 1-1.053L & -K^{.5} \\ .9025-L & -.0975+.1175L & 0 \end{bmatrix} .$$

If we add the completing equation as the bottom row of the matrix, and do not include any v term in the equation, the roots of the matrix will be just those of the 2x2 matrix in the lower left corner. The new equation cannot generate C from its own past alone or K from its own past alone. If the new equation involved only one of C or K, the determinant of the completed matrix would have one of the two elements of the second row above as one of its factors, guaranteeing one root inside the unit circle. But suppose we add an equation of the form

32) $$\log C_t = a + b \log C_{t-1} + c \log K_{t-1}.$$

(We include no error term in (29). We could add an arbitrary linear function of w_t to the right-hand side without affecting the analysis. However, since the system is singular, with only a one-dimensional driving process, there will in any case be some function of C_t, K_t, and v_t which can be predicted without error, so the presence or absence of an error term in (32) does not correspond to any qualitative difference in the model.)

With (32), linearized about steady state, generating the third row of the system's matrix polynomial, it takes the form

$$\begin{bmatrix} 1 & 1-1.053L & -K^{.5} \\ .9025-L & -.0975+.1175L & 0 \\ (1-bL)/C & -cL/K & 0 \end{bmatrix}$$

Multiplying the last row by C and using C/K=.2053, we get the determinant as

33) $$-K^{.5}[.0975-(.1853c+.0975b+.1175)L+(.2053c+.1175b)L^2]$$

For a second order polynomial $P(x)=1-mx-nx^2$, both roots lie outside the unit circle if and only if $m+n<1$, $n-m<1$, and $n<1$. This can be shown to imply that (b,c) in (33) lie in a triangle with vertices (1.23,-.23), (3.45,-2.45), and (46,-26). Thus b must exceed one and c must be below zero, with their absolute magnitudes roughly similar.

Equation (32) with b>1.23 can make sense only as part of a system. The polynomial in the lag operator applied to log(C) in it has no convergent one-sided inverse. I have run simulations of the system with parameters in the region which keeps the linearized system stable. For b and c values which are not too large it behaves very stably, with no apparent sensitivity of the stability to the sizes of assumed disturbance terms.

3. Invertibility problems.

When the backwards solution process turns out unstable, the problem is that we have completed the system in such a way that there is no stationary solution to the problem in which v_t, the exogenous disturbance, is a function of current and past values of w_t, the disturbance process we generate arbitrarily. The reverse of this can also happen -- current and past w's can turn out to contain more information than is contained in current and past shocks v. But while the instability problem invalidates the solution, the invertibility problem is only a cautionary note to the interpretation of the solution. While an engineer must take the amount of information available in making choices as given, an economist usually finds nothing unreasonable in supposing that economic agents have access to information that allows them to forecast the disturbances to the economy better than could be done from knowledge of past disturbances alone.

There are limits to this, of course. Usually it would violate the spirit of an economic model to suppose that agents know next period's technology shock exactly in this period, for example. In the most extreme case, we might end up with a "solution" to a stochastic model in which agents are in effect assumed to have perfect foresight. In most applications it will be valuable to know whether the solutions generated do imply agents have more information than is available from the history of the disturbances, even if we do not insist that the solutions be invertible.

One can check for invertibility by "reversing the reverse" solution procedure. The reverse solution procedure we have been discussing converts the problem into that of solving a set of difference equations in C, S, v, and w, in which w and initial values are taken as given in solving for C, S and v. These same equations can be used to find C, S, and w taking v and initial conditions as given. We cannot generate simulated paths for the economy this way, because to do so we would have to generate paths for v consistent with the expectation rules used in deriving the difference equations, and we do not know how to do this directly. But when we ask the question of whether dependence of w_t on initial conditions at t=0 dwindles to zero as t goes to infinity, this "reverse reverse" solution method is useful. Taking any of the simulated v paths obtained in the initial solution, we can modify the initial conditions, then find the corresponding w path. If w does depend only on current and past v, then for large t the w_t found by the reverse-reverse method based on the modified initial conditions should approach the w_t which generated the v path from the original initial conditions.

Note that in the most extreme case, where e.g. v_t is an exact function of initial conditions (hence known in advance), instead of showing up as an instability problem in the reverse-reverse solution, the invertibility problem shows up as an inability to freely modify the initial conditions. That is, while in the original reverse solution we took w's and initial condition's as given and found v from them, when we try to take initial conditions and later v's as given, solving for w, we discover we are not free to modify initial conditions without changing later v's.

It is not hard to check that, e.g., the first of our examples in this paper, defined by (4)-(8), can be reduced to a difference equation relating v to M of the form

$$34) \quad .15748\, M_t^{-2.6667} + .9512\, M_t^{-.2} = \log(.9512) - .2\log M_{t-1} + v_t \quad .$$

The linearization of (29) about M=1 is

$$35) \quad -.6029\, m_t + .2000\, m_{t-1} = v_t - 1.3560$$

where $m_t = M_t - 1$. This system is highly stable, obviously, so it appears likely that the solution is invertible.

In the example of equations (30)-(32), in solving for w from v we would generate a system whose linearization replaced the third column of its matrix polynomial by (0,1,0)', to produce

$$\begin{bmatrix} 1 & 1-1.053L & 0 \\ 1-L & .01488L & 1 \\ 1-bL & -.2053cL & 0 \end{bmatrix} .$$

It is easy to check that the determinant here is a second order polynomial in L with a coefficient of 1.053b on L^2 and a constant term of 1. Thus a necessary condition for all its roots to lie outside the unit circle is that $|b|<.95$, yet we know that all the stable solutions have $b>1.23$. Hence the solutions we generated this way are not invertible -- they imply that agents predict future v's better than could be done from knowledge of v's own past alone.

4. Higher order models.

The models we have looked at so far all have first-order lags only. Of course a model formulated with more lags can always be rewritten as a first-order model by treating the higher order lags as additional state variables. However when we expand the state vector this way, we obtain additional equations in the constraint (14), and these are "dummy equations", in which no disturbance term is appropriate. Each constraint equation in general generates a first-order condition with its own conditional expectation term. Our previous analysis would thus seem to indicate that, e.g., in a model with 5th order lags we would require a five-dimensional disturbance process, even if there is only one "non-dummy" constraint equation.

It turns out that there is no such difficulty. The first-order conditions for a high-order model with only m nontrivial constraint equations can always be reduced to m equations involving only m distinct expectational terms. Choice of the Z process for such a model will usually require some ingenuity, though.

Here's an example. The problem is

$$\max E\sum_{t=1} R^t \log C_t$$

subject to

31) $$C_t + K_t = K_{t-1}^a K_{t-2}^b + v_t \quad .$$

The FOC's are

32) $$1/C_t = M_t$$

33) $$M_t = E_t[RaM_{t+1}K_t^{a-1}K_{t-1}^b + R^2bM_{t+2}K_{t+1}^a K_t^{b-1}] \quad .$$

Equation (33) can be thought of either as obtained directly by variational principles, or can be derived from an expanded version of the problem in which the state vector is two-dimensional, consisting of K_t and K_{t-1}. If we simply set

34) $$Z_t = RaM_{t-1}K_{t-2}^{a-1}K_{t-3}^b + R^2bM_tK_{t-1}^a K_{t-2}^{b-1} ,$$

we give (33) a very simple form. However in that case (33) determines M_t from $E_t[Z_{t+1}]$, and (34) then becomes an equation relating K dated t-1 and earlier to Z dated t and earlier, which means solving it for K would contradict the information assumptions. (Here the solution would also be explosive). If we let $Y_t=K_t^a K_{t-1}^b$, however, (33) takes the form

35) $$M_t = E_t[RaM_{t+1}Y_t/K_t + R^2bM_{t+2}Y_{t+1}/K_t] \quad .$$

This suggests setting

36) $$Z_t = M_tY_{t-1} \quad .$$

Equation (36) allows us to convert (35) to

37) $$K_tM_t = E_t[RaZ_{t+1} + R^2bZ_{t+2}] \quad .$$

which, once we have a process specified for Z, determines K_tM_t. Evidently K_tM_t will be stable whenever the Z process is.

Now observe that (36) can be rewritten as

39) $$Z_t/M_tK_t = K_t^{-1}K_{t-1}^{a}K_{t-2}^{b} .$$

The left-hand side of (39) is stable if Z_t and K_tM_t are, and the right-hand side is log-linear in K, with a stable polynomial in the lag operator applied to log K so long as a+b<1. Thus the solution generates stable K for any stable Z process. It is then easy to see from (36) that C must be stable and from (31) that v must be as well.

In this kind of problem we can offer some general principles concerning the selection of a Z definition which avoids instability. The definition in (34) relates M_tY_{t-1}/K_{t-1} to Z_t through a noninvertible polynomial in the lag operator (unless b is much bigger than a). Thus when (34) becomes part of the system for solving for C and K from Z, it is likely to generate instability. We can correct this problem by modifying the right-hand side of (34) in such a way as to stabilize (34), while still allowing use of $E_t[Z_{t+1}]$ and $E_t[Z_{t+2}]$ to get rid of the expectational terms in (33). In a problem where the lags of K did not enter the constraint so symmetrically, it would not be so obvious how to do this as in this problem. It is worth something, though, to have located the equation where the difficulty occurs and provided a criterion for successful modification of the choice of Z.

5. Application of These Ideas to Estimation

In a sense, all economics is empirical. Research in pure theory aims at producing models which help us understand actual economies, just as does explicitly econmetric work. The difference is that for purely theoretical work the question of how one connects the model to observable facts about the economy is not addressed explicitly. "Goodness of fit" judgments are left to intuition. With complicated equilibrium models of the business cycle, there is a strong temptation to treat questions of fit more as a theorist does. Prescott and Kydland [1982], for example, "fit" their model on a criterion which looks at only a few moments of the data. While this kind of approach may be a practical necessity in some cases, we should recognize that a more rigorous treatment of inference is our goal. With the cheaper simulation methods suggested in this paper, a more rigorous approach to inference may prove possible.

One direction to go is to recognize the connection of these methods with the use of Hansen's generalized methods of moments methods for estimation based on stochastic first order conditions. Hansen's idea can be thought of as paralleling this paper in suggesting that the expectational equation (16) be replaced with (17), then observing that (17) is an equation which can be estimated by instrumental variables methods, so long as it involves no unobservables. It is sometimes an advantage of Hansen's method that it does not involve a complete system of generating equations for the data. Often we are confident of the first order condition, but building a complete model would require much more work. But equally often we need a complete model to make forecasts or to compare fit with another approach. This paper points out that we can embed the first order conditions in a complete model. We need not choose between, say a linear VAR complete model, and a nonlinear equilibrium model which we can estimate by Hansen's methods but not use for prediction.

In our approach we generate a system of difference equations which will produce, for a given stable path of serially uncorrelated shocks w, a stable path for the variables C, S, and v. For empirical work, we will certainly want the dimension of w to match the sum of the dimensions of C and S, except to the extent that C and S are related by some identities in the model. This may require that some of our appended equations be given disturbance terms of their own, in addition to those appearing in the first order conditions. Once we have dimensions matching, the mapping from w_t vectors to C_t,S_t vectors is likely to be one-to-one, hence in principle invertible. That is, the same equations we use at each t to find C, S, and v from w can be used to find w from C and S. A likelihood function value can then be developed from the computed w path, assuming that the model asserts a distribution for the w path. For example, if w were taken to be a gaussian process the likelihood of the sample would depend on the sample first and second moments of w.

We can illustrate the suggested method with our simplest model above, from equations (4)-(8). In those equations there is only one independent source of error, u or v. The model hypothesizes that u is serially uncorrelated and, say, stationary i.i.d. normal. If both of the variables K and C are observed, then equations (6)-(8) generate an exact relation between them, and in real data the model will be rejectable as untrue because that relation will not hold without error. If C alone is observed, then (6) and (8) together define a first order AR in a nonlinear function of C, or, in other words, a way to form u as a nonlinear function of C. From this a likelihood function for the data could be formed as a function of the parameters of the utility function and of equation (8). If K alone were observed, a nonlinear relation between K and u would emerge from (7) and (8), again allowing formation of a likelihood function.

In the more interesting case where C and K are both observed, we could make the model usable by, say, adding a random disturbance w_t in the marginal cost of investment goods, making (5) read

$$40) \qquad C_t^2 + (1+w_t)K_t = \log K_{t-1} + v_t$$

and converting (7) to

$$41) \qquad M_t(1+w_t) = .95E_t[M_{t+1}]/K_t \ .$$

Then (6), (8) and (41) allow calculation of the path of u_t and w_t from data on C and K. Of course (40) still allows calculation of a v_t path as well. We have already specified, as part of writing down (8), stochastic behavior for u. If this two-disturbance model is to have content, we need a two-dimensional stochastic specification. We could do this by adding a univariate AR equation for w or v (not both), for example, asserting that the innovations in one of these variables are normal and independent of u. More likely to have a chance of fitting would be a model which dropped (8) and postulated a joint linear vector autoregression for w and M (if it postulated a joint VAR for v and M, it would make solving for innovations in v and M from the data a little harder by breaking up some of the system's recursiveness).

It perhaps bears mentioning here that, though once we can calculate the innovations which are being postulated to be i.i.d. normal we can calculate the likelihood, the likelihood is not ordinarily a function of the sum of squared calculated innovations alone in these models. The Jacobean of the transform between observed data and innovations is generally not the identity in a nonlinear model, and this must be taken into account in calculating likelihood.

This suggested procedure could be expensive. In a complicated model it may be that the mapping from C and S to w requires a time-consuming iterative procedure for solution at each t. When we recognize that to evaluate the likelihood for each set of model parameters we have to run through the whole sample this way, we can see the potential for high computer costs. (The costs here are still orders of magnitude smaller than what would be needed for a direct dynamic programming approach to a discretized version of the model, however.)

The procedure is also limited in that it would be very difficult to apply to time-aggregated data. With time-aggregated data, the "state" S is not directly observable. Therefore no simple recursive scheme for computing w_t from C_t and S_t will be available.

An alternative approach is to "estimate by simulation". One would choose a vector H of statistics which can be calculated from the data. Letting H_0 be the statistics calculated from the data and H(b) be the statistics calculated from a long series of simulated data from the model with parameters set at b, one would search over b to minimize $(H_0-H(b))'W(H_0-H(b))$, where W is an estimate of the inverse of the variance matrix of H_0 (assuming the simulation is so long that the variance of H(b) is negligible). The resulting estimates have an asymptotic covariance matrix given by $(DH\ W\ DH')^{-1}$, where DH is the derivative of H(b) with respect to b. If, e.g., H is the vector of coefficients in a linear VAR fit to the observed data and the VAR is chosen to leave negligible serial correlation in the data, W can be taken from the usual regression printout of coefficient covariance matrices. One could also estimate W directly from the simulated data, using a b at or near the value that minimizes the criterion.

Estimation by simulation can approach the more direct method in providing a complete model if the H vector represents a complete linear model. A model which systematically misses some important dynamic interaction will fail to generate a linear structure close to the form of an unrestricted VAR. Also, this method can easily handle time aggregation, censoring, seasonal adjustment, and other messy transformations of the data. Sometimes, the calculations for simulation from given shocks may be much simpler than the calculations of shocks from given data (though the reverse can occur, or be engineered into a model, as well).

The method does require repeated simulations over many time periods, which may be costly in large models. The nature of the computation, however, is such that with some programming effort it might be done very rapidly on vector processors (similar monte carlo problems have been converted to run efficiently on vector processors by physicists and engineers) and with minor programming effort it could take advantage of highly parallel computers.

6. Conclusion.

There is more to be said about and done with these methods. They will work in continuous time models, for example, and invertibility and stability issues show up in such models in a different way. This topic is omitted here because it would take up considerable space and require a more advanced mathematical apparatus. The methods have already been applied by Alfonso Novales [1984] and Songdal Shim [1984], who discuss some practical details glossed over here. However the methods make simulation of solutions so easy that they cry out to be applied to actual estimation of nonlinear models, with formal statistical inference rather than the informal comparison of simulations to data which appear in Novales's work and that of Kydland and Prescott. We have already sketched how to do this here, and a more detailed discussion probably ought to accompany an actual, serious application. Perhaps this paper's treatment goes far enough, though, so that many economists can take a hand in applying the methods and increasing our understanding of their strengths and limitations.

REFERENCES

Prescott, Edward and Finn Kydland (1982). "Time to Build and Aggregate Fluctuations," Econometrica, v.50, no.6, 1345-1370.

Novales, Alfonso (1984). "A Stochastic Monetary Equilibrium Model of the Interest Rate," Discussion paper, S.U.N.Y. at Stony Brook, September (based on 1983 University of Minnesota Ph.D. thesis).

Shim, Sonsdal (1984). Inflation and the Government Budget Constraint: International Evidence, Ph.D. Thesis in Economics, University of Minnesota.