8

Using large non-linear models

One of the main practical applications of econometric work is the construction and use of systems of equations in the form of econometric models. The earlier parts of this book are typical of econometric theory generally in that it concentrates largely on linear systems and equations. So suppose we have the following set of simultaneous equations.

$$AY = BX + u \tag{8.1}$$

where Y is a vector of N endogenous variables, X is a vector of M exogenous variable, u is a vector of N independently normally distributed error terms with zero mean, A is an $N \times N$ matrix of parameters, and B is an $N \times M$ matrix of parameters. This model can be solved analytically to give

$$Y = A^{-1}BX + \varepsilon \tag{8.2}$$

where

$$\varepsilon = (A^{-1})u$$

ε will be normally distributed with $E(\varepsilon) = 0$ and a covariance matrix we denote S.

The simulation properties of the model may easily be calculated as

$$\frac{\partial Y_i}{\partial x_j} = a_{ij}$$

where a_{ij} is the i, j element of $A^{-1}B$.

If the parameter matrices A, B are estimated then the uncertainty attached to a given simulation is given by the standard error of the

reduced form coefficient a_{ij} and this may be calculated directly from the covariance matrices of A and B.

For a linear system we can therefore derive analytical solutions for the model and its simulation properties, as well as fully defining the stochastic nature of the model solution. Unfortunately almost all models which are designed for practical use are non-linear and do not fall within the scope of the analysis outlined above. Even when the modeller restricts himself to using standard linear estimation techniques the use of data transformations in estimation inevitably produces a final model which cannot be put into a linear framework. Often linear estimation is carried out on logarithms of the variables for a number of valid reasons. For example, it may reduce heteroscedasticity, or a constant elasticity relationship may often be regarded as more theoretically reasonable than a linear formulation, etc. However it is not generally possible to specify the whole model in logarithmic form. For example, we cannot express linear identities in this form (e.g. the GDP identity). Thus there is no transformation of the whole model which allows it to be put in the form of equation (8.1).

Once the linear form is abandoned then we must also abandon the whole range of solution techniques described above. It is no longer possible to derive explicit general solutions to the model or explicit results about the simulation or stochastic properties of the model. Instead, a range of numerical techniques has grown up for the solution and analysis of non-linear models. In this Chapter we outline a number of these techniques. In section 8.1 we examine the solution methods for deterministic models; section 8.2 considers deterministic simulation methodology; and in section 8.3 we deal with problems posed by rational expectations. The consequences of the stochastic nature of models are addressed in section 8.4 and in section 8.5 we give a brief description of optimal control.

8.1 Model solution procedures

Most econometric models which are used either for forecasting or for simulation are both large and non-linear. It is therefore necessary to resort to a numerical procedure in order to determine the solution to the model. There are two main types of solution technique which are available, Newton and Gauss–Seidel (see Froeberg 1981 for a general mathematical exposition). Of these two approaches Gauss–Seidel has been almost universally adopted as the most practical for large econometric models and will be the only method discussed here.

In this section we will discuss the solution of what might be termed 'conventional' econometric models, that is, those which do not include explicit expectations of future endogenous variables. In section 8.3 we will discuss extensions of the standard solution techniques to allow for full model consistent expectations.

The Gauss–Seidel solution technique

We represent an n-equation model in the following notation where a linear form is adopted for convenience without loss of generality.

$$Y_i = A_i Y + B_i X \qquad i = 1, 2, \ldots, n \tag{8.3}$$

so that there are n endogenous variables (Y) and m exogenous or predetermined variables (X) and A_i, B_i are suitably dimensioned matrices. The Gauss–Seidel method proceeds by first assigning a starting value to the Y vector. In practice this is often the actual values of the Ys in the previous period. It then uses these values to solve the equations, *one at a time*. After each single equation is solved, that solution value is used to replace the initial guess for that variable. So if $\bar{Y}$ is the initial guess and Y^* is the new value, then for any equation:

$$Y_j^* = A_{Kj}Y^* + A_{Mj}\bar{Y} + B_j X \tag{8.4}$$

Where

$$A_{Kj} = \begin{cases} A_{Kj} & K < j \\ 0 & K > j \end{cases}$$

$$A_{Mj} = \begin{cases} 0 & M > j \\ A_{Mj} & M > j \end{cases}$$

That is to say, we work our way through the model equations sequentially solving each in turn. Any other endogenous variables in a specific equation take either their original starting value $\bar{Y}$ if they are higher in the ordering (and so have not yet been solved) or they take their new solution values Y^* if they are lower in the ordering (i.e. they have already been solved). This process of continual updating distinguishes the Gauss–Seidel technique from other schemes such as the Jacobi method where the whole model is solved before any updating takes place.

When all the equations have been solved, a check is made according to some convergence criteria on $|\bar{Y}_i - Y_i^*|$. If the two estimates of each Y are satisfactorily close a solution to the model has been found; if not, then the Y^* are redefined as $\bar{Y}$ and the process is

repeated for another iteration. A more complete exposition of the Gauss–Seidel method may be found in Faddeev and Faddeva (1963), and an early example of its application to econometric models is Norman (1967).

In the linear case described above we know that if a solution exists it is unique and that a solution will exist if all the equations are linearly independent. The Gauss–Seidel technique in practice is not guaranteed to find such a solution even when it exists and is unique. The crucial factors in the success of the Gauss–Seidel approach is the order in which the equations are solved (this is referred to as 'the ordering' of the model) and the normalisation of the equations (that is, which variables are chosen as being the dependent variable for a particular equation).

There are a number of variants on the Gauss–Seidel technique which have received attention. A good survey of the recent literature may be found in Hughes-Hallett (1981). If we restate (8.3) in a more compact form as

$$AY = B \tag{8.5}$$

then the iteration procedure may be characterised as

$$Y^{S+1} = GY^S + C \tag{8.6}$$

with some arbitrary starting value Y^0. The various iteration procedures may be nested within this framework by varying the construction of G and C. If we define $A = (P - Q)$ then $G = P^{-1}Q$ and $C = P^{-1}b$. The way the A matrix is split determines the exact form of the iteration procedure. The simplest procedure is the Jacobi iteration which defines

$$P = \begin{cases} A_{IJ} & \text{if } I = J \\ O & \text{if } I \neq J \end{cases} \tag{8.7}$$

The Gauss–Seidel iteration is produced by setting

$$P = (D - E)$$

$$D = \begin{cases} A_{ij} & \text{if } i = J \\ O & \text{if } i \neq J \end{cases} \quad \text{and} \quad E = \begin{cases} -A_{ij} & \text{if } i < J \\ O & \text{if } i > J \end{cases} \tag{8.8}$$

The successive over-relaxation iterative method is defined by

$$P = \frac{1}{\alpha} D(I - \alpha D^{-1}E) \tag{8.9}$$

where D and E are defined above.

A particularly important variant on these techniques allows for the incorporation of a damping factor γ in the following way:

$$Y^{(S+1)} = \gamma(GY^{(S)} + C) + (1 - \gamma)Y^S \quad \textbf{(8.10)}$$

When this is applied to the Gauss–Seidel iteration (8.8), the resulting technique is often called 'fast' Gauss–Seidel. The importance of this development is that while (8.6) can be shown to converge only if the spectral radius of $G < 1$ (see Young 1971), (8.10) can be shown to converge on the much weaker assumption that the real parts of the eigenvalues of G are all greater ($\gamma < 0$) or less ($\gamma > 0$) than one (see Hughes-Hallett 1981).

To make some of these ideas a little clearer, a simple two-dimensional example of the Gauss–Seidel technique is given in Figure 8.1. An initial value is assigned to x_2 of x_2^A; the first equation is then solved for x_1 using x_2^A; this yields x_1^A. This value is used to solve the second equation to yield x_2^B. The new value of x_2^B is then used to solve the first equation again and this finds x_1^B. The solution procedure then converges in the direction of the arrows towards the solution. If the equations had been normalised arbitrarily in the reverse way so that equation 2 had been solved first for x_1, the algorithm would have moved away from the solution and would have diverged indefinitely.

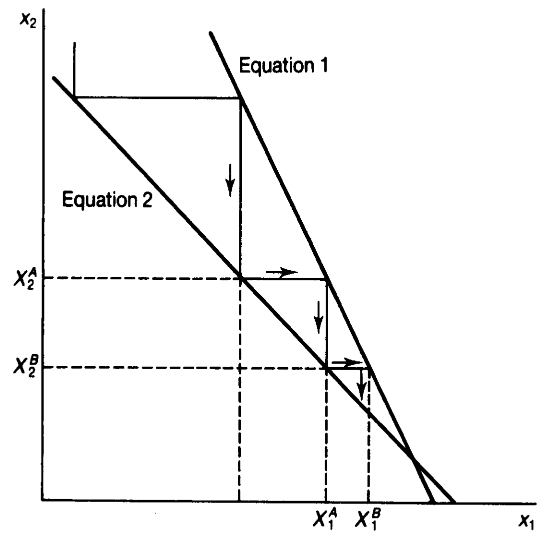

Figure 8.1 The Gauss–Seidel solution procedure.

8.2 Types of deterministic model solution

Users of large models have developed a range of different types of model solutions which are useful both in analysing a model and in using models in an applied framework. These techniques have not been discussed widely in the literature and so we will provide a summary of the more basic procedures here.

Suppose we write an N equation model in the following notation (again linearity is used for simplicity without loss of generality).

$$Y = B(L)Y + \gamma(L)Y + CZ \quad \textbf{(8.11)}$$

where Y is the vector of N endogenous variables, $B(L)$ is a matrix lag polynomial ($L = 0, 1, 2, \ldots$) where all the leading diagonals are zero, $\gamma(L)$ is a matrix lag polynomial ($L = 1, 2, \ldots$) where all off diagonal elements are zero and C is an $N \times M$ vector of coefficients on the M exogenous variable Z. In the notation $\gamma(L)Y$ will contain all the lagged dependent variables in each equation and $B(L)Y$ will contain all the lagged and contemporaneous endogenous (but not dependent) variables. This split simply isolates the lagged dependent variables in each equation in the γ matrix.

We may then define a dynamic solution to this model as Y^1 where

$$Y^1 = B(L)Y^1 + \gamma(L)Y^1 + CZ \quad \textbf{(8.12)}$$

That is, all lagged and contemporaneous endogenous variables take their solution values.

A very useful form of solution is defined as:

$$Y^2 = B(L)\bar{Y} + \gamma(L)\bar{Y} + CZ \quad \textbf{(8.13)}$$

where $\bar{Y}$ is a given value of Y, usually an historical realisation. A model solution such as (8.13) is often referred to as a single-equation residual solution, all the inputs to each equation take some known value. Each equation is therefore treated in isolation and $Y - Y^2$ will define a vector of single equation residuals which are exactly analogous to the residuals produced during estimation of a linear model.

A single-equation residual solution is particularly useful in assessing the recent performance of a model equation and it is a useful guide in tracking down the sources of errors which occur during a dynamic solution. If we define the dynamic residuals produced by a solution such as (8.12) as $U^D = Y - Y^1$ and the single equation residual as $U^S = Y - Y^2$ then it may easily be shown that for a linear model the information in U^D and U^S is identical. Each presents the same information in a different form; it is however, often much easier to deal with one than the other.

An important use for the single equation residuals U^S is in constructing a model simulation. Suppose we wish to increase Z by Δ then the solution outlined in (8.12) will become

$$Y^3 = B(L)Y^3 + \gamma(L)Y^3 + C(Z + \Delta)$$

The simulation effect will therefore be

$$Y^3 - Y^1 = B(L)(Y^3 - Y^1) + \gamma(L)(Y^3 - Y^1) + C\Delta$$

and the effect of Δ may be expressed as

$$Y^3 - Y^1 = [I - B(L) - \gamma(L)]^{-1}C\Delta \tag{8.14}$$

In the linear case defined above the simulation effect is a constant value and does not depend on the initial condition Y^1. For a general non-linear model this will not be true and so it is often desirable to calculate a simulation from a base which is particularly relevant, perhaps the actual historical data. In order to do this, Y^1, the model solution, must exactly reproduce the base data Y, which can be done by adding the *single-equation* residuals (U^S) to the dynamic model solution. This can be seen as:

$$Y = B(L)Y^* + \gamma(L)Y^* + CZ + U^S$$

$$Y = B(L)Y^* + \gamma(L)Y^* + CZ + Y - Y^2 \tag{8.15}$$

and by the substituting for Y^2 from (8.13) and rearranging, we get

$$(Y^* - Y) = B(L)(Y^* - Y) + \gamma(L)(Y^* - Y)$$

The solution to the equation is obviously $Y^* = Y$ and so the dynamic solution to the model with U^S added is Y (the base values used to calculate the single-equation residual). This result holds for both the linear and the non-linear case, the only problem being that in the non-linear case the solution to the model may not be unique.

A final solution method which is sometimes used as a model diagnostic is the *single-equation dynamic* solution, which may be characterised by

$$Y^4 = B(L)\bar{Y} + \gamma(L)Y^4 + CZ \tag{8.16}$$

where $\gamma(L)$ is a matrix lag polynomial with zero off diagonal terms and $B(L)$ is a matrix lag polynomial with zeros on the diagonal. This is exactly analogous to the dynamic forecast produced by a single equation. Each equation is treated in isolation and equation solution values are entered as lagged values only in the 'own' equation (but not in other equations). This may be a useful way of isolating dynamic instability in a model or detecting the source of a particularly bad dynamic performance.

8.3 Rational expectations and non-linear models

In this section we examine some of the special problems which arise when a model includes explicit expectations terms and when we wish to solve the model on the basis of model consistent or 'rational' expectations. The introduction of such terms raises both conceptual and practical problems. We will first discuss some of the conceptual problems of using expectations in non-linear models and will then examine some of the practical suggestions for dealing with the solution problem and the need for terminal conditions.

Before embarking on the details of model simulation and solution, there is an important conceptual problem which must be considered. The theoretical literature about rational expectations has evolved almost solely within a framework of small linear models. Within this framework it is accepted as axiomatic that a rational individual is interested in forming an estimate of the expected values of *all* relevant variables. That is to say, he will try to arrive at an estimate of the conditional mean of the probability distribution. Now, as the *deterministic* forecast of a linear model with normally distributed error processes coincides with the conditional mean of the probability distributor of the model, there is no conflict and the deterministic model solution may be used. Unfortunately (as discussed below) this is not the case for a non-linear model. The deterministic forecast of a stochastic non-linear model is not the mean of the probability distribution of the model. If the model represents a non-linear mapping from the error terms to the endogenous variables then the deterministic forecast may have no well-defined place on the probability distribution. This train of reasoning leads us towards carrying out stochastic simulations so as to *estimate* the *mean forecast* of the model. There is, however, a further complication; the expected values of any non-linear *identities* in the model are not given by the expected values of their component parts. Thus the expected real exchange rate will not equal the expected nominal exchange rate deflated by some expected relative price.

Of course, it does not follow that a set of expectations has to be consistent. If individuals have a quadratic loss function in their forecast errors and they use a non-linear model and are fully rational, then they should act on the basis of a *mutually inconsistent* set of expectations. Indeed, as we are dealing with many individuals, it may well be reasonable to think of these individuals as being different groups which hold inconsistent expectations about a number of variables. An exporting firm may form expectations about the real exchange rate while individuals hold price expectations and agents in financial markets have an expectation of the nominal exchange rate. All these expectations can be optimal, based on the same model and information set, and yet be inconsistent with each other.

These problems are perhaps most easily presented by stating a general non-linear model in the following form. Let

$$Y_t = f(Y_i, Y_j^e, X_k, B, \Omega) \quad \textbf{(8.17)}$$

$$i = 0, 1, \ldots t, j = t + 1, \ldots, T, k = 0, 1, \ldots, T$$

where Y_t is a set of N endogenous variables, X is a set of M exogenous variables, B is the full parameter set of the model and Ω is the variance–covariance matrix of all stochastic terms in the model (both parameters and error terms). In traditional macromodels the terms Y_j^e, future expected endogenous variables, may be viewed as having been substituted out of the model by some explicit expectations generating submodel:

$$Y_j^e = g(Y_i, X_k, \gamma, \phi) \quad \textbf{(8.18)}$$

$$i = 0, 1, 2 \ldots t, j = t + 1, \ldots, T, k = 0, 1, \ldots, t$$

where γ are parameters and ϕ is a covariance matrix of stochastic terms. We may substitute (8.18) into (8.17) to eliminate the future terms in the endogenous variables, Y_j. The model may then be solved in the traditional way. However, this procedure fails to identify explicitly the expectations formation procedure (8.18) so there is a loss of estimation efficiency. Further, if due to some regime change there is a shift in either the functional form of (8.18) or in its parameters, then, in the reduced form, of (8.17) and (8.18), the parameters will alter as the parameters in (8.18) alter under the new regime. However, if we deal with (8.17) and (8.18) *separately*, any change in the expectations formation mechanism is isolated in (8.18) and the structure (8.17) will be invariant to this form of structural change. This is the Lucas (1976) critique discussed in Chapter 6, but here in a whole model context.

Perhaps the simplest form of solution to this problem would be to derive an explicit model for expectations formation (8.18) and then use a complete structural model in the form of the set of equations (8.17) and (8.18) taken together. Certainly if we had a good idea of how expectations are actually formed the ideal situation would consist of explicit models of (8.18). However, in the absence of such information practitioners often invoke the rational expectations hypothesis. Under this assumption it is assumed that the expectations will coincide with the expected value of the actual forecast of the model:

$$Y_h^e = f(Y_i, Y_j^e, X_k, \beta, \Omega) \quad \textbf{(8.19)}$$

$$h = i, \ldots, T, i = 0, \ldots, h, j = h + 1, \ldots, T,$$

$$k = 1, \ldots, T$$

In fact, most implementations on large models do not conform fully to (8.19) as the solution is carried out in a deterministic fashion so that Ω is ignored. It is well known that for a non-linear model the deterministic forecast will differ from the mean (or expected value) of the model's density function. So under the REH assumption the usual procedure is to define

$$Y_h^e = f(Y_0, Y_j^e, X_k, \beta) \quad \textbf{(8.20)}$$

$$h = 1, 2, \ldots, T, i = 0, 1, \ldots, h, j = h + 1, \ldots, T,$$

$$k = 1, 2, \ldots, T$$

We will call an *explicit* expectations mechanism such as (8.18) an 'expectations model' solution. The deterministic model solution such as (8.14) we refer to as a 'consistent solution' and a stochastic solution such as (8.19) a 'rational solution'.

Carrying out a specific explicit expectations model solution involves no special problems, as the standard model solution procedures outlined above are quite able to cope with these models. The problems raised by the consistent solution have been the subject of recent attention in the literature. Very little attention has been paid to the rational solution of non-linear models, although Hall and Henry (1988) are an exception to this. The rest of this section will concentrate on the work which has dealt with consistent solution techniques.

There are currently a number of techniques in use for solving models with consistent expectations; the first to be used widely was the Fair (1979), Anderson (1979) iterative technique. A more recent approach using optimal control is the Holly and Zarrop (1983) penalty function method. An approach from the engineering literature is the multiple shooting technique. Finally there is the iterative technique outlined in Hall (1985b). All these techniques address the same problem, although the relationship between them is not always clear.

We will discuss the problem of model solution within a linear framework. This is done so that matrix notation may be used; none of the conclusions to be drawn are dependent on the assumption of linearity.

We begin by stating a general linear deterministic simultaneous model as

$$\alpha(L)Y_t = \beta(L)X_t \quad \textbf{(8.21)}$$

where $\alpha(L)$ and $\beta(L)$ are matrix lag polynomials (which may include lead terms), Y is a vector of N endogenous variables and X is a vector of M exogenous variables. Now, if we want to solve this model over a fixed time period, $1 \ldots T$, subject to suitable initial and terminal conditions Z, we may restate the problem, in a more explicit framework, as

$$AY' = BX' + CZ' \tag{8.22}$$

where Y and X are stacked vectors over all the time periods $1 \ldots T$ and CZ' is the initial and terminal conditions, that is, any lags which need values before the start of the solution (period 0) or expectations beyond the end of the solution (period T). It is worth actually writing out in full the left-hand side of (8.16):

$$\begin{bmatrix} \alpha & \alpha(L^{-1}) & \alpha(L^{-2}) & & & \alpha(L^{-(T-1)}) \\ \alpha(L) & \alpha & \alpha(L^{-1}) & & & \cdot \\ \alpha(L^2) & \alpha(L) & \alpha & & & \cdot \\ \alpha(L^3) & \alpha(L^2) & \alpha(L) & \alpha & & \cdot \\ \alpha(L^4) & \alpha(L^3) & \alpha(L^2) & \alpha(L) & \alpha \ddots & \\ \vdots & & & & & \\ \alpha(L^{T-1}) & & & & & \alpha \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ \cdot \\ \cdot \\ \cdot \\ y_T \end{bmatrix} \tag{8.23}$$

If the full A matrix is actually lower triangular, having only zeros above the leading diagonal, then the model contains no consistent expectation terms and it may be solved in the usual way, one period at a time. When the upper triangle is not empty, one of the special approaches mentioned earlier must be employed.

The approach outlined in Hall (1985b) is simply to deal directly with the equation system set out in (8.22) and (8.23). So we may normalise the model by defining $A = D - E$ and then use any of the standard iterative techniques (Gauss–Seidel, Fast, Gauss–Seidel, etc.) to solve the model.

Both the Fair–Anderson and the penalty function techniques make use of a separate split in the A matrix before the normalisation procedure is made. Both techniques begin by defining $A = (P - U)$, where P is the principal diagonal and all the lower triangular elements of A and U are minus the upper triangular elements of A. We can then rewrite (8.22) as

$$PY' = UY' + BX' + CZ' \tag{8.24}$$

This isolates all the lead terms and they can then be treated separately. This is done by defining a new vector, Y^e, where the consistent

solution is defined by $Y^e = Y$. The model may then be written as

$$PY' = UY^{e\prime} + BX' + CZ' \tag{8.25}$$

The Fair–Anderson procedure begins by setting arbitrary values for Y^e, solving (8.25), as a model without consistent expectations, and then updating the estimate of Y^e with the solution values. This procedure iterates until $Y = Y^e$.

The penalty function method proceeds in a similar fashion to achieve consistency by viewing the variables Y^e as control variables and then minimising a function $\Omega = \Sigma(Y - Y^e)^2$ using standard optimal control algorithms. This function has a minimum when $Y = Y^e$ and consistency is achieved.

The advantage of both these techniques is that the actual model solution procedure is reduced to a period-by-period problem without any consistent expectation terms entering. The added cost of this is obviously the cost of the extra iteration procedure in the Fair–Anderson technique and the cost of the optimal control exercise in the case of the penalty function approach. In effect these are both very sensible procedures to adopt while the upper triangle of A is very sparse. As A becomes more dense the costs can rise enormously.

The relationship between (8.23) and the multiple shooting techniques is a little less obvious. Any of the above techniques would proceed by normalising the model on the principal diagonal and then proceeding from there. The multiple shooting technique however first normalises the model on any lead terms. In terms of (8.23) this is rather like moving any rows with non-zero upper triangular elements down the model until the non-zero elements are on the diagonal. The model is then normalised on this new leading diagonal, which leads to some variables being determined twice. The initial period variables are then chosen so as to make the terminal values of the endogenous variables conform with the terminal conditions.

A simple example makes this more clear. Suppose we have an equation

$$E_t = E_{t+1} + \alpha X_t \tag{8.26}$$

where $E_T = \Sigma$. We renormalise this equation to give

$$E_{t+1} = E_t - \alpha X_t \tag{8.27}$$

This equation can now be used to solve the whole path of E_t, given E_0 and X_t. We search over alternative values of E_0 so that the solution for E_T, the terminal value, is equal to the (pre-set) terminal condition (see below).

The advantage of the multiple shooting technique is that it emphasises the importance of model normalisation and suggests ways in which the normalisation can be improved. The disadvantage is that it is in only very special cases where renormalisation is actually possible. If a single equation can be renormalised as a single unit, as in the case of (8.26), then the approach is quite straightforward. However, most cases would involve renormalising whole blocks of the model and this would not generally be feasible. An employment equation which includes expected output cannot be renormalised as a single unit, for example.

Specifying terminal conditions for a model with consistent expectation

Before one can solve a model which involves future expectations to yield a consistent solution, a suitable set of terminal conditions must be supplied. There has for some time been confusion over the distinction between terminal conditions and transversality conditions, in fact the two are quite different. This may be appreciated best by the following example, suppose we wish to minimise the following intertemporal cost function, where X_t^* is the desired, or target, value:

$$C = \sum_{t=1}^{\infty} \frac{a}{2} (X_t - X_t^*)^2 + \frac{b}{2} (X_t - X_{t-1})^2 \qquad \textbf{(8.28)}$$

which implies the following Euler equation (i.e. set $\delta C/\delta X_t = 0$)

$$aX_t^* = (a + 2b)X_t - b(X_{t+1} + X_{t-1}) \qquad \textbf{(8.29)}$$

A suitable transversality condition for this problem is

$$\lim_{t\to\infty} (X_t - X_t^*) = 0 \qquad \textbf{(8.30)}$$

However a finite-time horizon problem does not require a transversality condition. Instead, the Euler equations take on a special form as they approach the terminal date. For example, consider the three-period problem:

$$C = \sum_{t=1}^{3} (a/2)(X_t - X_t^*)^2 + (b/2)(X_t - X_{t-1})^2 \qquad \textbf{(8.31)}$$

The three first-order conditions are

$$aX_1^* = (a + 2b)X_1 - b(X_0 - X_2) \qquad \textbf{(8.32a)}$$

$$aX_2^* = (a + 2b)X_2 - b(X_1 + X_3) \qquad \textbf{(8.32b)}$$

$$aX_3^* = (a + b)X_3 - bX_2 \qquad \textbf{(8.32c)}$$

The transversality condition (8.30) may be derived by letting 't' in (8.32) go to infinity. But in the finite horizon case no transversality condition is necessary; instead the problem is dealt with by special equations appearing towards the terminal period.

The proper analogy for a macromodel would seem to be that we may interpret terminal conditions as transversality conditions if we solve the model over an infinite horizon. This is obviously impractical. It is wrong, however, to view the finite solution to the model with a terminal condition as being a version of (8.32) unless we recognize that this implies that all planning horizons end at the terminal period, as in (8.31) above.

A better interpretation of the terminal condition is that they should force the model solution to be on the infinite time horizon solution path at period T. Let us define $\bar{Y}$ to be the solution path of the model solved over an infinite time horizon subject to a set of transversality conditions derived by driving the model's own equations to infinity. Then, if we solve the model over the finite period $1, 2, \ldots, T$ subject to $Y_T = \bar{Y}_T$, the finite solution Y_i, $i = 1, \ldots, T$ will be equal to the infinite solution path for the first T periods. So we may achieve part of the infinite horizon solution path without solving the model to infinity.

The obvious difficulty here is that we cannot know what $\bar{Y}$ is until after an infinite model solution has been achieved. However, bearing this interpretation of the terminal conditions in mind we are able to make a more precise interpretation of the various suggestions which have been made. In particular, the Minford and Mathews (1978) suggestion that equilibrium values should be used, is based on the idea that they are using a market clearing model which quickly moves towards its equilibrium. So after a few initial periods have been passed, the infinite time solution path should be the steady-state equilibrium. Similarly the Holly and Beenstock (1980) suggestion of projecting constant growth rates as a terminal condition may be seen as a suggestion that the infinite time solution path is characterised by steady growth rates. The Fair (1979) idea of testing the terminal condition by extending the solution period until no significant change occurs in the early part of the solution period may also be seen as a way of fixing the terminal conditions on the infinite time solution path.

8.4 The analysis of stochastic models

By their very nature models are stochastic simply because no description of the world can ever be so complete that the models fit the data perfectly. So the full specification of an econometric model must include a set of error terms on the behavioural equations. For a linear model, as long as the error terms are normally distributed with zero mean, the stochastic part of the model is largely redundant. Ignoring the error terms completely gives rise to a deterministic forecast which is identical to the mean forecast of the stochastic model and which is optimal on almost any criterion. However, as soon as the model becomes non-linear this is no longer the case. There is then no general analytical relationship between the deterministic solution and the solution to the full stochastic model. In this section we explore the consequences of the stochastic nature of large models and discuss some of the numerical techniques for analysing non-linear models.

Stochastic simulation is a numerical computer technique which allows us to investigate the uncertainty which is associated inevitably with any large econometric model. Because such models are generally non-linear and highly complex, an analytic investigation of the effects and importance of their stochastic nature is impossible. Stochastic simulations bypass the analytic problems by simply performing large numbers of model simulations; each simulation differs from the others because of the different set of 'shocks' administered to the model. These shocks may be added to the equations, the parameters, or even the exogenous variables; the shocks are random drawings from a particular distribution. Given this repeated experiment it is then possible to calculate a range of statistics such as the mean, the standard deviation and the higher moments of the solution of the model variables. As the number of simulations undertaken increases, these summary statistics should provide a good guide to the stochastic performance of the whole model.

For any behavioural equation of a macromodel there is always some degree of uncertainty about its general specification, the actual values of its parameters and the importance of any error term. Typically, when an econometric model is used either for forecasting or simulation the stochastic nature of the model will be ignored. All error terms will be set, at least initially, to zero and the parameter estimates will be taken as known with certainty. It is natural to ask what the standard error of the deterministic forecast is and stochastic simulation can provide this answer. However, a much more important problem lies in the meaning of the deterministic forecast itself. It is well known that if the model is non-linear then the mean of the forecast will differ from the deterministic solution value. It has recently been pointed out (Hall 1984, Wallis 1984) that for some types of non-linearity the deterministic forecast may be quite meaningless and highly misleading as to the model's true forecast. A simple example can demonstrate this:

Let

$$\begin{aligned} Y &= \alpha X + u \\ W &= \beta Y + v \\ Z &= Y.W \end{aligned} \tag{8.33}$$

where u, v are stochastic error processes, α, β are parameters and X, Y, W, Z are variables. The reduced form solution is

$$\begin{aligned} Z &= \beta\alpha^2 X^2 + 2\alpha\beta X u + X v + \beta u^2 + u v \\ Y &= \alpha X + u \\ W &= \alpha\beta X + \beta u + v \end{aligned} \tag{8.34}$$

The equations for Y and W are simple linear equations, so assuming $E(u) = E(v) = 0$, the expected value of Y and W will be equal to the deterministic model forecast. This is not true for Z however, as the term in $E(u^2)$ will be positive. So the deterministic forecast, which sets $u^2 = 0$ will be an extreme point on the probability distribution of the random variable u^2. Any error at all will make $u^2 > 0$ and so the deterministic forecast is a highly biased and misleading indication of the stochastic model forecast.

It will be shown below that there are three broad classes of model. First, there are linear models and the deterministic forecast of such models is equal to the mean of the stochastic linear model, and all endogenous variables are normally distributed around this point (assuming normal error processes). Second, there are non-linear models which represent bijective mapping from the error terms on to the endogenous variables. A bijective mapping is a unique one-to-one mapping in both the function and its inverse. (The quadratic term discussed above is not bijective as its inverse is not a true one-to-one function.) The deterministic forecast of such a model can be shown to be the median of a (generally) skewed probability distribution. In this case the median, the mode and the mean of the probability density functions of the model are different. Forecasting the median seems a reasonable option especially considering some undesirable properties of the mean and the mode, discussed below. Finally, the third category is a non-linear model which is also non-bijective; in this case the deterministic forecast has no well-defined place on the probability density functions of the model. It can even lie at some highly unrepresentative extreme point, as shown above.

The example given above shows that a fairly simple form of non-linearity, which certainly exists in most large models, can give rise to non-bijective terms in the reduced form. So unless considerable work is undertaken to define and investigate the shape of the probability function of such models we have great difficulty interpreting any deterministic model results.

Stochastic simulations are useful therefore in defining and quantifying the uncertainty associated with a model forecast or simulation. But far more importantly they allow us to have a firm basis for interpreting the results of a deterministic model solution. If we know that the deterministic forecast is close to the mean value and that the probability distribution is near to being normal, then the model may be used in deterministic solution mode with some confidence. Until we have that information a serious problem of interpretation exists.

Interpreting the deterministic solution

When we are faced with the problem of having to choose a single point forecast from a skewed probability distribution there is no single point on the distribution which should be chosen in all circumstances. Instead, the optimal predictor will depend on the specific loss function of the forecaster (see Dunham Jackson 1921). For example, with a quadratic loss function and if a_i $(i = 1, \dots N)$ is a set of real numbers:

$$S_1 = \sum_{i=1}^{N} (x - a_i)^2 \quad \textbf{(8.35)}$$

then S_1 may be minimised with respect to x by setting x equal to the arithmetic mean of the a_i. In a forecasting context, if x is a point forecast and the a_i are all possible outcomes, then the optimal forecast is the mean of the probability distribution of the a_i.

The quadratic loss function is perhaps the most immediately appealing choice but it is by no means the only one. A clear alternative is to minimise the absolute error of the forecast:

$$S_2 = \sum_{i=1}^{N} |(x - a_i)| \quad \textbf{(8.36)}$$

S_2 will take a minimum value when x is equal to the median of the distribution of a_i.

Both of the above loss functions consider the whole set of possible errors. A more restrictive loss function might be to maximise the probability of picking the correct value:

$$S_3 = -|\mathrm{Max\,PR}(x - a_i) = 0| \quad \textbf{(8.37)}$$

This function will be minimised when x is set equal to the mode of the a_i.

Clearly, in the case of a normal distribution all three loss functions will deliver the same point estimate. The final function (S_3) is in

general unappealing as it gives no weight to the shape of the density function and in a highly perverse case could lead to extreme forecasts on the boundary of the density function. When considering the other two functions it may be argued that it is desirable to penalise large errors with a proportionately greater weight than small errors, so at first sight we may prefer the quadratic function.

There is, however, only highly undesirable property of the mean which makes it difficult to accept as a coherent forecast. This is that the mean forecast of the model is likely to violate any non-linear identities in the model. We can see quite easily that linear identities will hold in the mean forecast as

$$E(\Sigma x_i) = \Sigma E(x_i) \quad \textbf{(8.38)}$$

But we know that

$$E(XY) = E(X).E(Y) + \mathrm{Cov}(XY) \quad \textbf{(8.39)}$$

So any relationships which involve deriving a variable from the product of two other endogenous variables which are not independent of each other will not hold in expected values. This is not a trivial problem as most large macro models have many such identities, in particular the nominal value of a variable is often derived as the product of the real quantity of the variable and its price (for example, real disposable income, real wages or the real exchange rate). In general we would not expect the price (P) of a good to be independent of the quantity (Q) traded. The covariance of the two must therefore be non-zero and the mean value of revenue ($R = PQ$) will not equal the mean quantity multiplied by the mean price.

There are, of course, several alternatives which could be used to derive a coherent forecast based on the expected values of the model. One would be to derive the expected values of the behavioural equations, $E(X)$ and $E(Y)$, and then calculate any identities on the basis of these values, i.e. set arbitrarily $E(XY) = E(X)E(Y)$. There are two objections to this. First, if the identity feeds back into the model then the value calculated will not be the same as the value used in the model. Second, if we report the means because our loss function is quadratic, then to impose the identities is to behave sub-optimally. This point raises the second major objection to requiring coherency; it may be that rather than abandon the mean forecast we should actually abandon the coherency requirement. Part of the popular appeal of large models among forecasters is that they ensure that a large number of accounting identities are observed simultaneously. This may, however, be a mistake if the forecaster is simply interested in minimising his squared forecast error. However, if a forecasting group places some weight on the coherency of its forecast then it may well be that the use of mean forecasts is simply too simplistic.

The importance of the non-linearities present in large econometric models should not be underestimated. The interpretation of the stochastic nature of the endogenous variable is rendered particularly difficult by this problem. While we appear to have a good deal of information about the density function of the error terms of the model, the only information usually available on the endogenous variables is the deterministic forecast. Generally we have no way of even knowing where the deterministic solution lies on the density function.

Hall (1989) provides an analysis of this question and he establishes a proof for an important class of models that the deterministic solution is in fact the median of the distribution of the endogenous variables. When the distribution of the endogenous variables is skewed the normal measures of central tendency (the mean, the mode and the median) will of course differ and there is no strong reason to chose one measure over another. Indeed each can be justified as the optimal choice for a particular loss function. So establishing that the deterministic solution is the median of the distribution is an important justification for the use of non-stochastic model solutions, although clearly only a stochastic solution procedure can provide information on the overall shape of the distribution.

The numerical procedure of stochastic simulation

In this section we discuss a range of techniques which are known generally as stochastic simulation. Conceptually this is a very simple procedure. Suppose we have a non-linear model expressed in a general non-linear final form as

$$Y = Y(X, A, U) \qquad (8.40)$$

where Y is a vector of endogenous variables, X is a vector of exogenous variables, A is a set of parameters and U is a set of stochastic error terms. If both A and U are stochastic with mean zero and covariance matrix Σ_A and Σ_U then Y will also be stochastic with a vector of means Y' and a covariance matrix Σ_Y. Unfortunately the analytical calculation of Σ_Y and Y' is impossible for anything but the most simple form of non-linearities and even in these cases the size of a large model would often render the problem intractable.

The technique of stochastic simulation avoids the analytical calculation of Σ_Y and Y' by constructing a numerical approximation which is asymptotically equivalent to the true density function of Y. This approximation is carried out simply by repeatedly simulating (8.40) with values of A and U drawn from their distributions, defined above. The solution values for Y are then collected together and the moments of the distribution may be calculated. As the number of repetitions becomes very large the estimates of the moments of the distribution converge on Σ_Y and Y'. Further details and application of this approach may be found in Hall and Henry (1988), Fair (1984), Bianch and Calzolari (1982). We will now discuss a few of the details of the application of this technique.

Structural errors and additive errors

Despite the fact that most large models are non-linear they are generally estimated by single equation linear techniques, typically OLS. This is done by subjecting the variables to various transformations, for example, by taking the log of a variable, $\ln X$. When the equations are coded into the actual computer model the dependent variable is always transformed back into the 'pure' variable, X. This means that a random error added to the end of such an equation will not play the same role, or have the same properties as the estimated residual. An example will make this clear. If an equation of the form

$$\Delta \log(Y) = \alpha \Delta \log(X) + U$$

is estimated, then this will often be coded as

$$Y = \exp[\log Y_{t-1} + \alpha \Delta \log(X) + B] + A \qquad (8.41)$$

A is an additional 'residual' used for shocking the equation. B is the structural error term, normally set to zero, which will be a transformation of the estimation error U. Other forms of non-linearity are treated analogously. It is possible therefore to apply random shocks to either the A- or the B-residuals. The B or structural residuals depend on the estimation assumption of normality but there is no general reason to expect the B-residuals to be normally distributed, rather than the A-residuals.

Univariate and multivariate residual shocks

The distinction between structural (B) and additive (A) residuals has been made above but when we apply shocks to either of these sets of residuals we must also decide whether these shocks are to be univariate or multivariate ones. Univariate shocks are simply random, normally distributed shocks which have a given variance but are completely independent of each other.

Multivariate shocks will also generally be distributed normally with a given variance but they will also have some covariance structure between the individual shocks. In its simplest form we may allow for the fact that the error terms of different equations have some non-zero contemporaneous covariances. As an extension we may allow also for the covariance of the error terms in different equations to be related over different time periods.

The main argument for considering the covariances of the error terms in a model which has been estimated by OLS on the assumption of zero covariance in the equation error terms is that often the estimation assumptions are not actually fulfilled. An equation may be subject either to simultaneous equation bias or to omitted variable bias, or both, and the covariance structure of the error terms across equations may contain a great deal of information on this misspecification. For example, if current income were incorrectly omitted from the consumption function, then the covariance of the error term in the consumption equation and the other income-generating equations should pick up this omission.

There are currently three main techniques used to generate additive residual shocks which follow the covariance structure of the error terms of the whole model. Only one of these techniques can be used for large models however. The simplest technique is the Mariano and Brown (1981) approach; they use observed residuals from an N period model solution to carry out N static, one-period replications. This limits the number of replications to thirty or forty at the most as well as allowing the calculation of only the one-quarter-ahead static error bounds. A more useful technique is Nagar (1969); this uses an estimate of the full covariance structure of the model to apply shocks to the residuals. The problem here is that the covariance matrix must be estimated from observed residuals so that there must be more data points available than equation residuals. This will not generally be the case for a large model and so the initial covariance matrix cannot be defined. The final, and more useful, technique is the McCarthy algorithm (1972). This approach generates a vector of shocks by using the formula:

$$S = T^{0.5} rU$$

where S is the vector of random shocks, r is a $1 \times T$ vector of random numbers which are distributed $N(0,1)$ and U is a $T \times M$ matrix of disturbances from T observations of M true structural equations.

This technique therefore only requires a set of equation errors over T periods; T may be any length although the properties of S only tend to those of the true structural errors as T tends to infinity. Therefore this gives an asymptotic estimate of the true covariance matrix. The McCarthy technique has been extended to take account also of serial correlation in the error terms, although this extension will not be discussed here.

Handling parameter uncertainty

The variance of the forecast errors is made up from two sources, the variance of the true error term (U) and the parameter uncertainty, represented by the covariance matrix of the parameters. In stochastic simulation exercises it is relatively easy to take account of the variance of U but it is extremely difficult to make proper allowance for the variance of A in a satisfactory manner when the model is large. It is, of course, easy to shock the parameters by applying random shocks which are normal and have the parameters' estimated standard error. This procedure is, however, not satisfactory as it ignores the covariances between the parameters in any given equation as well as the covariances of the parameters across different equations. When these covariances are ignored there is a significant possibility that all the shocks in a given equation may be applied in the same direction, causing the dependent variable to change by an enormous amount, even changing sign. This need happen to only one equation in any run for the model to fail. Making allowance for the parameter covariance is therefore vital as this will mean that, on average, if one parameter falls then another will move in a compensating fashion so that the level of the dependent variable is maintained within 'sensible' bounds.

Three main techniques are used to deal with the problem of stochastic parameters, none of them being entirely satisfactory. These techniques are:

1. *Stochastic simulation and re-estimation (see Schink 1971)* Random shocks are added to the error term of the model so as to generate new values for the endogenous variables. These new values are then used to re-estimate the entire model and carry out a forecast run. The process is repeated many times so that the forecast errors can be

calculated. This technique is almost completely satisfactory in the sense that it generates sets of parameter values which take full account of all the covariances between the parameters themselves and between the parameters and the error terms. The disadvantage is, of course, that it is almost infeasible to consider 500 or 1000 replications of this technique for a large model.

2. *Monte Carlo on coefficients (see Cooper and Fisher 1974)* Shocks are applied to the parameters as well as to the random errors of each equation. The disadvantage here is that in the case of a large model where system estimation techniques are impractical, it is very hard, if not impossible, to carry out the necessary decomposition of the parameter covariance matrix. The normal technique used here when dealing with a large model is simply to ignore the cross-equation covariances and deal only with variance of the parameters. This clearly represents an important loss of information.

3. *Analytical simulation of coefficients (see Bianchi and Calzolari 1980)* An analytical formula is involved for the parameter uncertainty term which concerns the partial derivative of the parameters with respect to the endogenous variables. These partial derivatives are evaluated by using finite difference which involves many model simulations. The analytical formula also involves using an estimate of the variance–covariance matrix of the parameters.

It seems that the only feasible method in the case of a large model is to use procedure 2 and follow the assumption of Cooper and Fisher (1974), Fair (1980), Haitovsky and Wallace (1972) and assume the cross-equation covariances are all zero.

Variance reduction techniques

The main procedure used to reduce the uncertainty of the estimate of the mean of the distribution is the technique of antithetic errors. This means that the sets of residual errors to be applied in each simulation are not completely independent of the other sets, but instead are generated in pairs, where the second set of each pair is minus the first set. This produces a group of errors which are perfectly symmetric around the mean of the error process. A substantial increase is given in the efficiency of the estimate of the mean of the endogenous variables but it does not increase the efficiency of the estimate of the variance.

Estimating the uncertainty of a model's simulation properties

From the point of view of economic policy formation, the main interest in any macro model is its simulation properties. It is these properties which determine the policy prescriptions which are given by the model, no matter whether a simple set of policy alternatives is examined or if a complex analysis involving optimal control is used. When evaluating a large model an important aspect of its properties, which is often ignored, is the density function of the simulation effects. To say that the deterministic effect of a rise in government expenditure is to raise GDP is of little use until we are able to say what the margin of error surrounding this estimate is.

The original work in this area was undertaken by Fair (1980) and the approach is summarised in Fair (1984).

An analytical framework

Let Y_{it} be the set of i endogenous variables in a general non-linear model, X_{nt} be a set of n exogenous variables, Ω represents the variance–covariance matrix of all stochastic elements in the model (error terms and parameters) and B is a vector of parameter estimates. It is then possible to state the model in reduced form as

$$Y_{it} = Y_{it}(\Omega, B, X) \qquad \textbf{(8.42)}$$

The deterministic model solution would be given by ignoring the stochastic parts of the model as:

$$Y_{it}^{D} = Y_{it}(B, X) \qquad \textbf{(8.43)}$$

Conventional stochastic simulation techniques allow us to estimate the expected value of the endogenous variables conditional on an estimate of the variance–covariance matrix.

$$Y_{it}^{e} = Y_{it}(\hat{\Omega}, B, X) \qquad \textbf{(8.44)}$$

A model simulation exercise consists of solving the model for some base set of exogenous values (X^{I}) and then comparing this with another solution carried out on the basis of a different set of exogenous variables (X^{II}). So the effect of the deterministic simulation will be

$$d_{it}^{D} = Y_{it}(B, X^{II}) - Y_{it}(B, X^{I}) \qquad \textbf{(8.45)}$$

and similarly the difference in expected values of the stochastic simulation will be

$$d^e_{it} = Y^e_{it}(\Omega, B, X^{\text{II}}) - Y^e_{it}(\Omega, B, X^{\text{I}}) \quad \textbf{(8.46)}$$

In order to assess the uncertainty of a model's simulation properties we need to investigate the probability density function of d_{it}. As with conventional stochastic simulations, if the model is non-linear we will generally expect d^D_{it} to differ from d^e_{it}. Also it is clear that when we are dealing with non-linear models the variance of d_{it} will depend on both the stochastic parameters and the stochastic error terms. It is only in the case of a linear model that the variance of d_{it} is due to only the parameter uncertainty. This point can be appreciated easily by referring back to the simple model of (8.33). The reduced form equation for W, a linear part of the model, is

$$W = \alpha\beta X + \beta U + V \quad \textbf{(8.47)}$$

A simulation on X would give

$$d^W = \alpha\beta(X^{\text{II}} - X^{\text{I}}) \quad \textbf{(8.48)}$$

the error terms U, V drop out, and the density function of d^W is due solely to the stochastic nature of α and β. However, the situation is different for Z, the reduced form equation here is

$$Z = \beta\alpha^2 X^2 + 2\alpha\beta XU + \alpha XV + \beta U^2 + UV \quad \textbf{(8.49)}$$

So

$$d^Z = \beta\alpha^2((X^{\text{II}})^2 - (X^{\text{I}})^2) + 2\alpha\beta U(X^{\text{II}} - X^{\text{I}}) + \alpha V(X^{\text{II}} - X^{\text{I}}) \quad \textbf{(8.50)}$$

Here both the second and third term include the stochastic variables U and V, so the density function of d^Z depends in part on the density function of U and V.

Calculating the uncertainty of a model's simulation properties

Here we present the algorithm of Hall (1985) which efficiently provides estimates of the density function of a model's simulation properties.

1. Given the covariance matrices of the parameters and the error terms, draw a set of random parameters B^* and a set of residuals U^*.
2. Using the set of parameters and errors (B^*, U^*), solve the model for a base set of exogenous variables X^{I} to give $\hat{Y}^{\text{I}}$. The outcome of the model conditional on B^*, U^* and X^{I}.

3. Using the same set of parameters and errors (B^*, U^*), solve the model for a simulation set of exogenous variables X^{II} to give $\hat{Y}^{\text{II}}$, the outcome of the model conditional on B^*, U^* and X^{II}.
4. Compute $\hat{d}^J = \hat{Y}^{\text{II}} - \hat{Y}^{\text{I}}$
5. Repeat steps 1 to 4, J times, when J is the desired number of trials.
6. Given the J values of d, compute the mean and variance of d.

8.5 Optimal control of non-linear models

Wherever a model is used for policy analysis we are essentially trying to find the best setting for some group of instruments, given our understanding of the economy which is formalised in the model. The formal framework for any such analysis is clearly that of optimal control, even if in practice the analysis is conducted in a less formal way using only simulation methodology. The problem statement in its most general form is quite straightforward; let the model be

$$f_i(Y, X, A, \Omega) = 0 \quad \textbf{(8.51)}$$

where Y is a vector of endogenous variables, X are the exogenous variables, A the parameters and Ω is the full covariance matrix of the stochastic elements (from both the error terms and parameter estimates). The problem statement then simply involves specifying an objective function which is to be minimised.

$$\text{Min } E(J) = E[J(Y, X, \Omega)] \quad \textbf{(8.52)}$$

Note that we are minimising the expectation of some general function (J) of the stochastic model. We then minimise (8.52) subject to the model (8.51) with respect to a set of control variables C which are some subset of the exogenous variables X such that $(X) = (Z, C)$ where Z is all the exogenous variables not under the control of the policy maker.

For the case where the model is linear and the objective function is quadratic, a well-defined analytical solution exists which is detailed in a number of books, Intriligator (1971) or Hall and Henry (1988), and will not be discussed here. When the model is non-linear however, analytical solutions no longer exist and we must again resort to numerical procedures. The deterministic case is relatively easy to deal with, if we normalise (8.51) with fixed parameters such that

$$Y_i = h_i(Y, X) \quad \textbf{(8.53)}$$

and then state the reduced form of the system (assuming that this exists) as

$$Y_i = h_i'(X) \quad \textbf{(8.54)}$$

We may then state the problem (8.52) as the unconstrained minimisation problem

$$\text{Min} = J[h_i'(X), X]$$

or

$$\text{Min}\, J = g[h_i'(Z, C), Z, C] = g'(Z, C)$$

which is simply a problem of minimising a non-linear function with respect to a set of variables C; this problem was dealt with earlier when we considered maximum likelihood procedures. In practice because although many econometric models are large they are also fairly simple systems and so a number of particularly efficient algorithms have been developed (for example, Fair 1984, Holly *et al.* 1979 discuss such algorithms). Conceptually however we are simply maximising a non-linear function and any of the standard techniques could be used.

The problem takes on a different order of complexity when the non-linear model is stochastic. In this case there is no widely accepted procedure for calculating the optimal solution. This problem has been addressed by Chow (1976) from a theoretical standpoint and he outlines an algorithm which calculates optimal control rules for stochastic non-linear models. The Chow algorithm, in essence, works by iterating over a number of linearisations of the stochastic model using standard dynamic control theory to optimise the stochastic linearised model at each iteration. The key feature of the algorithm is that it is the stochastic model which is linearised not the deterministic model. To linearise a large stochastic model once would be enormously difficult and to include this as part of an iteration procedure would be an order of magnitude more complex, and as far as we know the Chow (1976) algorithm has never been implemented in its full form.

A few applications exists of stochastic optimal control of fairly small models, such as Bray (1975), but this work has generally proceeded by linearising the deterministic model rather than the full stochastic model. These applications then tend to produce solutions close to the deterministic solution (as we would expect) – indeed, if they were performed using fixed parameters and only error term uncertainty this algorithm converges on the deterministic solution.

Hall and Stephenson (1989) propose an algorithm which combines the technique of stochastic simulation with optimal control. It enables

one to calculate a very close approximation to the full stochastic optimal control solution. Their algorithm has the following form.

If the model has the general form of (8.51) then we may define Y^* to be the solution to (8.51) subject to the full stochastic processes of the model.

$$E[f_i(Y^*, X, \Omega, A, U)] = 0 \qquad i = I, N \quad \textbf{(8.55)}$$

and Y^* will be the mathematical expectation of Y, and U is a vector of error terms. Now define $\hat{Y}$ to be the deterministic model solution,

$$f = [\hat{Y}, X, 0, E(A), 0] = 0 \qquad i = 1, N \quad \textbf{(8.56)}$$

That is the variance–covariance matrix if the parameters are set to zero and the error terms take their mean value, which is assumed to be zero without loss of generality.

We know that, when the model is not linear:

$$\hat{Y} \neq Y^*$$

We may extend this framework to include optimal control by splitting the X vector into two sections Z, a vector of exogenous variables and C a vector of control variables. We then need only to specify a suitable objective function which is to be minimised.

We will examine the standard case of a conventional quadratic objective function:

$$E(J) = E\left[\sum_{i=1}^{n} A_i(Y_i - \bar{Y}_i)^2\right] \quad \textbf{(8.57)}$$

where $\bar{Y}_i$ is the desired value for variable Y_i and (8.57) is to be minimised subject to the model.

$$f(Y, Z, C, \Omega, A, U) = 0 \quad \textbf{(8.58)}$$

with respect to the control variables C. Again without loss of generality, we assume a one-period time horizon so as to simplify the notation, the multi-period extension is trivial.

Now we may rewrite (8.57) in the following way:

$$E(J) = \sum_{i=1}^{N} A_i E(Y_i^2 + \bar{Y}_i^2 - 2Y_i\bar{Y}_i) \quad \textbf{(8.59)}$$

$$= \sum_{i=1}^{N} A_i[E(Y_i^2) + \bar{Y}_i^2 - 2\bar{Y}E(Y_i)] \quad \textbf{(8.60)}$$

and given that $E(Y_i)^2 = E(Y_i)E(Y_i) + \text{Var}(Y_i)$

$$E(J) = \sum_{i=1}^{N} A_i[E(Y_i)E(Y_i) + \text{Var}(Y_i) + \bar{Y}_i^2 - 2\bar{Y}_iE(Y_i)] \quad \textbf{(8.61)}$$

and we may define $E(Y_i) = \hat{Y}_i + E(d_i)$, the expected value of Y_i equals the deterministic model solution $\hat{Y}_i$ plus the expected deviation of the deterministic value from the mean value $E(d_i)$. Then, substituting this into (8.61) gives:

$$E(J) = \sum_{i=1}^{N} A_i[\hat{Y}_i\hat{Y}_i + E(d_i)E(d_i) + 2\hat{Y}_iE(d_i) + \mathrm{Var}(Y_i) + \bar{Y}_i^2 - 2\bar{Y}_i\hat{Y}_i - 2\bar{Y}_iE(d_i)] \quad \textbf{(8.62)}$$

The advantage of (8.62) over (8.57) is that the stochastic elements of the solution have been isolated in the terms $\mathrm{Var}(Y_i)$ and $E(d_i)$ and we are able to provide numerical estimates for both of these terms through the use of stochastic simulation. This suggests the possibility of an algorithm to solve the stochastic problem which has the following step-by-step form:

1. Calculate the optimal solution to the deterministic problem given by (8.57) subject to (8.51), let the solution be C^*.
2. Perform a set of stochastic simulation around the base given by C^* to produce estimates of $\mathrm{Var}(Y_i)$ and d_i $(i = 1, 2, \ldots, N)$.
3. Using these estimates of d_i and $\mathrm{Var}(Y_i)$ we can now minimise (8.62) subject to (8.51) to produce a new optimal solution C'. If C' is within a convergence criteria of $C^*(|C' - C^*| < \text{EPS})$ for EPS suitably small then stop; if the convergence criteria is not met then set $C^* = C'$ and return to step 2.

This algorithm, at convergence will, still entail a small degree of approximation although this will be much less than the usual method of producing a linear approximation to the non-linear model. The conventional procedure of linearising the deterministic model, discussed in Kendrick (1981) would involve producing a linear approximation to the model and then appealing to the certainty equivalence theorem to solve the resulting deterministic quadratic–linear model. The problem with this approach is that when the objective function is quadratic and the parameters are known this procedure simply reproduces the deterministic solution.

We can see the source of the above approximation by noting that in general $\mathrm{Var}(Y_i)$ and d_i are both functions of the control variables C. We may simplify the notation by considering an example with only one control variable (C) and one state variable Y. Then, following the notation in (8.62) we may define

$$\hat{Y} = f(C) \quad \textbf{(8.63)}$$

$$\mathrm{Var}(Y) = g(C) \quad \textbf{(8.64)}$$

$$E(d) = h(C) \quad \textbf{(8.65)}$$

These terms may then be substituted into (8.62) to give

$$E(J) = f(C)f(C) + h(C)h(C) + 2f(C)h(C) + g(C) + \bar{Y}^2 - 2\bar{Y}f(C) - 2\bar{Y}h(C) \quad \textbf{(8.66)}$$

This is now an unconstrained function in C which will be minimised when the following FOC is met:

$$2f(C)f' + 2h(C)h' + 2f(C)h' + 2h(C)f' + g' - 2Yf' - 2Yh' = 0 \quad \textbf{(8.67)}$$

In the algorithm given above during the calculation of the optimal solution the partial derivatives g' and h' are set to zero so the solution which is calculated will be characterised by

$$2f(C)f' + 2h(C)f' - 2\bar{Y}f' = 0$$

The standard technique of linearising the model would also set $h(C) = 0$ and so this term would also be lost in the approximation. It must be appreciated at this point that $h(C)$, the deviation between the deterministic value of Y and its expected value, is of a quite different order of magnitude to g' and h', the derivatives of the deviation and the variance with respect to C. For most model applications g' and h' are likely to be so small that ignoring them is a reasonable approximation to make. However, if it is felt that a particular model is so non-linear that this is a damaging assumption then it is possible to reduce this level of approximation by estimating a simple linear approximation of $g(C)$ and $h(C)$. Two sets of stochastic simulation could be performed for different levels of C and a simple linear function for $g(C)$ and $h(C)$ could be calculated. Under normal circumstances however the main effect of the stochastic parts of the model will be captured by the term $h(C)$.

Finally, it is perhaps worth noting that the well-known certainty equivalence theorem can be demonstrated via equations (8.62) and (8.67). Certainty equivalence states that if the objective function is quadratic and the model is linear then the optimal and control trajectory for the stochastic problem is identical to the solution to the deterministic problem when all stochastic terms take their expected value. When the model is linear, $h(C) = g' = h' = 0$ and so (8.67) reduces to

$$2f(C)f' - 2\bar{Y}f' = 0 \quad \textbf{(8.68)}$$

which is identical to the FOC for the deterministic model.

8.6 Summary

This chapter has reviewed a range of techniques which allow large non-linear models to be analysed in much the same way that we are familiar with for small linear models. We have shown how model solutions can be obtained, how the stochastic properties of models can be investigated and how various forms of simulation and optimal control procedures may be defined. While these procedures may often be extremely complex from a numerical perspective, modern computers bring such techniques within the realms of feasibility even for very large models.

References

Aitchison, J. and Silvey, S. D. (1960), 'Maximum likelihood estimation procedures and associated tests of significance', *Journal of the Royal Statistical Society*, series B, 134–74.

Amemiya, T. (1981), 'Qualitative response models: a survey', *Journal of Economic Literature*, **19**, 1483–1536.

Anderson, B. D. O. and Moore, J. B. (1979), *Optimal Filtering*, Englewood Cliffs, N. J.: Prentice Hall.

Anderson, G. and Hendry, D. F. (1984), 'An econometric model of United Kingdom building societies', *Oxford Bulletin of Economics and Statistics*, **46** (3), 185–210.

Anderson, P. A. (1979), 'Rational expectations forecasts from non-rational models', *Journal of Monetary Economics*, **5**, 67–80.

Armijo, L. (1966), 'Minimization of functions having continuous partial derivatives', *Pacific Journal of Mathematics*, **16**, 1–3.

Athans, M. (1974), 'The importance of Kalman Filtering Methods for Economic Systems', *Annals of Economic and Social Measurement*, **3** (1), 49–63.

Baba, Y., Hendry, D. F. and Starr, R. M. (1985), 'A stable US money demand function, 1960–1984', Mimeo, Oxford: Nuffield College.

Baba, Y., Hendry, D.F. and Starr, R. M. (1988), 'US money demand 1960–1984', Oxford: Nuffield College, Discussion Paper no. 27.

Banerjee, A., Dolado, J. J., Hendry, D. F. and Smyth, G. W. (1986), 'Exploring equilibrium relationships in econometrics through state models: some Monte Carlo evidence', *Oxford Bulletin of Economics and Statistics*, **42** (3).

Barr, D. G. and Cuthbertson, K. (1990), 'Neoclassical consumer demand theory for the demand for money', *Economic Journal*, **101**, (407), July, 855–76.

Barr, D. G. and Cuthbertson, K. (1991), 'An interdependent error feedback model of UK company sector asset demands', *Oxford Economic Papers*, November.

Barro, R. J. (1978), 'Unanticipated money, output and the price level in the