Solution Methods for Nonlinear Models*

Richard G. Pierse

November 1997 (Revised June 1999)

1 Introduction

Economic models are used for many purposes, including forecasting and simulation. Economic models are generally nonlinear. Even simple macroeconomic models, where all the stochastic equations in the system may be linear in logarithms, need to be completed by accounting identities that are linear in levels. As a consequence, although a model may be estimated using linear techniques, forecasts with that model (which need to take the identities into account) will normally involve using nonlinear solution techniques.

Unfortunately, nonlinear models cannot in general be solved analytically and so solution techniques involve iterative numerical methods. When models include rational expectations, then solution techniques are made more complicated.

2 Linear Models

A linear model can be written in the standard *structural* form

$$\mathbf{B}_0\mathbf{y}_t = \mathbf{\Gamma}\mathbf{z}_t + \mathbf{u}_t \quad , \quad t = 1, \cdots, T \tag{2.1}$$

where $\mathbf{y}_t$ is an $n \times 1$ vector of *endogenous* variables in time period t, $\mathbf{z}_t$ is an $m \times 1$ vector of *predetermined* variables and $\mathbf{u}_t$ is a vector of structural errors on the equations with $E(\mathbf{u}_t) = 0$ and $\text{var}(u_t) = \mathbf{\Sigma}$. $\mathbf{B}_0$ and $\mathbf{\Gamma}$ are matrices of coefficients of dimensions $n \times n$ and $n \times m$ respectively.

If the model is dynamic then the predetermined variables $\mathbf{z}_t$ will comprise both current and lagged *exogenous* variables $\mathbf{x}_t$, not explained in the system

*This handout was originally prepared for a lecture given at the Bank of England. It should not be quoted without permission.

and *lags of the endogenous variables* $\mathbf{y}_t$**.** Separating out these two different types of variable we can rewrite (2.1) as

$$\mathbf{B}_0\mathbf{y}_t = -\mathbf{B}_1\mathbf{y}_{t-1} - \cdots - \mathbf{B}_p\mathbf{y}_{t-p} + \mathbf{C}_0\mathbf{x}_t + \cdots + \mathbf{C}_q\mathbf{x}_{t-q} + \mathbf{u}_t$$

or

$$\mathbf{B}(L)\mathbf{y}_t = \mathbf{C}(L)\mathbf{x}_t + \mathbf{u}_t \tag{2.2}$$

where L is the lag operator defined by $L^j x_t = x_{t-j}$ and $\mathbf{B}(L)$ and $\mathbf{C}(L)$ are matrices of polynomials in the lag operator of order p and q respectively defined by

$$\mathbf{B}(L) = \mathbf{B}_0 + \mathbf{B}_1 L + \mathbf{B}_2 L^2 + \cdots + \mathbf{B}_p L^p$$

and

$$\mathbf{C}(L) = \mathbf{C}_0 + \mathbf{C}_1 L + \mathbf{C}_2 L^2 + \cdots + \mathbf{C}_q L^q\,.$$

The conditions for dynamic stability of the model (2.2) are that the np roots of the determinantal polynomial $|\mathbf{B}(L)|$ all lie within the unit circle.

The *reduced form* of the system can be written as

$$\mathbf{y}_t = -\mathbf{B}_0^{-1}(\mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \mathbf{C}_0\mathbf{x}_t + \cdots + \mathbf{C}_q\mathbf{x}_{t-q}) + \mathbf{B}_0^{-1}\mathbf{u}_t\,. \tag{2.3}$$

This represents the solution of the model in terms of lagged endogenous variables and current and lagged exogenous variables.

Substituting out the lagged endogenous variables defines what is called the *final form* of the model:

$$\mathbf{y}_t = \mathbf{B}(L)^{-1}\mathbf{C}(L)\mathbf{x}_t + \mathbf{B}(L)^{-1}\mathbf{u}_t$$

which expresses the time path of $\mathbf{y}_t$ in terms of (generally infinite) lags of the exogenous variables and past errors.

The reduced form can be used for prediction. Two modes of solution are possible: *static* and *dynamic*, depending on how the lagged endogenous variables are treated. Static solution sets the lagged endogenous variables to their observed historical values. Dynamic solution uses predicted values for lagged endogenous variables as these become available. Dynamic solution can be used recursively to forecast future values of $\mathbf{y}_{t+j}$, $j > 0$ given initial values of $\mathbf{y}_t$ and future values of the exogenous variables $\mathbf{x}_{t+j}$.

Setting $\mathbf{u}_t$ to its expected value of $\mathbf{0}$, the static solution is given by

$$\widehat{\mathbf{y}}_t = -\mathbf{B}_0^{-1}(\mathbf{B}_1\mathbf{y}_{t-1} + \cdots + \mathbf{B}_p\mathbf{y}_{t-p} + \mathbf{C}_0\mathbf{x}_t + \cdots + \mathbf{C}_q\mathbf{x}_{t-q})$$

and the dynamic solution by

$$\widehat{\mathbf{y}}_t = -\mathbf{B}_0^{-1}(\mathbf{B}_1\widehat{\mathbf{y}}_{t-1} + \cdots + \mathbf{B}_p\widehat{\mathbf{y}}_{t-p} + \mathbf{C}_0\mathbf{x}_t + \cdots + \mathbf{C}_q\mathbf{x}_{t-q})$$

where $\widehat{\mathbf{y}}_t$ represents predicted value of $\mathbf{y}_t$. Note that its conditional expectation is given by

$$E(\widehat{\mathbf{y}}_t|\mathbf{y}_{t-1}, \cdots, \mathbf{y}_{t-p}, \mathbf{x}_t, \cdots, \mathbf{x}_{t-q}) = \mathbf{y}_t$$

so that $\widehat{\mathbf{y}}_t$ is an unbiased predictor.

3 Nonlinear Models

A general dynamic nonlinear model can be written in the implicit form

$$\mathbf{f}(\mathbf{y}_t, \mathbf{y}_{t-1}, \cdots, \mathbf{y}_{t-p}, \mathbf{x}_t, \cdots, \mathbf{x}_{t-q}, \mathbf{u}_t; \theta) = \mathbf{0} \tag{3.1}$$

where $\mathbf{f}$ is an $n \times 1$ vector valued function and θ is a vector of parameters. This represents a set of n nonlinear equations. For conciseness it is useful to define

$$\mathcal{Y}_{t-1} = \{\mathbf{y}_{t-1}, \cdots, \mathbf{y}_{t-p}\}$$

and

$$\mathcal{X}_t = \{\mathbf{x}_t, \cdots, \mathbf{x}_{t-q}\}$$

and rewrite (3.1) as

$$\mathbf{f}(\mathbf{y}_t, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) = \mathbf{0}\,. \tag{3.2}$$

4 Solution Methods for Nonlinear Models

Unfortunately, there is no analytic solution for the nonlinear model (3.2) corresponding to the reduced form (2.3) for linear models. Consequently, the reduced form solution

$$\mathbf{y}_t = \mathbf{g}(\mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) \quad , \quad t = 1, \cdots, T \tag{4.1}$$

has to be found using iterative numerical methods. Starting from an initial guess $\mathbf{y}_t^0$, each iteration s, $s = 1, \cdots, S$ generates a new vector of values $\mathbf{y}_t^s$ based on the previous iteration. Convergence is achieved when either

$$\max_i \left| y_{it}^s - y_{it}^{s-1} \right| < \varepsilon_1 \quad \text{or} \quad \max_i \left| \frac{y_{it}^s - y_{it}^{s-1}}{y_{it}^{s-1}} \right| < \varepsilon_2\,.$$

Two general solution methods exist: *Newton's method* which uses first derivatives of the model, and the first-order methods, the most famous of which is the *Gauss-Seidel method*.

4.1 Initial Conditions

The model (3.2) includes current and lagged values of the endogenous variables but, at each point in time t, the lagged values of the endogenous variables, $\mathcal{Y}_{t-1}$ are predetermined and can be treated as fixed. Solution can thus proceed sequentially for each period $t = 1, \cdots, T$. Note however, that in the first period, we need values for $\mathcal{Y}_0$ which are outside the model solution period. These are known as *initial conditions* and are generally set to actual historical values. The supplied values $\mathcal{Y}_0$ are then treated as exogenous in model solution.

4.2 Newton's method

Consider the Taylor expansion of (3.2) around the point $\mathbf{y}_t = \overline{\mathbf{y}}_t$:

$$\begin{aligned} \mathbf{f}(\mathbf{y}_t, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) &= \mathbf{f}(\overline{\mathbf{y}}_t, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) + \frac{\partial \mathbf{f}}{\partial \mathbf{y}_t'}(\mathbf{y}_t - \overline{\mathbf{y}}_t) \\ &+ \quad \text{higher order terms} \quad = \quad 0 \end{aligned} \tag{4.2}$$

where

$$\frac{\partial \mathbf{f}}{\partial \mathbf{y}_t'} = \begin{bmatrix} \frac{\partial f_1}{\partial y_{1t}} & \cdots & \frac{\partial f_1}{\partial y_{nt}} \\ \vdots & \ddots & \vdots \\ \frac{\partial f_n}{\partial y_{1t}} & \cdots & \frac{\partial f_n}{\partial y_{nt}} \end{bmatrix} = \mathbf{J} \tag{4.3}$$

is the $n \times n$ matrix of first order model derivatives evaluated at the point $\overline{\mathbf{y}}_t$. This matrix is known as the *Jacobian matrix*.

Neglecting the higher order terms, equation (4.2) can be rearranged to give

$$\mathbf{y}_t = \overline{\mathbf{y}}_t - \mathbf{J}^{-1}\mathbf{f}(\overline{\mathbf{y}}_t, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta)\,.$$

This equation will hold exactly in a linear model (where the higher order terms in the Taylor expansion are zero and where $\mathbf{J} = \mathbf{B}_0$) and will hold approximately in nonlinear models.

It forms the basis of Newton's iterative method, Newton (1686), where in iteration s,

$$\mathbf{y}_t^s = \mathbf{y}_t^{s-1} - \mathbf{J}_{s-1}^{-1}\mathbf{f}(\mathbf{y}_t^{s-1}, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta)\,.$$

Newton's method is said to be *quadratically convergent* in that it will find the solution to a linear model in a single iteration. Since all models will be approximately linear in the neighbourhood of a solution, this makes it a very powerful solution method. The method is sometimes also known as Newton-Raphson.

4.3 First Order Methods

First order solution methods require that the nonlinear system (3.2) be written in the explicit form

$$\mathbf{y}_t = \mathbf{h}(\mathbf{y}_t, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta)\,. \tag{4.4}$$

Each equation in (4.4) has been normalised on one of the endogenous variables. Such a normalisation will generally be possible but will not be unique. Convergence of the first order iterative methods will depend on the particular normalisation that has been adopted. This is a disadvantage of these methods compared with Newton's method. On the other hand, first order methods do not require the inversion of a matrix of model derivatives and are considerably cheaper to implement.

The *Jacobi solution method* is defined by the iteration

$$\mathbf{y}_t^s = \mathbf{h}(\mathbf{y}_t^{s-1}, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta)\,. \tag{4.5}$$

This method generates a new solution using the previous iteration's solution on the right-hand side of each equation. Iteration proceeds until a fixed point is found where $\mathbf{y}_t^s = \mathbf{y}_t^{s-1}$.

A generalisation of this method is the *Jacobi over-relaxation* or *JOR* method

$$\mathbf{y}_t^s = \alpha\mathbf{h}(\mathbf{y}_t^{s-1}, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) + (1-\alpha)\mathbf{y}_t^{s-1} \tag{4.6}$$

where the scalar α is the relaxation parameter. With $0 < \alpha < 1$, the new solution is a linear combination of the solution of this iteration and last iteration. The effect of this is to dampen the change between iterations and this can speed up convergence. The optimal value for α will depend on the model and can be found by experimentation.

The *Gauss-Seidel* method solves each of the n equations in the order i, $i = 1, \cdots, n$ using the iterations

$$y_{it}^s = h_i(y_{1t}^s, \cdots, y_{i-1t}^s, y_{i+1t}^{s-1}, \cdots, y_{nt}^{s-1}, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta)\,. \tag{4.7}$$

This method differs from the Jacobi method in that equation i uses the solution from *this iteration* for all variables defined by equations appearing *earlier* in the model: *i.e.* y_{jt}, $j < i$. For this reason the method is also known as the method of successive substitution. This variation on the Jacobi method has been found to speed up convergence considerably. On the other hand, it makes convergence dependent on the equation ordering. Don and Gallo (1987) suggest an algorithm for the optimal ordering of a system of equations for Gauss-Seidel model solution. This orders the model equations

to maximise the recursivity of the system which is equivalent to ordering the equations to make the model Jacobian matrix (4.3) as close as possible to being lower triangular.

Adding a relaxation parameter to the Gauss-Seidel method defines the *Successive Over Relaxation* or *SOR* method

$$y_{it}^{s} = \alpha h_i(y_{1t}^{s}, \cdots, y_{i-1t}^{s}, y_{i+1t}^{s-1}, \cdots, y_{nt}^{s-1}, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) + (1-\alpha)\, y_{it}^{s-1}\,. \quad (4.8)$$

Finally the *Fast Gauss-Seidel* or *FGS* method uses the iterations

$$\begin{aligned} y_{it}^{s*} &= \alpha h_i(y_{1t}^{s*}, \cdots, y_{i-1t}^{s*}, y_{i+1t}^{s-1}, \cdots, y_{nt}^{s-1}, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) + (1-\alpha)\, y_{it}^{s-1} \\ y_{it}^{s} &= \gamma y_{it}^{s*} + (1-\gamma) y_{it}^{s-1}\,. \end{aligned} \quad (4.9)$$

which takes a linear combination of the *SOR* iteration and last iteration's solution with weights γ and $1-\gamma$ respectively. The *FGS* method has been found to be faster than *SOR* for judicious choice of parameters α and γ. Note that if $\gamma = 1$ or $\alpha = 1$ then *FGS* collapses to *SOR*.

5 Prediction with Nonlinear Models

As with linear models, prediction can be based on the model reduced form

$$\mathbf{y}_t = \mathbf{g}(\mathbf{y}_{t-1}, \cdots, \mathbf{y}_{t-p}, \mathcal{X}_t, \mathbf{u}_t; \theta)$$

which is obtained from numerical solution of the model. As before, both *static* and *dynamic* prediction is possible. Setting $\mathbf{u}_t$ to its expected value of $\mathbf{0}$, the static solution is given by

$$\widehat{\mathbf{y}}_t = \mathbf{g}(\mathbf{y}_{t-1}, \cdots, \mathbf{y}_{t-p}, \mathcal{X}_t, \mathbf{0}; \theta)$$

and the dynamic solution by

$$\widehat{\mathbf{y}}_t = \mathbf{g}(\widehat{\mathbf{y}}_{t-1}, \cdots, \widehat{\mathbf{y}}_{t-p}, \mathcal{X}_t, \mathbf{0}; \theta)\,.$$

However, unlike the linear model case, these predictors are no longer unbiased since $E(a(x)) \neq a(E(x))$ when $a(\cdot)$ is a nonlinear function, so that in general

$$E(\mathbf{y}_t | \mathcal{Y}_{t-1}, \mathcal{X}_t) \neq \mathbf{g}(\mathcal{Y}_{t-1}, \mathcal{X}_t, E(\mathbf{u}_t); \theta)\,.$$

Unbiased prediction can be obtained by the method of *stochastic simulation*. This generates repeated solutions to the model

$$\widehat{\mathbf{y}}_t^{r} = \mathbf{g}(\widehat{\mathbf{y}}_{t-1}, \cdots, \widehat{\mathbf{y}}_{t-p}, \mathcal{X}_t, \mathbf{e}_t^{r}; \theta) \quad , \quad r = 1, \cdots, R$$

for R successive draws of a vector of pseudo-random disturbances $\mathbf{e}_t^r$, with distribution given by

$$\mathbf{e}_t^r \quad \sim \quad iid \quad N(\mathbf{0}, \widehat{\mathbf{\Sigma}}) \quad , \quad r = 1, \cdots, R$$

where $\widehat{\mathbf{\Sigma}}$ is a consistent estimate of $\text{var}(\mathbf{u}_t)$.

Then, for large enough R, the mean of the replications

$$\widetilde{\mathbf{y}}_t = \frac{1}{R} \sum_{r=1}^{R} \widehat{\mathbf{y}}_t^r$$

will approximate $E(\mathbf{y}_t | \mathcal{Y}_{t-1}, \mathcal{X}_t)$.

6 Models with Forward Expectations

We define expectations using the notation

$$x_{t+k|t-1}^e \equiv E(x_{t+k} | \mathfrak{I}_{t-1})$$

where $x_{t+k|t-1}^e$ is the expected value of variable x_{t+k} formed at the end of period $t-1$ and based on information available at that time, represented by the information set $\mathfrak{I}_{t-1}$. The rational expectations hypothesis of Muth (1961) asserts that expectations should satisfy

$$x_{t+k} = x_{t+k|t-1}^e + \eta_{t+k} \tag{6.1}$$

where $E(\eta_{t+k}) = 0$. Most solution methods set η_{t+1} to its expected value of zero and adopt the stronger assumption of *model consistent expectations* that

$$x_{t+k} = x_{t+k|t-1}^e \, . \tag{6.2}$$

Models including rational expectations present special solution problems.

6.1 Linear Models

Defining the forward expectations operator $F^j x_t = x_{t+j|t-1}^e$, a general linear model including forward expectations of the endogenous variables $\mathbf{y}_t$ can be written as

$$\mathbf{B}(L)\mathbf{y}_t + \mathbf{D}(F)\mathbf{y}_t = \mathbf{C}(L)\mathbf{x}_t + \mathbf{u}_t \tag{6.3}$$

where

$$\mathbf{D}(F) = \mathbf{D}_0 + \mathbf{D}_1 F + \mathbf{D}_2 F^2 + \cdots + \mathbf{D}_k F^k$$

is a matrix polynomial of order k with $\mathbf{D}_0 = \mathbf{0}$.

The assumption (6.2) that expectations are *model consistent* implies that $F = L^{-1}$ so that (6.3) can be rewritten as

$$(\mathbf{B}(L) + \mathbf{D}(L^{-1}))\mathbf{y}_t = \mathbf{\Phi}(L, L^{-1})\mathbf{y}_t = \mathbf{C}(L)\mathbf{x}_t + \mathbf{u}_t$$

where

$$\mathbf{\Phi}(L, L^{-1}) = \mathbf{D}_k L^{-k} + \cdots + \mathbf{D}_1 L^{-1} + \mathbf{B}_0 + \mathbf{B}_1 L + \mathbf{B}_2 L^2 + \cdots + \mathbf{B}_p L^p$$

is a polynomial of order $p + k$ with $n(p + k)$ roots. Assuming no unit roots, this polynomial can be factorised as

$$\mathbf{\Phi}(L, L^{-1}) = \mathbf{\Phi}_1(L)\mathbf{\Phi}_2(L^{-1})$$

where the roots of $|\mathbf{\Phi}_1(L)|$ all lie inside the unit circle and the roots of $|\mathbf{\Phi}_2(L^{-1})|$ lie outside the unit circle. The stability condition for the model (6.3) is that there are no more roots in $\mathbf{\Phi}_2(L^{-1})$ than there are expectational terms. If there are exactly the same number of roots outside the unit circle as there are expectations then the saddlepath condition holds and the stable solution is unique. See *inter alia* Sargent (1979) or Blanchard and Kahn (1980). If not, then an infinite number of stable solutions exist.

As long as the stability conditions holds, the inverse of $\mathbf{\Phi}(L, L^{-1})$ will exist and the model final form is given by

$$\mathbf{y}_t = \mathbf{\Phi}_2(L^{-1})^{-1}\mathbf{\Phi}_1(L)^{-1}\mathbf{C}(L)\mathbf{x}_t + \mathbf{\Phi}_2(L^{-1})^{-1}\mathbf{\Phi}_1(L)^{-1}\mathbf{u}_t\,. \tag{6.4}$$

This equation expresses the time path of $\mathbf{y}_t$ in terms of (generally infinite) lags and leads of the exogenous variables and errors.

6.2 Nonlinear models

The general form of a nonlinear dynamic model with forward expectations is given by

$$\mathbf{f}(\mathbf{y}_t, \mathbf{y}^e_{t+1|t-1}, \cdots, \mathbf{y}^e_{t+k|t-1}, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) = \mathbf{0}\,. \tag{6.5}$$

On the assumption of model consistent expectations, this becomes

$$\mathbf{f}(\mathbf{y}_t, \mathbf{y}_{t+1}, \cdots, \mathbf{y}_{t+k}, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) = \mathbf{0}$$

or

$$\mathbf{f}(\mathbf{y}_t, \mathcal{E}_t, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) = \mathbf{0} \tag{6.6}$$

where

$$\mathcal{E}_t = \{\mathbf{y}_{t+1}, \cdots, \mathbf{y}_{t+k}\}\,.$$

Clearly, this system of equations can no longer be solved period by period, using the nonlinear techniques considered so far, except in the special case where there are no lags in the system so that $\mathcal{Y}_{t-1}$ can be dropped and it is possible to use the standard techniques to solve the model backwards from $t = T, T-1, \cdots, 1$.

6.2.1 Terminal conditions

Note that a model solution over a finite time horizon $t = 1, \cdots, T$ requires values of the variables $\mathbf{y}_{T+1}, \cdots, \mathbf{y}_{T+k}$ which are outside the solution period. These are called *terminal conditions* and are analogous to the initial conditions required for any dynamic model. Generally, terminal values are set to fixed exogenous values, or according to an automatic rule such as constant level $\mathbf{y}_{T+j} = \mathbf{y}_T$, $j = 1, \cdots, k$ or constant growth rate $\mathbf{y}_{T+j} = \mathbf{y}_T^{j+1}\mathbf{y}_{T-1}^{-j}$, $j = 1, \cdots, k$. However, it is also possible to choose terminal conditions to reflect the equilibrium properties of the model as suggested by Minford *et al.* (1979). This uses terminal conditions in a way analogous to the transversality conditions that occur in infinite time horizon problems, and implies that the model will have reached an equilibrium solution by the terminal date. Fisher (1992) shows how terminal conditions may help pick one solution among an infinite number of solutions.

To some extent, terminal conditions are arbitrary and it is important to know how sensitive the model solution is to them. One way to investigate this is to vary the terminal date by extending the solution period beyond T. As the solution period is extended, the terminal condition will have less influence on the first T periods.

6.2.2 Stacked solution methods

Consider stacking the systems of equations (6.6) for each time period $t = 1, \cdots, T$ to form the system of equations

$$\begin{aligned}
&\mathbf{f}(\mathbf{y}_1, \mathcal{E}_1, \mathcal{Y}_0, \mathcal{X}_1, \mathbf{u}_1; \theta) = \mathbf{0} \\
&\mathbf{f}(\mathbf{y}_2, \mathcal{E}_2, \mathcal{Y}_1, \mathcal{X}_2, \mathbf{u}_2; \theta) = \mathbf{0} \\
&\qquad\qquad \vdots \\
&\mathbf{f}(\mathbf{y}_T, \mathcal{E}_T, \mathcal{Y}_{T-1}, \mathcal{X}_T, \mathbf{u}_T; \theta) = \mathbf{0}\,.
\end{aligned} \tag{6.7}$$

This can be regarded as a system of nT equations in the nT variables $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_T$ with np initial values $\mathcal{Y}_0$ and nk terminal values $\mathcal{E}_T$. This system can be solved using any of the nonlinear solution methods already considered as was pointed out by Hall (1985). The ordering of the equations

is important if one of the order-dependent methods such as *Gauss–Seidel* or *FGS* is being considered since ordering by time is no longer necessarily optimal.

The system (6.7) can also be solved using Newton's method. It has a Jacobian matrix of a block tridiagonal form. For the case $p = k = 1$ this looks like

$$\begin{bmatrix} \mathbf{J}_1 & \mathbf{F}_1 & & & \\ \mathbf{B}_2 & \mathbf{J}_2 & \mathbf{F}_2 & & \\ & \mathbf{B}_3 & \ddots & \ddots & \\ & & \ddots & \mathbf{J}_{T-1} & \mathbf{F}_{T-1} \\ & & & \mathbf{B}_T & \mathbf{J}_T \end{bmatrix}$$

where

$$\mathbf{J}_t = \frac{\partial \mathbf{f}}{\partial \mathbf{y}'_t} \quad , \quad \mathbf{F}_t = \frac{\partial \mathbf{f}}{\partial \mathbf{y}'_{t+1}} \quad , \quad \mathbf{B}_t = \frac{\partial \mathbf{f}}{\partial \mathbf{y}'_{t-1}}$$

are all matrices of dimension $n \times n$.

The *Stacked Newton method* applies Newton's method to the stacked system (6.7). This involves solving each iteration a system of linear equations of order nt, and, as a consequence has long been regarded as impracticable. However, efficient algorithms that take into account the special form of the Jacobian matrix have been suggested by Laffargue (1990) and Boucekkine (1995), or, alternatively, general black-box techniques for dealing with sparse matrices can be applied, as described in Duff *et al.* (1986). Gilli and Pauletto (1998) and Pauletto (1997) show how nonlinear Krylov methods can be used to solve the sparse system of equations iteratively. Recent applications of some of these approaches are described in Armstrong *et al.* (1998) and Juillard *et al.* (1998).

6.2.3 The Fair-Taylor solution method

Fair and Taylor (1983) propose a first-order solution method for a nonlinear dynamic model with consistent expectations (6.6). Rewriting the system in normalised form gives

$$\mathbf{y}_t = \mathbf{h}(\mathbf{y}_t, \mathcal{E}_t, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta)\,.$$

The method is as follows. First make an initial guess at the consistent expectations $\mathcal{E}^0_{t+1}, t = 1, \cdots, T$. Then, in iteration v, $v = 1, \cdots, V$ solve the system of equations

$$\mathbf{y}^v_t = \mathbf{h}(\mathbf{y}^v_t, \mathcal{E}^{v-1}_t, \mathcal{Y}_{t-1}, \mathcal{X}_t, \mathbf{u}_t; \theta) \quad , \quad t = 1, \cdots, T \tag{6.8}$$

treating the expectations $\mathcal{E}_t^{v-1}$as *fixed*. This problem can be solved using any of the first order solution methods considered above, or indeed by Newton's method. Then update the expectations $\mathcal{E}_t^v = \{\mathbf{y}_{t+1}^v, \cdots, \mathbf{y}_{t+k}^v\}$. Continue to iterate until

$$\max_t \left|\mathcal{E}_t^v - \mathcal{E}_t^{v-1}\right| < \varepsilon_3 \,.$$

Fair and Taylor call the iterations on v type II iterations to distinguish them from the type I iterations on s that comprise the solution of (6.8). It is often computationally efficient to set loose convergence criteria ε_1 and ε_2 for the type 1 iterations until the expectational variables are close to convergence.

It should be noted that Fair and Taylor suggested that the solution method described above should be nested in a further outer loop in which the solution period is extended until this has no more effect on the first T periods. The result of these type III iterations is to remove the effect of the terminal conditions on the solution. In practice, this third iteration loop is often ignored.

6.2.4 Shooting methods

Consider the reduced form of the stacked system (6.7) given by:

$$\begin{aligned} \mathbf{y}_1 &= \mathbf{g}(\mathcal{E}_1, \mathcal{Y}_0, \mathcal{X}_1, \mathbf{u}_1; \theta) \\ &\vdots \\ \mathbf{y}_T &= \mathbf{g}(\mathcal{E}_T, \mathcal{Y}_{T-1}, \mathcal{X}_T, \mathbf{u}_T; \theta) \,. \end{aligned} \tag{6.9}$$

By substituting out for the variables $\mathbf{y}_1, \mathbf{y}_2, \cdots, \mathbf{y}_T$ this can be interpreted as a system of equations expressing the terminal conditions $\mathcal{E}_T$ as a function of the initial expectations $\mathcal{E}_1$, the initial conditions $\mathcal{Y}_0$, the whole time path of the exogenous variables $\mathcal{X}_1, \cdots, \mathcal{X}_T$ and the $T \times n$ matrix of errors $\mathbf{U} = (\ \mathbf{u}_1 \ \cdots \ \mathbf{u}_T\ )'$:

$$\mathcal{E}_T = \phi(\mathcal{E}_1, \mathcal{Y}_0, \mathcal{X}_1, \cdots, \mathcal{X}_T, \mathbf{U}; \theta) \,. \tag{6.10}$$

All the variables in (6.10) are fixed except for the initial expectations $\mathcal{E}_1$. This equation writes the model solution problem as a *two-point boundary value problem* where the boundary values are the initial values $\mathcal{Y}_0$ and the terminal values $\mathcal{E}_T$.

Taking a first order Taylor expansion of (6.10) around the point $\mathcal{E}_1^0$ we have

$$\phi(\mathcal{E}_1) \simeq \phi(\mathcal{E}_1^0) + \frac{\partial \phi}{\partial \mathcal{E}_1^{0\prime}}(\mathcal{E}_1 - \mathcal{E}_1^0) = \mathcal{E}_T$$

suggesting the Newton iteration

$$\mathcal{E}_1^s = \mathcal{E}_1^{s-1} - \left(\frac{\partial \phi}{\partial \mathcal{E}_1^{s-1\prime}} \right)^{-1} (\phi(\mathcal{E}_1^{s-1}) - \mathcal{E}_T)\,. \tag{6.11}$$

This method for updating $\mathcal{E}_1$ to solve the model is known as the *single shooting method* and is a standard method for solving two-point boundary value problems. (See for example Roberts and Shipman (1972)). The Jacobian matrix in (6.11) will in general be impossible to calculate analytically but may be computed numerically or approximated by a matrix of constants.

Lipton *et al.* (1982) find that the single shooting method does not work well in rational expectations models. Instead they propose a method of *multiple shooting* which involves splitting the solution period into overlapping sub-periods and solving these sub-problems using a shooting technique. Fisher (1992) criticises shooting methods, finding them to be fundamentally unreliable for models that satisfy the usual stability conditions, and difficult ot apply to nonlinear macroeconomic models.

7 Further reading

Pauletto (1997) is a recent and excellent treatment of the computational techniques considered here and is strongly recommended. Fisher (1992) concentrates on first order solution methods for nonlinear rational expectations models, applying these methods to large models of the UK economy. A short and non-technical treatment in given in chapter eight of Cuthbertson, Hall and Taylor (1992).

References

[1] Armstrong, J., R. Black, D. Laxton and D. Rose (1998), 'A robust method for simulating forward-looking models', *Journal of Economic Dynamics and Control*, 22, 489–501.

[2] Blanchard, O. J. and C. M. Kahn (1980), 'The solution of linear difference models under rational expectations', *Econometrica*, 48, 1305–1311.

[3] Boucekkine, R. (1995), 'An alternative methodology for solving nonlinear forward-looking models', *Journal of Economic Dynamics and Control*, 19, 711–734.

[4] Cuthbertson, K., S. G. Hall and M. P. Taylor (1992), *Applied Econometric Techniques*, Harvester Wheatsheaf, Hemel Hempstead, UK.

[5] Don, F. J. H. and G. M. Gallo (1987), 'Solving large sparse systems of equations', *Journal of Forecasting*, 6, 167–180.

[6] Duff, I. S., A. M. Erisman and J. K. Reid (1986), *Direct Methods for Sparse Matrices*, Oxford University Press, Oxford, UK.

[7] Fair, R. C. and J. B. Taylor (1983), 'Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models', *Econometrica*, 51, 1169–1186.

[8] Fisher, P. G. (1992), *Rational Expectations in Macroeconomic Models*, Kluwer Academic Publishers, Dordrecht, Netherlands.

[9] Gilli, M. and G. Pauletto (1998), 'Krylov methods for solving models with forward-looking variables', *Journal of Economic Dynamics and Control*, 22, 1275–1289.

[10] Hall, S.G. (1985), 'On the solution of large economic models with consistent expectations', *Bulletin of Economic Research*, 37, 157–161.

[11] Juillard, M., D. Laxton, P. McAdam and H. Pioro (1998), 'An algorithm competition: first-order iterations versus Newton-based techniques', *Journal of Economic Dynamics and Control*, 22, 1291–1318.

[12] Laffargue, J-P. (1990), 'Résolution d'un modèle macroéconomique avec anticipations rationnelles', *Annales d'Economie et Statistique*, 17, 97–119.

[13] Lipton, D., J. Poterba, J. Sachs and L. Summers (1982), 'Multiple shooting in rational expectations models', *Econometrica*, 50, 1329–1333.

[14] Minford, A. P. L., K. G. P. Matthews and S. S. Marwaha (1979), 'Terminal conditions as a means of ensuring unique solutions for rational expectations models with forward expectations', *Economic Letters*, 4, 117–120.

[15] Muth, J. F. (1961), 'Rational expectations and the theory of price movements', *Econometrica*, 29, 315–335.

[16] Newton, I. (1686), *Philosophiae Naturalis Principia Mathematica*, translation by A. Motte (1729), revised by F. Cajori (1934), Univerisity of California Press, Berkeley, CA.

[17] Pauletto, G. (1997), *Computational solution of large-scale macroeconometric models*, Kluwer Academic Publishers, Dordrecht, Netherlands. Also available electronically on the world-wide web at *http://www.unige.ch/ses/metri/pauletto/t.ps*

[18] Roberts, S. M. and J. S. Shipman (1972), *Two-point Boundary Value Problems: Shooting Methods*, Elsevier, New York, NY.

[19] Sargent, T. J. (1979), *Macroeconomic Theory*, Academic Press, New York, NY.