SIMULATION AND CONTROL TECHNIQUES FOR NONLINEAR RATIONAL EXPECTATION MODELS

by

Paul Gregory Fisher

Thesis submitted for the degree of
Doctor of Philosophy

to the

UNIVERSITY OF WARWICK
Department of Economics

January, 1990.

CONTENTS

LIST OF TABLES AND FIGURES

ACKNOWLEDGEMENTS

This thesis has been developed during the course of the research programme of the ESRC Macroeconomic Modelling Bureau, partially in response to the needs of that work. Indirect support by the ESRC is therefore acknowledged. Special thanks are due to my supervisor and the director of the Bureau, Professor K.F. Wallis and to the Bureau's principal research fellow Dr. J.D. Whitley. I am also grateful for the comments and encouragement of Bureau colleagues M.J. Andrews, D.N.F. Bell, A.J. Longbottom, S.K. Tanna, and D.S. Turner. Thanks are also due to co-authors on particular topics: Professors A.J. Hughes Hallett and M.H. Salmon.

Finally I should like to thank S. Holly (London Business School), S. Hall and B. Henry (Bank of England), A. Rastogi and K. Matthews (Liverpool), R. Whittaker (Her Majesty's Treasury) and B. Rustem (Imperial College) for informal discussions on these issues and the models over a number of years.

DECLARATION

Some of the results presented in this thesis have already been published in jointly authored work. Specific acknowledgements are due as follows. Chapter 3, sections 3.1, 3.2 and 3.4 and the LPL and LBS results of section 3.5 are revised material from Fisher and Hughes Hallett (1988). Professor Hughes Hallett was primarily responsible for developing the analysis in sections 3.1, 3.2 and 3.4. These sections are the only material in the thesis not primarily the work of the author and are included for completeness. The author was responsible for the proposal and implementation of the general approach to incomplete inner iteration strategies including the precise technique employed and the generation of the results in section 3.5. The NIESR results in 3.5 and all of sections 3.6, 3.7, 3.8, 3.9 and 3.10 are new work.

Elsewhere in the thesis, results and material are used that have been previously published in volumes of *Models of the U.K. Economy* (Wallis *et al.*, 1985, 1986, 1987) or joint discussion papers. Although all the included material was originally generated by the current author and has since been revised, the work benefits substantially from the comments of Professor Wallis and Dr. Whitley in particular. Material in this category includes section 4.4 and parts of 5.2, 5.4, and 7.4. Sections 6.3–6.8 are largely material from a joint discussion paper with Professor Wallis. Any errors remaining either in these sections or in others are the sole responsibility of the present author.

SUMMARY

This thesis presents a comprehensive set of techniques for solving, simulating, analysing and controlling large scale, nonlinear, econometric models that contain rational expectations of future dated variables. These expectations are generally treated as model consistent, whereby the expectation is set to the deterministic projection of the model.

Solutions to such models are distinguished from those of conventional models by the fact that they are not recursive in time. The outcome for the current period depends on the expected outcome for future periods as well as past periods. This property means that all of the basic numerical procedures need to be altered.

We consider the following topics: solution algorithms for the two–point boundary value problem; terminal conditions, uniqueness and stability; experimental design and stochastic simulation; model forms, solution modes and historical tracking; control methods including optimal control. We find that suitable procedures allow us to undertake all of the experiments usually conducted with conventional models.

Each topic is illustrated by application to three large scale models of the United Kingdom economy which contain rational expectations terms. Only one of these models is constructed following the new–classical paradigm and hence their comparative properties revealed by our experiments are of some interest.

ABBREVIATIONS

BE	Bank of England
CUBS	City University Business School
EEC	European Economic Community
FGS	Fast Gauss–Seidel
FIML	Full information maximum likelihood
FOI	First order iterations
FT	Fair–Taylor
FWW	Fisher, Wallis and Whitley
GDP	Gross domestic product
HMT	Her Majesty's Treasury
III	Incomplete inner iterations
IV	Instrumental variables
JOR	Jacobi–over–relaxation
LBS	London Business School
LPL	Liverpool Research Group in Macroeconomics
MAE	Mean absolute error
NIESR	National Institute of Economic and Social Research
NPV	Net present value
OLS	Ordinary least squares
PSBR	Public sector borrowing requirement
RMSE	Root–mean–square error
SOR	Successive–over–relaxation
UK	United Kingdom
US	United States of America
VAT	Value added tax
2SLS	Two stage least squares

Chapter 1

INTRODUCTION

Throughout the last two decades, expectations formation has been at the heart of macroeconomic research. In particular, following the work of Muth (1961), attention has been concentrated on the concept of rational expectations i.e. that agents form expectations consistent with their knowledge of the underlying processes of the economic system and taking into account all available information. The rational expectations hypothesis has sufficiently penetrated all areas of macroeconomic theory and applied work to have caused a "revolution in macroeconomics" according to Begg (1982). Developments in macroeconomic theory are usually incorporated, sooner or later, into large–scale macroeconomic models. In the United Kingdom there are now three large, nonlinear, empirical, macroeconomic forecasting models which incorporate rational expectation terms as part of their basic structure.

The Liverpool model (LPL) is an annual model which has been based on rational expectations since its inception in 1979. It is new classical in structure with a set of long–run equilibrium equations determined only by the supply side. The London Business School (LBS) and National Institute of Economic and Social Research (NIESR) models are quarterly and used rational expectation models for their forecasts for the first time during 1985 (*Economic Outlook*, October; *National Institute Economic Review*, November). The LBS model has three expectation terms in its large financial sector. It is based around the income–expenditure framework and has a tradition as an international monetarist model. A more accurate description of current versions would be a sluggish price adjustment model. The NIESR model remained largely Keynesian despite the incorporation of the expectations of eleven variables with leads of up to four periods. The model is therefore a quantity adjustment model, being driven more by expenditures than relative prices.

The actual processes by which agents form expectations are generally unknown and their precise expectations concerning future macroeconomic quantities are usually unobserved. The assumption of rational expectations in economic theory therefore leads to complications both in estimating equations which contain explicit expectation terms and in the numerical procedures which are designed to solve, simulate and analyse large–scale models which contain such terms. The econometric literature now suggests a range of statistical methods which allow the estimation of equations containing rational expectation terms (e.g. see Begg, 1982, Ch.5). This thesis concerns numerical methods which allow the use of such equations in large–scale models.

1.1 The objectives of the thesis

The primary objective of this thesis is to establish a set of numerical methods for the simulation, analysis and optimal control of large–scale, nonlinear, macroeconomic models that contain rational expectations of future–dated variables. These methods allow us to undertake on rational expectations models all of the experiments usually performed with conventional large–scale models such as forecasting, policy analysis or stochastic simulation.

Numerical methods are often designed for specific problems or for specific models. In this thesis we aim to establish a general set of techniques which can be applied to a variety of models and problems. Starting from these general methods we can then modify the procedures to suit the requirements of particular experiments.

The secondary objective of the thesis is to use our procedures to investigate the properties of the three publicly available models of the U.K. economy which incorporate rational expectations terms. As noted by Wallis (1987):

> "In the United States the eventual adoption of the rational expectations hypothesis was associated in addition with the new classical equilibrium business cycle models and the policy ineffectiveness proposition."

In the U.K. this association has not been universally adopted and the properties of those models with rational expectations may not coincide with the new–classical model. The elucidation of the properties of these models is therefore of some interest.

In the three different U.K. models containing rational expectations, the only common expectations term is that of the sterling effective exchange rate. In a relatively small open economy, operating under a floating rate regime, fluctuations in the exchange rate can be a particularly important transmission mechanism. Expectations of future exchange rates changes appear to play an important behavioural role in determining its current value (Isard, 1988) and these terms are therefore of fundamental importance to the properties of U.K. models (e.g. see Wallis *et al.*, 1987, pp44–48). When explaining our results we pay particular attention to the behaviour of the exchange rate.

1.2 Outline of the research

In Chapter 2 we present a wide–ranging literature survey covering all aspects of large–scale modelling but paying particular attention to the numerical methods used. The role of expectations is discussed and we consider the problems caused by including expectation terms of future–dated variables. We conclude that the issues raised have only partially been resolved by the existing literature. Some problems, such as terminal condition choice, do not appear to have been considered in depth and some existing numerical methods, such as those for the optimal control of nonlinear rational expectations models, do not appear to be well understood.

In the third chapter of the thesis we consider the basic problem of solving a rational expectations model such that the expectations terms are consistent with the model's solution. We develop a family of first–order iterative techniques which encompasses some specific solution algorithms suggested elsewhere in the literature. These methods are tested on three large–scale models to derive the most efficient forms of solution algorithm. We contrast and compare our methods with two alternative approaches to solving nonlinear rational expectations models.

In Chapter 4 we address the issues of uniqueness and stability in the solutions of rational expectations models. Having established the conditions for a unique stable solution to exist, we are faced with the problem of locating that solution. This comes down to the choice of terminal values for the expectation terms in the final solution period. A number of possible terminal conditions are proposed and evaluated on both a small demonstration model and the three large-scale models. Particular attention is paid to the implications of different choices when there is not a unique stable solution.

The following chapter begins by examining the implications of different assumptions concerning input shocks. Such shocks may be introduced for the purpose of policy analysis or simply to evaluate the partial responses of the model. The shock can be treated as anticipated or unanticipated, temporary or permanent. The practical implications of such distinctions are evaluated for each of the three large-scale models.

The analysis of a sequence of temporary, unanticipated shocks leads us to a proposed method of stochastic simulation which differs from two methods proposed elsewhere in the literature. The differences are critically assessed and our preferred method is applied in an experiment to reveal the stochastic implications of alternative assumptions for the financing of the PSBR.

In Chapter 6 we present a general discussion of alternative model forms (structural form, reduced form, final form) and solution modes (single-equation, static, dynamic) extended to the rational expectations case. In particular we develop an appropriate method for the static simulation of rational expectations models. We argue that static rather than dynamic simulation is the correct procedure to be used in evaluating the historical tracking performance of models. A comprehensive historical tracking exercise is presented for six U.K. models, three containing rational expectations and three without.

In the seventh chapter we develop algorithms for the optimal control of nonlinear rational expectations models. Three algorithms are investigated which produce the optimal solutions corresponding to different formulations of the policy optimization problem. The differences in these solutions are discussed and critically assessed. One of the three algorithms is then used to derive optimal inflation–unemployment trade–offs for the three large–scale rational expectations models. Attention is paid both to the observed economic properties of the models and the costs of the optimization procedure. Finally, the implications are examined of alternative formulations of the problem.

The concluding Chapter 8 contains directions for future research and final remarks. The models used in this thesis are those made available by the ESRC Macroeconomic Modelling Bureau between Autumn 1985 and Autumn 1989. A precise list of the model vintages used in each chapter is given in an appendix. All the calculations have been made using software developed for the purpose and this software is attached to the macroeconomic model User service provided by the Bureau to U.K. academics over the Joint Academic Network.

Chapter 2

LARGE–SCALE MACROECONOMIC MODELS AND FORWARD EXPECTATIONS

In this chapter we begin by reviewing the nature and purpose of large–scale macroeconomic models. We then survey the development of the rational expectations literature and lead in to large–scale macroeconomic models with forward expectations. We consider the problems posed for conventional numerical methods in solving, simulating and analysing these models and the extent to which these problems are overcome by existing procedures.

2.1 Large–scale models of the economy

An economy is formed by a large number of agents (e.g. firms, consumers) engaging in economic activity: the production, distribution and consumption of goods and services. Each individual agent may have a unique behavioural pattern. Using economic theory, we can predict how representative agents will behave under given conditions – such as perfect competition in the product market. However, if we wish to model the behaviour of an entire economy, it would be a Herculean task to model the actions of each individual agent. Apart from the sheer size of the model, the realized decisions of any one actual agent may not conform to any known economic theory. In a macroeconomic model of a national economy, we therefore model the aggregate (or average as appropriate) behaviour of groups of agents. The size of any particular model is then dependent on the degree of aggregation chosen by the model builder or specified by the model user. In modelling the behaviour of economic agents, we are seeking to explain or predict the values of their decisions as realized in the economic data. These decisions are determined for each agent by their own behavioural pattern and by their conditioning information. That is to say the agents respond to the state of the world. In an aggregate or average representation of an economy, we are trying to explain (or predict) the movement of a set of macroeconomic aggregates such as consumption expenditure, the general price level or the exchange rate between domestic and foreign currency. We condition our model by choosing not to explain certain other aggregates on economic or statistical grounds. We may exclude certain policy variables (e.g. the rate of interest in some models); those which are determined outside the geographical economy of interest (e.g. world production if we are modelling the U.K.); or simply those variables which are within the system but which are not of interest to the user of the model (e.g. some demographic factors such as the birth rate).

The relationships between the variables being explained by the model (the endogenous variables) and those not being explained (the exogenous variables) are modelled using algebraic representations. These models then consist of a system of simultaneous equations. These equations may be derived entirely from theoretical analysis or they may have functional forms and/or numerical parameters which are based on empirical exercises. Models based on empirical research, often distinguished by the term "macroeconometric models" will usually be based on time series data recorded over some historical period. This gives the model builder a choice of temporal aggregation, usually annual or quarterly although continuous time models have also been developed (e.g. Bergstrom, 1967; Gandolfo, 1981).

The first macroeconomic model was presented by Jan Tinbergen over fifty years ago (Tinbergen 1936). In the period since, there have been many models developed for a range of purposes. In particular we identify three main uses of a large—scale macroeconomic model: understanding the processes of the economy; predicting the future course of the economy (forecasting); analysing the effects on the economy of external shocks or changes in policy variables. For the U.K. some recent developments in these areas have been surveyed by Wallis (1988, 1989).

A wide variety of models have been constructed, with different emphasis on the use to which they are put, at different levels of temporal and sectoral aggregation. In each case however, the model is treated as a framework for analysis to provide quantifiable evidence when all the relevant relationships in the economy are taken into consideration. Given this framework, the results obtained from any individual model may still depend on the way in which it is used: the answer depends on how the question is asked. This aspect is discussed by Turner, Wallis and Whitley (1989). A classic example is that the effects of an increase in government expenditure depend on how it is financed: by printing money, issuing debt or raising taxes. This particular example has its roots in the seminal papers of Christ (1968) and Blinder and Solow (1973).

Macroeconomic models are built to give a detailed representation of the economy and may range from just one or two equations up to many thousands (the project LINK model of the world economy contains some 20000 equations: Petersen, 1987). The definition of what size model comprises large–scale is somewhat arbitrary. The smallest model used in this thesis contains just 30 equations and the largest just over 1200. Large models are often multi–purpose, designed for all three of the activities noted above. The costs associated with building and maintaining a model increase with size and there is an obvious incentive to extract the maximum use value.

Each equation in a large model will be grounded to a greater or lesser extent in an economic analysis of a particular behavioural relationship. Typically the majority of equations in the model will be estimated. That is to say, they have functional forms, numerical coefficients and a choice of explanatory variables which are determined to some extent by the use of econometric methods. In this thesis we shall be largely treating the structure of models as given. The process of constructing models is covered in texts such as Fair (1984) and Holden, Peel and Thompson (1982).

The simplest representation of an economic system is a set of linear, static equations which we may write as:

$$B\, y_t + C\, x_t = u_t \,, \tag{2.1}$$

where y_t is an $(n \times 1)$ vector of observations on the endogenous variables at time t; x_t is an $(m \times 1)$ vector of observations on the exogenous variables and u_t is an $(n \times 1)$

vector of disturbance terms. The matrices B and C are $(n \times n)$ and $(n \times m)$ matrices of coefficient values respectively (derived or estimated). The n elements of u_t (u_{it}, $i=1,\dots,n$) are non–zero whenever equation i does not perfectly fit the observed data in period t. An equation j for which $u_{jt} = 0$ in every period $t=1,\dots,T$ is defined as an identity. A system for which B is diagonal is un–coupled or non–simultaneous. Off–diagonal elements in B represent the direct effect of the outcome of one decision on another.

In general, economic behaviour is dynamic in the sense that decisions made in period t depend on past decisions and conditioning information from earlier periods as well as x_t. A general dynamic model can be represented by use of the lag operator L which is defined such that $Ly_t = y_{t-1}$. We can then generalize equation (2.1) to:

$$B(L)\, y_t + C(L)\, x_t = u_t, \tag{2.2}$$

where B(L), C(L) are now matrices of polynomials of order p, q repectively such that $B(L)\, y_t = B_0 y_t + B_1 y_{t-1} + B_2 y_{t-2} + \dots + B_p y_{t-p}$ and C(L) is similarly defined.

Equation (2.2) still presents the model in linear form whereas in practice, most large–scale models involve nonlinear relationships. Examples include log–linear equations in which the variables are first transformed by taking logs or ratio transformations such as taking the current account balance as a fraction of total output in forming an explanatory variable. A combination of linear identities, log–linear equations and multi–variable transformations together with the occasional extreme nonlinearity (such as raising to a power) will typically yield a system which cannot be expressed in the form of equation (2.2). We therefore adopt a more general notation:

$$f(y_t, Y_{t-1}, X_t; \theta) = u_t \tag{2.3}$$

The vector of general functions $f(\cdot)$ is restricted only in that it is a real–valued function of real variables. The vectors defined with capital letters X_t and Y_{t-1} represent all observations up to the date indicated i.e. $X_t' = \{x_t', x_{t-1}', \ldots, x_{t-q}'\}$ and $Y_{t-1}' = \{y_{t-1}', y_{t-2}', \ldots, y_{t-p}'\}$. The vector θ represents all the parameters of $f(\cdot)$ and θ_i will represent the parameters of equation $f_i(\cdot)$. Although the vast majority of large–scale models are nonlinear, the linear forms (2.1) and (2.2) can still be useful as an analytical device for considering the behaviour of the nonlinear system (2.3). The linear system (2.2) could be considered as a first order approximation to the nonlinear system (2.3) in some local neighbourhood of its solution. The elements of B(L) and C(L) are then interpreted as partial derivatives of $f(\cdot)$. Hence in some parts of the exposition we continue to use the linear form (2.2) and this has the benefit that linear systems are easily manipulated to give closed form expressions whereas general nonlinear systems are not.

2.2 Solution and simulation of macroeconomic models

The use of a model for understanding the processes of the economy is partly a matter of examining the specific relationships embodied in the system. However, there are also various experiments which we may wish to conduct such as full system analysis, policy analysis or forecasting and these exercises require a set of numerical techniques for solving and simulating the model. The computational complexity of these techniques generally increases as we move from models of the form (2.1) through (2.2) to (2.3). Throughout this thesis, a detailed presentation of techniques will be made as we extend them to cover the forward expectations case. In this section we briefly summarise these procedures for models of the form so far considered.

The basic technique for using a large scale model is simulation. That is to say, the simultaneous equation system is solved for the endogenous variables (y_t) conditional on the form of the model, its coefficients and the values of the exogenous variables. The input may also include direct adjustments to y_t to account for off–model information. Hence solution of the nonlinear equation system (2.3) is an essential requirement. Solution procedures are covered by many authors e.g. Hughes Hallett and Rees (1983).

The static linear system (2.1) can be used to solve for y_t, $t=1,\dots,T$ for each t by obtaining the reduced form solution:

$$y_t = -B^{-1} C x_t + B^{-1} u_t. \tag{2.4a}$$

Hence the only numerical technique needed is the inversion of B which is required to be of full rank. In what follows, we assume that the disturbances (u_t) are set to zero for simplicity. For the dynamic model (2.2) we can solve for a static or dynamic sequence. In the static sequence the lagged values $y_{t-1},\dots,y_{t-p}$ are set to observed historical values and the system is solved for y_t. In the dynamic sequence, the solution value for y_1 is calculated using observed historical values but the solution for y_2 then depends on the solved value for y_1. The solution for y_t, $t>p$ uses solution values for all the lags on y. These solutions can be expressed as:

$$\hat{y}_t = -B_0^{-1} [B_1 y_{t-1} + \dots + B_p y_{t-p} + C_0 x_t + \dots + C_q x_{t-q}] \tag{2.4b}$$

$$\overset{*}{y}_t = -B_0^{-1} [B_1 \overset{*}{y}_{t-1} + \dots + B_p \overset{*}{y}_{t-p} + C_0 x_t + \dots + C_q x_{t-q}] \tag{2.4c}$$

$$t=1,\dots,T$$

$$\overset{*}{y}_{t-s} = y_{t-s},\ s\geq t,$$

for the static and dynamic solutions respectively. For dynamic as well as static systems the reduced form solution is obtained by matrix inversion. For large systems this inversion can be costly and the solution is usually obtained by solving the simultaneous equation system by direct methods which will be discussed below. In both cases we require B_0 to have full rank. The dynamic system need not be stable but the stability condition requires that the determinantal polynomial $|B(L)|$ has all its roots of modulus less than unity.

The nonlinear system (2.3) does not have a general solution with a convenient expression. Instead we assume that the conditions of the implicit function theorem hold and we write the reduced form of equation (2.3) as:

$$y_t = g(Y_{t-1}; X_t; u_t; \theta) \tag{2.5}$$

In general there may be more than one solution to equation (2.3) which satisfies equation (2.5) but of all the large-scale models there has been only one case of multiple solutions arising from nonlinearities reported in the literature (Friedman, 1971). Since the function $g(\cdot)$ is not usually obtainable analytically, we solve equation (2.3) directly by numerical methods. For some of the methods considered below we require that we can re-write equation (2.3) as:

$$y_t = h(y_t; Y_{t-1}; X_t; u_t; \theta). \tag{2.6}$$

As in the linear case, we set $u_t = 0$ for simplicity. In the nonlinear case however, this causes complications which we consider in the next section.

Solutions to the system (2.3) or (2.6) are typically generated either by derivative-based methods or first-order iterative methods. The commonest derivative-based technique is the Newton method. To obtain this numerical algorithm we consider an expansion of the i'th equation in system (2.3) around some trial solution $y^{(0)}$:

$$f_i(y_t^{(0)}, Y_{t-1}, X_t; \theta) - \sum_{j=1}^{n} [\partial f_i/\partial y_j]\Big|_{y^{(0)}} \cdot (y_j - y_j^{(0)}) = 0. \tag{2.7}$$

Solving equation (2.7) for i=1,...,n simultaneously yields the iteration:

$$y^{(s)} = y^{(s-1)} - F^{-1} f(y^{(s-1)}; Y_{t-1}; X_{t-1}; \theta) \tag{2.8}$$

where $F = [\partial f/\partial y]\Big|_{y^{(s-1)}}$ is the matrix of partial derivatives evaluated at iteration (s–1).

First–order methods are based on solving equation (2.6). These methods will be considered in greater depth in Chapter 3 but the simplest form, the Jacobi method, can be obtained by simply writing equation (2.6) as:

$$y_t^{(s)} = h(y_t^{(s-1)}; Y_{t-1}; X_t; u_t; \theta) . \tag{2.9}$$

For both derivative–based and first–order methods, the solution y_t^* is found when

$$\max_i \text{abs}[(y_i^{(s)} - y_i^{(s-1)})/y_i^{(s-1)}] < \tau,$$

where τ is some chosen tolerance level.

In the two algorithms given above, the ordering of the equations $i=1,...,n$ is irrelevant. Gabay *et al.* (1980) and Don and Gallo (1987) have shown that the efficiency of the Newton solution procedure can be improved by exploiting any sparseness (zero entries) of the F matrix to reduce the dimensionality of the matrix inversion in equation (2.8). Hughes Hallett and Fisher (1987) have shown that any such observed sparseness can be used to re–order the equations to improve the efficiency of general first–order methods.

The solution methods discussed in this section are the basic tool used in the simulation of macroeconomic models. They allow us to calculate the values of the endogenous variables y_t conditional on alternative values for the exogenous variables x_t – which might be actual or forecast values; an hypothesised policy setting or simply a shock to the system . In addition we will consider two further types of simulation which require more complicated techniques: stochastic simulation and optimal control.

2.3 Stochastic simulation

The solutions delivered by equations (2.4) and algorithms (2.8) and (2.9) are obtained by setting the disturbance terms to their expected values $E(u_t) = 0$. That the observed disturbances (residuals) have mean zero over the estimation period is a property of most econometric estimators. If an equation is observed to have residuals with a non-zero mean (e.g. post-sample) then an estimate of the mean value is sometimes used rather than zero. In either case, these solutions are defined as deterministic because they offer a single point estimate of y_t ignoring the stochastic nature of the system. A stochastic simulation is an (Monte Carlo) experiment undertaken with a model in order to approximate the distribution of y_t^*.

It has been known, at least since Howrey and Kelejian (1971), that the deterministic solution to a nonlinear model is generally a biased estimate of its conditional expectation. This arises because the expectation of a nonlinear function of a random variable is not generally equal to the same function of the expectation of that variable i.e.

$$E(g(Y_{t-1}; X_t; u_t; \theta)) \neq g(Y_{t-1}; X_t; E(u_t); \theta). \tag{2.10}$$

Furthermore, the deterministic solution to equation (2.3) gives us no information on the distribution of y_t unlike the linear model (2.1) in which, for example, the conditional variance-covariance matrix of y_t can be obtained directly as $B_0^{-1}\Psi_t B_0^{-1}$ (where $\Psi_t = E(u_t u_t')$) and hence estimated from the residuals $\hat{u}_t = y_t - \hat{y}_t$ (where $\hat{y}_t$ denotes the fitted value of the estimated equation). Stochastic simulation can be used to obtain an unbiased estimate of the conditional expectation of $g(\cdot)$ and estimates of the higher order moments.

The stochastic simulation generates repeated solutions to the model for R successive draws of an $(n\times1)$ vector of pseudo-random disturbances $s_t^{(r)}$, $r=1,\ldots,R$. One vector of these shocks is then introduced in place of u_t for each replication. The successive solutions generate an empirical distribution for y_t from which we can estimate the parameters of interest. In order for the input to approximate the correct distribution of u_t, various schemes have been suggested for calculating the pseudo-random shocks. The Nagar method (Nagar, 1969) suggests drawing deviates from the multivariate normal distribution using an estimate of the residual covariance matrix based on residuals over some historical period. For an $(n\times1)$

vector of pseudo–random standard normal deviates $v_t^{(r)}$ and for an estimated variance–covariance matrix $\hat{\Psi} = \sum_{t=1}^{T} \hat{u}_t \hat{u}_t' / T$ we obtain a set of disturbances $s_t^{(r)\prime} = v_t^{(r)\prime} W$ where $W'W = \hat{\Psi}$. This method is unfortunately not generally feasible because the required decomposition of the covariance matrix is only obtainable when that matrix ($\hat{\Psi}$) is non–singular. We seldom have enough degrees of freedom to identify all the $(n*(n+1))/2$ parameters involved in the covariance matrix.

The McCarthy method (McCarthy, 1972) is the most commonly adopted residual generation technique for stochastic simulation of large systems. It overcomes the problem of the Nagar procedure by using the residuals directly instead of decomposing the variance–covariance matrix to generate the pseudo–random shocks i.e. $s_t^{(r)\prime} = T^{-1/2} v_t^{(r)\prime} \hat{U}$ where $\hat{U}$ is the $(T*n)$ matrix of observed residuals and $v_t^{(r)}$ is now a $(T*1)$ vector of pseudo–random standard normal deviates. This procedure generates disturbances with a covariance matrix which, for large R, tends to that of the residuals. This procedure can be extended to include the serial correlation properties of $\hat{u}_t$; details are given in Schink (1971). For estimating the bias of the deterministic solution we calculate the difference between that value and the mean of the stochastic replications (which is our estimate of the conditional expectation). This estimates the extent of the inequality in equation (2.10). Calzolari (1979) has suggested a method of negative antithetic variates which helps to reduce the experimental variance in estimating the bias. Calzolari and Sterbenz (1989) have recently proposed a generalization of the McCarthy procedure which explicitly maintains the empirical third moment of the disribution of the residuals.

The methods of stochastic simulation noted above are all parametric in that they rely on an underlying assumption of normality in the disturbances with well defined moments. The method of Mariano and Brown (1984) uses a non–parametric procedure which does not require these assumptions to be made. In this approach a sample of the residuals is directly used as the set of input disturbances and the resulting distribution of the shocks is therefore identical to the historical distribution. By sampling randomly and repeatedly from the observed residuals, this procedure becomes a boot–strap type method (Mariano, 1985).

Fisher and Salmon (1986) present a survey of stochastic simulation experiments and examine the implications of changes in the experimental design for the measurement of the bias. This survey is reproduced and slightly extended in Hall and Henry (1987). Fisher, Wallis and Whitley (1985) use stochastic simulation to estimate second moments and their results are extended in this thesis.

2.4 Optimal control

Control theory is a subject area that covers a wide spectrum of techniques which, in the economics literature, is aimed primarily at policy optimisation (e.g. Chow, 1975, 1981). In the context of large–scale, nonlinear macroeconomic models, optimal control is a term usually used to indicate a particular numerical technique for the analysis of policy which is optimal in some explicit sense. However, optimal control can also be viewed as a method of generating simulations or as an aid to full system analysis.

When forecasting with a model we are usually interested simply in obtaining a solution for y_t, $t=1,\ldots,T$ conditional on x_t, $t=1,\ldots,T$. This solution will then be treated as a base or central forecast around which we conduct sensitivity or policy analysis. In the variant forecast or policy simulation of a model we then choose a subset of the elements (sometimes only one) of the vector of exogenous variables and perturb these elements by a particular value i.e. we replace x_t by $x_t+\delta_t$. We then obtain the solution for y_t conditional on $x_t+\delta_t$, $t=1,\ldots,T$. This new solution will then be compared to the base solution (in which $\delta_t=0$). The design of δ_t may represent a particular policy proposal or it may be chosen simply to elicit a good estimate of the partial derivative $\partial y_{it}/\partial x_{js}$. Control of a model entails the design of δ_t so as to achieve, as near as possible, some desired outcome say $\bar{y}_t$, $t=1,\ldots,T$. In this case the input δ_t is calculated as part of the simulation process rather than pre–specified. As part of the problem we may also have specified a desired trajectory, *ceteris paribus*, for x_t^*, t=1,...,T.

For linear models there is a substantial literature covering methods for the optimal and sub–optimal control of economic systems. These techniques were originally developed in the engineering literature and adapted for economic systems because of their mathematical similarity to physical systems (see e.g. Chow 1975, 1981; Holly, Rustem and Zarrop (eds.) 1979). We therefore begin by summarising the basic results of this literature.

We begin by re–writing the general dynamic model (2.2) as:

$$y_t = B\, y_{t-1} + C\, x_t + D\, z_t + u_t \tag{2.11}$$

where we have separated out a vector of policy variables z_t from the other exogenous variables (x_t). The dynamics in (2.2) are reduced to first–order terms only by defining new variables dated t for the longer lags. The y and x vectors in equation (2.11) are then no longer the same as those in equation (2.2). There is no unique way of reducing a general dynamic model to a first–order system but various state–space forms are popular (e.g. see Aoki, 1976).

In formulating a control problem, we set up a formal explicit objective function which specifies the desired trajectories for both endogenous variables (henceforth targets) and exogenous variables (henceforth instruments) as well as the costs associated with deviating from the desired paths. The most common approach is to use a quadratic function such as:

$$J_t = \tfrac{1}{2} \sum_{i=t}^{T} \left[(y_t - \tilde{y}_t)' W_{yt} (y_t - \tilde{y}_t) + (z_t - \tilde{z}_t)' W_{zt} (z_t - \tilde{z}_t) \right], \tag{2.12}$$

where tilde (e.g. $\tilde{y}_t$) denotes a desired or target value and W_{zt}, W_{yt} are positive semi–definite and positive definite weighting matrices respectively. We then define the optimal values of z_t as those which minimise the cost function J_t, t=1,...,T subject to the model (2.11). Strictly, we minimise the expectation of J_t, t=1,...,T since equation (2.11) has a stochastic disturbnance term. However, we ignore the stochastic component in what follows. The optimal values can be determined in the form of a linear control rule:

$$x_t = K_t y_{t-1} + h_t, \tag{2.13}$$

where K_t, known as the feedback gain, and h_t known as the tracking gain (or feedforward term) can be expressed in closed form as functions of the model parameters, the weighting matrices and the desired values (for all time periods) see e.g. Holly and Hughes Hallett (1989, pp41–42). The feedback rule is complicated and time–varying. The precise form of these expressions is therefore only of relevance in obtaining the optimal values.

In 1978 the Committee on Policy Optimisation chaired by Professor R. J. Ball reported to Parliament on the possible application of optimal control techniques to Her Majesty's Treasury' macroeconomic model (Cmnd. 7148, 1978). For such nonlinear models there are no closed form analytical solutions available. In the following presentation we therefore resort to a linear model from which we can derive nonlinear methods.

We now write the economic model in stacked form as:

$$D\,Y + E\,X = U, \tag{2.14}$$

where D and E are the parameter matrices containing the system matrices (2.2) stacked over time and (Y, X) denote vectors of observations stacked over time such that: $Y' = \{y_1', y_2', \dots, y_T'\}$ and y_t and x_t are now the same vectors as in equation (2.2). We also stack the objective function over time:

$$J = \tfrac{1}{2}(Y-\tilde{Y})' W_y (Y-\tilde{Y}) + \tfrac{1}{2}(X-\tilde{X})' W_x (X-\tilde{X})\,. \tag{2.15}$$

The tilde (e.g. $\tilde{Y}$) denotes a given target value. The weighting matrices W_y, W_x

are (nT×nT) and (mT×mT) repectively and are chosen to be positive semi-definite and elements of y or x that are not targets or instruments have zero weights attached. The objective function J is then minimized subject to the constraint imposed by the model (2.14). The optimal solution can be found by substituting the constraint (2.15) into the model (2.14) and differentiating the resulting function directly to obtain optimal values (Y^o, X^o) from the first order conditions and hence:

$$X^o = [\Pi' W_y \Pi + W_x]^{-1} [W_x \bar{X} + \Pi' W_y \bar{Y} - \Pi' W_y D^{-1} U] \qquad (2.16a)$$

$$Y^o = -D^{-1} (E X^o - U) \qquad (2.16b)$$

where $\Pi = -D^{-1}E$. For a nonlinear system, we obtain the minimum of the objective function J by numerical means as follows. Taking the function (2.15) and substituting in the model (2.14) we have a general nonlinear function. This can be minimised using derivative-based methods as for the basic solution problem in solving the nonlinear model (2.3). In practice the dimension of this resulting equation system is so huge (even for just a few instruments/targets) that more efficient approximations to Newton's method are used. Rustem and Zarrop (1979, 1981) propose various quasi-Newton algorithms for minimizing J subject to a general nonlinear model.

In general the equations given by (2.16) are simply another way of writing the feedback control rule. However, in the numerical approach there is no attempt to formulate the solution as a linear control rule – the result is simply a set of optimal values for x_t and hence y_t. The other distinguishing feature is that (2.16) solves for all time periods simultaneously whereas (2.13) can solve for a single period (as a function of desired values and weights in all periods).

A recent example of the application of optimal control to economic models may be found in Ghosh, Gilbert and Hughes Hallett (1987) who looked at the problems of stabilising the copper market. Applications to large-scale models have been relatively few. Henry *et al.* (1982), Mellis (1984) and Wallis *et al.* (1987) all use optimal control analysis to derive output–inflation trade–offs in the manner of Chow and Megdal (1978). This form of analysis is to be examined and extended in Chapter 7.

2.5 The role of expectations

The role of expectations in macroeconomic relationships has long been recognized. In Keynes' General Theory (1936) the influence of expectations received much explicit attention. There is one chapter considering "Expectation as determining output and employment." and another on "The state of long–term expectation." . In his "General theory of the rate of interest.", Keynes offered a detailed example which is now used as a standard introductory explanation of investment behaviour. An investment project will be profitable if its discounted stream of net revenues is positive. This Net Present Value may be defined in terms of revenues R, costs C and discount factor d over the life of a project (T years) by the formula:

$$NPV = \sum_{t=0}^{T} (R_t - C_t)/(1+d_t) \qquad (2.17)$$

However, in an uncertain world the values of R_t, C_t and d_t are not known in the initial period 0. Hence the NPV is evaluated using expectations of these variables. Therefore investment decisions must depend on expectations of the future.

The treatment of expectations in the economics literature initially developed slowly. An important early contribution was made by Grunberg and Modigliani (1954) who showed how forecasts could be self–fulfilling if one takes into account the reactions of agents to the forecasts. The modern literature on "rational" expectations is often assumed to start with Muth (1961). It was Muth who coined the phrase and declared a rational expectation to be (p46):

" ... essentially the same as the predictions of the relevant economic theory. ".

In the context of a system of algebraic equations subject to a random disturbance term, this is interpreted in mathematical terms as the appropriate conditional expectation.

The rational expectation of a variable y_t formed in period $s<t$ and based on a particular theory or model is distinguished from other mechanisms for generating expectations such as extrapolative forecasts. Adaptive expectations are a particular kind of extrapolative forecast which are also considered by Muth. The adaptive expectation denoted y_t^a may be expressed in the following manner:

$$y_t^a = y_{t-1} + \lambda(y_{t-1}^a - y_{t-1}) \tag{2.18}$$

for some parameter $1 \geq \lambda \geq 0$. Hence the adaptive expectation of y_t is a function of its own past and is essentially an adjustment mechanism designed to use information on past errors to improve current forecasts.

The rational expectation of variable y_t formed in period $s \leq t$ will be denoted $y_{t|s}$ and defined as $E(y_t | \Omega_s)$: the conditional expectation of y_t given the information set Ω_s. The information set is assumed to contain all data on the variables of the model which are available at the end of period s and the relevant model (or theory). The information set could alternatively be defined for the beginning of a period and dated appropriately for the context. In Muth's simple example the adaptive expectation and the rational expectation coincide but this is a very special case.

The expansion of economic research incorporating explicit expectations variables could be attributed to the introduction of the expectations augmented Phillips curve. This is, in turn, often attributed to work by Friedman (e.g. 1968) and was incorporated into an empirical model by Lucas and Rapping (1969). By allowing expectations of inflation to shift the relationship between output and inflation (the Phillips curve), Lucas and Rapping showed that it is possible to reconcile the observed existence of such a relationship with the classical proposition that no such relationship exists. The long–run trade–off between output and inflation in Lucas and Rapping's model becomes non–existent as inflation expectations match actual inflationary changes. This model then helped to explain the observed breakdown in the output–inflation relationship which occurred in the 1960's. Lucas and Rapping explicitly reject the assumption of rational expectations but the primary importance attached to the role of expectations opened the way for later developments.

During the 1970's, the rational expectations hypothesis was adopted as an integral part of the new–classical economics. Particularly important contributions were made, *inter alia*, by Lucas (1972 a,b; 1973; 1975), Sargent (1973, 1976), Sargent and Wallace (1973, 1975, 1976), Barro (1976, 1977) and Kydland and Prescott (1977). These authors combined rational expectations with new–classical macroeconomics and the policy ineffectiveness proposition. An excellent survey and critique of this literature is given by Shiller (1978). Following the theoretical literature, rational expectations began to appear in small macroeconomic models (e.g. Sargent, 1976; Taylor, 1979).

The estimation of behavioural relationships which incorporate expectations is complicated by the fact that these expectations are not observable. Although some survey data exist these may not be a good proxy for actual economy–wide expectations (Mishkin, 1981). This complication led to a line of econometric research with important contributions by, *inter alia*, Nelson (1975); McCallum (1976 a,b) who proposed an errors–in variables (instrumental variables) approach; Hansen and Sargent (1980); Wallis (1980) who presented a summary of the econometric implications of the rational expectations hypothesis; Wickens (1982) who generalized McCallum's errors–in–variables approach to sub–system FIML; Hansen (1982) proposed a class of generalized methods of moments estimators; Fair and Taylor (1983) offered the Full Information Maximum Likelihood method; and, more recently, Nijman and Palm (1989) examine generalised least squares methods. Pagan (1986) also gives a survey of estimation methods with constructed regressors which integrates much of this literature.

As the explicit treatment of expectations became more common, an important distinction was made between different types of expectations variables. In particular we shall now consider forward expectations (e.g. $y_{t+1|t-1}$) and current expectations e.g. ($y_{t|t-1}$). Let us consider a wage equation in which nominal wages (W_t) depend on expectations of prices in the current period formed in the previous period so that we have $W_t=w(p_{t|t-1})$. We call this a current expectations term. As mentioned above, we may alternatively consider the current expectation to have been formed using information available at the start of period t and hence write $W_t=w(p_{t|t})$. This difference is simply a question of dating the information set according to the underlying economic theory. However, as discussed by Aoki and Canzoneri (1979), the forward expectation is different in kind. In this case we assume that wages are a function of the expectation of next period's prices: $W_t=w(p_{t+1|t-1})$ (or $W_t=w(p_{t+1|t})$ as appropriate). The information set can be the same in the forward expectations case as in the current expectation case but the implications for modelling are quite different.

Forward and current expectations are both used in the rational expectations literature, as appropriate to the underlying model. The models of Sargent and Taylor both used current expectations only. As shown by Aoki and Canzoneri (1979), and Anderson (1979) (and our simple example which follows) the use of current expectations is more tractable since the conditional expectations $y_{t|t-1}$ or $y_{t|t}$ can usually be substituted out using the reduced form of the model and leaving the solution as a function of lagged values of exogenous variables only. If, on the other hand, we substitute out for $y_{t+1|t}$ or $y_{t+1|t-1}$, then we introduce forward dated terms in the exogenous variables. To obtain a solution as a function of lagged values only, we then need to specify separate processes for the expectations of the exogenous variables.

To illustrate this important distinction we consider the price equation from Muth's (1961) paper which can be written as $p_t = -\alpha p_{t|t-1} + u_t$, where p is the deviation of price from its equilibrium and u may be either a random error or known process. We can solve this equation for the expectation by conditioning both sides and collecting terms to give $p_{t|t-1} = (1+\alpha)^{-1} u_{t|t-1}$. We can then solve for p_t if we know u_t i.e. $p_t = -(1+\alpha)^{-1} \alpha u_{t|t-1} + u_t$. Hence p_t is a function of lagged and current terms only. However, suppose the price equation had instead been $p_t = -\alpha p_{t+1|t-1} + u_t$. Now to obtain $p_{t+1|t-1}$ we lead this equation and condition both sides on t−1 but this leaves $p_{t+1|t-1}$ as a function of $p_{t+2|t-1}$. Repeating the process of leading the equation successively yields a geometric expansion which solves as $p_{t+1|t-1} = \sum_{i=1}^{\infty} -\alpha^{i-1} u_{t+i|t-1}$. Substituting this back into the price equation gives p_t as a function of expected u into the infinite future. This solution then requires us to specify a process for the expectations of u_t or its known future values.

Estimation methods for the forward expectations case are included in the procedures of McCallum, Wickens and Fair and Taylor. The distinction between forward and contemporaneous expectations is particularly important in simulation since the presence of forward expectations introduces a dependency of the solution in period t on anticipated changes at future dates. This forward dependency requires changes to our numerical procedures which are the subject of this thesis.

Anderson (1979) uses a model with current expectations only and introduces a solution procedure for the forward expectations case which, at the time that paper was written, had never been implemented. The first model incorporating forward expectations is that of Fair (1979) who constructed a medium sized model (84 equations) with rational expectations in the bond and stock markets. In the U.K., the first forward expectations model was that of P. Minford and known as the Liverpool (LPL) model. This model, based on annual data had its inception in 1979 and is described in Minford *et al.* (1984). The LPL model has a new classical approach with a set of supply–side equations which are causally prior to the demand–side equations which depend on it.

Since the LPL model was introduced, rational expectations have been adopted as the standard operating mode for forecasting and policy analysis for two much larger models of the U.K. economy. Both the London Business School (LBS) and National Institute of Economic and Social Research (NIESR) models incorporated forward expectations during their forecasting rounds of 1985 (Economic Outlook, October; National Institute Economic Review, November). The LBS model contains 3 forward expectations in its large financial sector. The NIESR model has a more pervasive influence from forward expectations of 11 variables with leads of up to four periods ahead. Recent Australian models have also adopted forward expectations (Murphy, 1989) as have some European models (e.g. Lahti and Viren, 1989). In the U.S. large–scale modelling work with forward expectations seems to have remained with variants of Fair's model (Fair, 1984). Other groups have introduced forward expectations as a variant solution. The in–house version of the U.K. Treasury model (e.g. Westaway and Whittaker, 1986; Melliss *et al*, 1989) is solved for forward consistent expectations whereas the public release contains explicit, backward–looking expectations generation mechanisms.

The explicit treatment of expectations in large–scale models has several implications. Firstly it allows relationships to be incorporated which are more in accordance with the underlying economic theory – such as that underpinning the LBS and LPL models. Secondly forward expectations allow for a different type of dynamic response in models; in particular the dependence on the anticipated future course of the economy as noted above. Commenting on the NIESR model, Hall and Henry (1985 a,b) claim that the presence of forward expectations enables the model to track the U.K. recession of 1981 better than the earlier versions of the model – a claim repeated in the historical tracking exercise of Hall (1987).

A further aspect of these models is that they help to defuse the well–known Lucas critique of policy making with macroeconomic models (Lucas, 1976). Lucas argued that the behaviour of agents is conditioned on their expectations of economic policy variables and the rules that drive those policies. A model which ignores this dependence may nevertheless be adequate as long as policy does not change. However, any substantial change in policy would potentially invalidate the model's structure. An example is the output–inflation relationship described in the article by Lucas and Rapping (1969). An observed trade–off between output and inflation may be conditioned on expectations of a government policy target of zero inflation. If the government tries to exploit the observed trade-off by generating higher inflation, the relationship may shift. Policy proposals based on a model which does not incorporate this shift could be seriously in error. The obvious riposte to this criticism is to treat expectations explicitly wherever they are theoretically appropriate. The current generation of U.K. models therefore has a defence to the Lucas critique.

2.6 The solution of forward expectations models

We can express the simplest form of forward expectations model as:

$$y_t = Ay_{t+1|t-1} + By_{t-1} + Cx_t + u_t \tag{2.19}$$

Equation (2.19) is a simplified dynamic model of the form of equation (2.2), augmented by two features. We add the one-period-ahead expectation $y_{t+1|t-1}$ (an n×1 vector) as an explanatory variable. In estimation we usually explicitly condition the expectation term on an information set dated at the end of period t–1. This information set Ω_{t-1} includes Y_{t-1}, X_{t-1}, U_{t-1}, the model, its parameter matrices, and forecasts (or values known at t–1) of x_r, u_r r=t,...,T as appropriate to the method. Hence, even in estimation, we can vary the degree of information which enters Ω_t. The basic dynamic solution problem is that of obtaining the dynamic forecast sequence $\hat{y}_t$, t=s+1,...,T. In this case we keep the information set fixed at Ω_s for all periods and thus actually obtain $\hat{y}_{t|s}$, t=s+1,...,T . The expectations terms could be generated by an adaptive expectations mechanism or other extrapolative forecast. In this case we could add the relevant equations to the system and the expanded system would then be backward looking. Hence it could be be solved in a recursive, period-by-period manner as before.

The rational expectation $y_{t+1|s}$ is given by the conditional expectation of the model. For a linear system this is equivalent to the model's own forecast $\hat{y}_{t+1}$. In setting $\hat{y}_{t+1|s}=\hat{y}_{t+1}$ we obtain model consistent expectations. A distinction between model consistent and rational expectations occurs only if the model is nonlinear when, as we showed earlier, the deterministic forecast is generally a biased estimate of the conditional expectation. In generating model consistent expectations over a finite horizon, we require a terminal value for y_{T+1} which is outside our solution period. Choice of this value and its implications are discussed in Section 2.7 below.

In order to generate the sequence y^*_t, t=s+1,...,T, we can no longer use recursive, period–by–period procedures since the solution for each period depends on that for all future periods as well as lagged values. Instead we must obtain the entire sequence simultaneously. There are four generic methods of solution suggested in the literature. The first of these are methods for solving linear systems and we examine the best–known approach. Other methods, for nonlinear systems, are dealt with in depth in Chapter 3.

The Blanchard and Kahn solution (Blanchard and Kahn, 1980) gives a general solution to forward expectations models and considers the uniqueness and stability of the solution. This analytical approach requires us to write the model (2.11) as a first–order linear difference equation:

$$z_{t+1} = D\, z_t + B\, x_t. \tag{2.20}$$

Equation (2.11) can be written in this form by defining new variables for the lagged terms in y (or x) as in the classical optimal control problem examined in Section 2.4. These new variables are added to the y vector to give z and the matrix D equates the forward values of these variables to their current values. Disturbance terms may be incorporated in x. In order to generate the dynamic solution we condition on an information set Ω_0 and solve for consistent expectations. Blanchard and Kahn derive the solution to this first order difference equation (2.20) by using the Jordan canonical form of D:

$$D = K^{-1} J\, K \tag{2.21}$$

where the diagonal elements of J are the eigenvalues of D ordered by increasing absolute value. The vector z_{t+1} is partitioned into its predetermined and non–predetermined components. The $(m_1 \times 1)$ predetermined part $z_{1,t+1}$ will be the constructed variables and those endogenous variables whose expectations do not appear in the model; $z_{2,t+1}$ is then $(m_2 \times 1)$ and $m_2+m_1=m$. The matrix J is partitioned such that the first p_1 diagonal elements lie within the unit circle and the next p_2 lie outside; $p_1+p_2=m$. Hence D has p_1 eigenvalues within the unit circle. The matrices K and K^{-1} are then partitioned conformably with J.

Blanchard and Kahn then show that if $p_2 = m_2$ i.e. the number of non–predetermined variables is equal to the number of eigenvalues of D outside the unit circle, then there exists a unique solution which is obtained by pre–multiplying (2.20) by K and using a partitioned inverse to obtain separate closed form expressions for z_{1t}, z_{2t} as a function of lagged and future values of x and elements of the canonical form of D. It can then be shown that if $p_2>n_2$, then there is no solution to the model which is non–explosive (i.e. stable). Blanchard and Kahn's non–explosion condition rules out exponential growth in the endogenous variables, having been applied to the exogenous variables by assumption. Finally they show that if $p_2<n_2$ then there exists an infinite number of non–explosive solutions.

To summarize, these results yield a closed form, analytical expression for the solution to a linear forward expectations model formulated in discrete time. Buiter (1984) gives the solution for continuous time models. Both Blanchard and Kahn, and Buiter give us the conditions for uniqueness and stability. A solution satisfying these conditions is generally called a saddlepoint path.

The solution forms given above require the calculation of eigenvalues and so can only be applied to a linear system. A linear model can usually be manipulated to give the required first–order difference equation form – Preston and Pagan (1982) suggest "shuffle" algorithms to obtain this form from a general structural model such as system (2.19). Computer based programs are available for obtaining both the continuous and discrete time solutions (Austin and Buiter, 1982). One approach would therefore be to linearize the large–scale nonlinear models and use one of these solution packages. Such an approach is used by Gaines *et al.* (1987).

In this thesis we are concerned with obtaining the direct solutions for nonlinear models for which the precise forms of these linear solutions are not directly relevant. However, the conditions for uniqueness and stability can still be useful since we have no simple equivalent for general nonlinear systems. Hence we may wish to examine these issues for a local linearization of the nonlinear system. This will be addressed briefly in the next section.

For solving the nonlinear system directly there are three general types of approach. These will all be examined in Chapter 3 and so we briefly summarise them here. Shooting methods are a general class of algorithms for solving two–point boundary value problems and have been known for some time in the numerical analysis literature (e.g. Keller, 1968; Roberts and Shipman, 1972). Lipton, Poterba, Sachs and Summers (1982) suggested their use in economic models in a form known as multiple shooting. This approach takes the model in a form similar to the first–order difference equation representation of Blanchard and Kahn but then solves numerically. Spencer (1985) has suggested an improvement to the basic technique to overcome some observed numerical instabilities. There appear to be no large–scale nonlinear models which actually use shooting algorithms on a regular basis.

There are a variety of first–order iterative techniques available based on the methods usually employed to solve large–scale models. The extension to forward expectations models was first noted by Anderson (1979), used by Fair (1979) and developed independently by Minford *et al.* (1979, 1980). The most comprehensive version of the algorithm is given by Fair and Taylor (1983). These methods are based on simple two–part iterations (i) solving a model for fixed expectations and then (ii) updating those expectations to be consistent. This two–part iteration is repeated until expectations are consistent after step (i). Hall (1985) gives a variation which was based on stacking the time periods to generate one huge equation system. Fisher, Holly and Hughes Hallett (1985, 1986) develop a complete family of first–order methods which include the Fair–Taylor and Hall approaches as special cases.

Finally we have derivative–based algorithms which follow the two–part scheme of first–order iterations but which update the expectations at step (ii) using Newton's method instead of a simple updating formula. This approach includes the penalty function method which adapts the optimal control problem, treating the consistent expectations condition as a target and the expectations variables as instruments. An explicit objective function is then minimized using approximations to Newton's method. This approach was first proposed by Holly and Zarrop (1979, 1983) and by Holly and Beenstock (1980). Holly and Zarrop used this solution method in conjunction with a policy optimization problem. That joint procedure will be examined in Section 2.9.

Beyond the basic dynamic solution problem, we also need methods for obtaining a static or one–step–ahead path which may be used, for example, in historical tracking. The problem in this context is of defining the unobserved expectation variables. This issue does not appear to have been directly addressed by the modelling literature although it is similar to the estimation problem in that an historical value is required for an unobserved variable. Historical tracking has been undertaken on forward expectations models by Hall (1987) using the NIESR model and Matthews and Minford (1987) using the LPL model. Hall uses a number of dynamic solutions starting in the first quarter of successive years and simply graphs all the solutions. Matthews and Minford similarly use dynamic solutions, one starting in each period of the horizon considered. Neither paper details the information assumptions being made.

The generation of a stochastic simulation has been addressed in practice by Hall and Henry (1985a,b) and in theory by Fair (1984, pp383–384). However neither of these authors gives a detailed discussion of the issues involved and conflicting procedures are suggested. This conflict is partly explained by the fact that the Hall and Henry procedure was not intended as a means of generating a full distribution of the endogenous variables but as a way of obtaining rational rather than consistent expectations from the nonlinear model. As noted above the bias in deterministic solutions of nonlinear systems as estimates of conditional expectations creates a difference between consistent and rational expectations. Allowing for this distinction it remains unclear that either the Hall and Henry procedure or that of Fair generate the intended results. We attempt to resolve these issues in Chapter 5 and suggest an alternative procedure.

2.7 Uniqueness, stability and terminal conditions

The Blanchard and Kahn solution shows that linear forward expectations models only have a unique stable solution if the state–space form (2.2) has as many unstable eigenvalues (modulus greater than unity) as there are non–predetermined (expectations) variables (proposition 1, p1308; proof, Appendix pp1310–11). Buiter (1984) derives the same result for the state–space form in continuous time. Gourieroux, Laffont and Montfort (1982) develop the same condition by examining all the solutions of a small theoretical example. Sargent (1979, Chapter 8) and Minford *et al.* (1980) also obtain the condition by examining the general and specific solutions to systems of difference equations. Fisher, Holly and Hughes Hallett (1985, appendix) generate the condition by examining the sensitivity of the solution to the terminal condition when the number of time periods is large. All of these results apply to linear systems only. In a nonlinear system, the corresponding analysis can only be applied to a local linearisation of the model. However, the same possibility of many stable solutions or none carries over to the nonlinear case.

The existence of multiple equilibria in rational expectations models has been recognised by many authors prior to Blanchard and Kahn. Black (1974), Taylor (1977), Calvo (1979), Burmeister (1980) and Gourieroux, Laffont and Montfort (1982) have all considered the implications of these multiple equilibria. Taylor's (1977) stochastic model yielded multiple stable equilibria and he recommended selecting the path with minimum variance. Gourieroux *et al.* considered a complete set of solutions for a general difference equation. Burmeister considers some of the philosophical implications of multiple equilibria. A view emerged (see e.g. Begg,

1982) that multiple solutions were not a problem as long as only one is stable. Economic systems will usually have such a saddlepoint property by construction and the economy will have an incentive to locate itself on the saddlepoint path. If this view is correct then the only problem for solving a large–scale model is to locate the saddlepoint solution. The problem is given a physical dimension by the choice of terminal condition. In solving the model for y_t, $t=s+1,\ldots,T$ we need to specify a value for $y_{T+1|s}$ since it will not be provided by the model. Different choices for this terminal value will give different dynamic trajectories and thus we have multiple solutions. If the system has a saddlepoint path then the choice of terminal condition problem revolves around locating that path (or at least its first T–s periods).

The choice of terminal condition appears to have had little consideration in the literature. We can identify two general approaches. Minford *et al.* (1979) suggest using terminal conditions obtained from an equilibrium analysis of the model. Over a finite time period, these equilibrium conditions then act in a similar way to the transversality conditions in some models of optimizing behaviour in economic theory. Transversality conditions ensure that behaviour is optimal even at the boundary points. This approach is also taken by Beenstock and Holly (1980).

The second general approach is that of Fair and Taylor (1983). Their solution algorithm, called the "Extended Path" makes arbitrary choices of terminal condition for each dynamic solution and then re–calculates the solution successively over longer time periods. When the solution period of interest is insensitive to the terminal date the process stops. This view of terminal conditions is essentially that they do not matter as long as they do not affect the solution values.

Wallis *et al.* (1985, 1986) consider the sensitivity of several models to different types of terminal condition. They conclude that terminal conditions should be of a form that characterises the unique stable path and that the choice of terminal condition can be tested against a longer solution horison to see if the conditions correctly select that path. This approach is essentially a synthesis of the two views outlined above. There remain several issues to be resolved concerning the choice of terminal conditions, particularly in cases where there is no unique saddlepoint trajectory.

2.8 Time inconsistency

Before applying control techniques or conducting policy analysis with models containing forward expectations, we first address the issue of time-inconsistency. An "optimal" policy given by a series of decision variables x_t, $t=1,\ldots,T$, is time-inconsistent if the value of x_{t+s} (say x^*_{t+s}) planned as optimal in period t for execution in period t+s, is not the same as the optimal value x^*_{t+s} calculated when re-optimizing in period t+s-j, s>j≥0 for some t,j,s *even when there is no uncertainty*. That is to say the mere passage of time makes a previously "optimal" plan sub-optimal when viewed from a later date. The policy which is time-inconsistent contravenes Bellman's (1957) principle of optimality:

> "An optimal policy has the property that, whatever the initial state and the initial decision are, the remaining decisions must constitute an optimal policy with regard to the state resulting from the first decision."

It is thus possible to argue that an optimal policy cannot be time-inconsistent. In what follows we define an optimal policy to be that set of policy values or policy rule which minimizes an explicit objective function under stated assumptions regardless of whether it generates potential time inconsistency.

To illustrate time-inconsistency we use a simple example which follows the presentation of several authors (e.g. Levine and Holly, 1987). Let us consider the simplest case T=2, s=1, j=0. We need to define an optimality criterion and we use the general form of welfare function:

$$\max_{x_1, x_2} S(x_1, x_2, y_1, y_2). \qquad (2.22)$$

The decision (policy) variable x is chosen for each period and we assume that the output variable y depends on its own lagged values and lagged, current and future policy variable values:

$$y_1 = f_1(y_0, x_0, x_1, x_{2|1}), \quad y_{1|2} = y_1, \tag{2.23a}$$
$$y_{2|1} = f_{2|1}(y_0, y_1, x_0, x_1, x_{2|1}), \tag{2.23b}$$
$$y_2 = f_2(y_0, y_1, x_0, x_1, x_2). \tag{2.23c}$$

The usual formulation of this problem ignores y_0 and x_0 which are treated as known constants and we henceforth drop these terms. Expectation terms are denoted as e.g $y_{2|1}$ and hence the second equation (2.23b) represents the outcome for period 2 forecast in period 1. We assume that the information set includes all lagged terms and current exogenous variables. We further assume that $S(\cdot)$ and $f_i(\cdot)$ are differentiable, that $S(\cdot)$ is concave and that there exists a solution to each equation in the system (2.23) for arbitrary values of x_1 and x_2. We start by considering the optimal values of x_1, x_2 planned at the beginning of period 1 and the first–order conditions are given by differentiating (2.22) given the model constituted by (2.23a) and (2.23b):

$$\frac{\partial S}{\partial y_{2|1}} \frac{\partial f_{2|1}}{\partial x_1} + \frac{\partial S}{\partial x_1} + \frac{\partial f_1}{\partial x_1} \left[\frac{\partial S}{\partial y_1} + \frac{\partial S}{\partial y_{2|1}} \frac{\partial f_{2|1}}{\partial y_1} \right] = 0 \tag{2.24}$$

$$\frac{\partial S}{\partial y_{2|1}} \frac{\partial f_{2|1}}{\partial x_{2|1}} + \frac{\partial S}{\partial x_{2|1}} + \frac{\partial f_1}{\partial x_{2|1}} \left[\frac{\partial S}{\partial y_1} + \frac{\partial S}{\partial y_{2|1}} \frac{\partial f_{2|1}}{\partial y_1} \right] = 0$$

The solutions to these equations will maximise the expected value of S with respect to $x_1, x_{2|1}$ viewed from period 1. However if we now re–optimise our plan for period 2 at the beginning of period 2, then the values of x_1, y_1 are in the information set and are fixed. We thus have the first order conditions when the model is (2.23c):

$$\frac{\partial S}{\partial y_2}\frac{\partial f_2}{\partial x_2} + \frac{\partial S}{\partial x_2} = 0 \tag{2.25}$$

This change in the first order conditions reflects the fact that the last term in equation (2.24) is a function of derivatives taken with respect to y_1 which is now a fixed value and so that term drops out. These two first–order conditions will generally lead to different solutions for x_2 when optimizing from periods 1 and 2 and hence time–inconsistency. Sufficient conditions for consistency are either that

$$\frac{\partial f_1}{\partial x_{2|1}} = 0 \text{ or } \left[\frac{\partial S}{\partial y_1} + \frac{\partial S}{\partial y_{2|1}}\frac{\partial f_{2|1}}{\partial y_1}\right] = 0\,. \tag{2.26}$$

The first of these conditions implies that the outcome in period one is independent of what happens in period 2 i.e. the model is not forward–looking. The second condition requires that the direct effects of y_1 on S are offset by its indirect effect through y_2 and hence y_1 does not affect S. The second of these conditions is unlikely to hold either by design or chance and is generally neglected.

The potential existence of time–inconsistency in economic optimization has been recognized at least since Strotz (1956). It is important to stress that time–inconsistency can arise even when there are no random shocks and there is no uncertainty about the model and the initial policy is carried out as planned. Strotz suggested that there were three possible responses to a potentially time–inconsistent situation. First, agents may, and often do, engage in time–inconsistent behaviour. Strotz argues that this is irrational because the incentive to change plans can be forecast and hence agents are acting in period 1 as if they expect themselves to act in period 2 in a way in which they can predict they will not i.e. their expectations are predictably incorrect. If such behaviour is realised then (as pointed out by Levine and Holly) it will actually be the globally optimal solution yielding a maximum of the welfare function. Secondly, agents may pre–commit themselves to avoid changing plans at a later date. This solution will generally be the optimal policy amongst all rational policies and so there is an incentive for rational agents to undertake contracts of this kind. Finally, the agents may recognize the existence of time–inconsistency and re–optimise taking this into account i.e. optimise over the set of those time–consistent strategies which do not involve contracting. This is done using dynamic programming in the following manner for our example. First we find the optimal value for period 2 viewed from period 2. Conditioning on this value we then find the optimal value for x_1. This description is only applicable to models which are not state dependent e.g. linear systems. In general nonlinear systems, the optimal values must be found simultaneously but with each period satisfying the appropriate first–order conditions.

Kydland and Prescott (1977) use the concept of time–inconsistency to challenge the use of control theory for dynamic economic planning when expectations are rational. In particular, they proposed a classic macroeconomic example which we shall find useful in explaining some model properties and which ties in our earlier discussion of the expectations augmented Phillips curve.

The Kydland–Prescott model is presented in a single period context and so is not strictly consistent with the two–period general formulation given above (which they also use). However, by treating period t as the second period of a two–period model and concentrating on the second period we can use the same first–order conditions. The endogenous variable y is unemployment. The policy variable x is the inflation rate, assuming that this can be chosen using monetary policy. The function $f(\cdot)$ is a supply curve (or Phillips curve) which for period t can be written as:

$$y_t = \psi(x_t - x_{t|t-1}) + \bar{y} \tag{2.27}$$

where $\bar{y}$ is the natural rate of unemployment and $x_{t|t-1}$ is expected inflation. The coefficient ψ is positive so that unemployment is inversely related to inflationary surprises. In the new–classical model this relationship arises because the real wage is unexpectedly reduced when inflation surprises occur and so more workers are employed.

Kydland and Prescott then define an objective function which is separable over time $S_t = s(x_t, y_t)$ and we illustrate by choosing:

$$S_t = -1/2\,[x_t^2 + (\bar{y}-y_t)]\,. \qquad (2.28)$$

If inflation announced in period t–1 for period t is anticipated, then the optimal inflation rate viewed from period t–1 for period t is zero and so we start by assuming that this is the announced policy and that the announced policy is carried out. In this case our optimization yields a "no–loss" outcome:

(i) "no–loss" policy

$$x_t = x_{t|t-1} = 0\ ;\ y_t = \bar{y}\ ;\ S_t = 0. \qquad (2.29)$$

However if we re–optimise in period t treating expectations as fixed then we find the optimal value of inflation is no longer zero hence solution (i) is time–inconsistent. Since this new solution involves going back on the announced policy it is called the cheating or reneging solution:

(ii) Optimal cheating solution

$$x_t = \psi/2\ ;\ x_{t|t-1} = 0\ ;\ y_t = \bar{y} + \psi^2/2\ ;\ S_t = \psi^2/8. \qquad (2.30)$$

However this solution is irrational because agents, who understand how the policy–maker optimizes, can predict that their forecasts will be wrong. If the plan was achieved it would yield positive gains of $\psi^2/4$ on unemployment against losses of $\psi^2/8$ on inflation – an unambiguous gain of $\psi^2/8$. Note that this precise result is dependent on the particular form of model and objective function chosen.

This plan is optimal only from the policy–maker's view. If agents suffer any loss from the irrationality of expectations then they will react to the government's plan. At this stage we are introducing a two–player game but, following Kydland

and Prescott we do not specify the agent's objective function and simply assume that expectations become rational.

We now assume that agents set the expected inflation rate and the policy maker the actual inflation rate simultaneously. If agents predict that the optimal inflation rate in period t is $\psi/2$ then the solution is:

(iii) Rational, time–consistent solution

$$x_t = x_{t|t-1} = \psi/2\,;\; y_1 = \bar{y}\,;\; S_t = -\psi^2/8. \tag{2.31}$$

This strategy is time–consistent since the optimal strategy for period t viewed from period t–1 is now $\psi/2$ and that is also the optimal strategy as viewed from period t. The solution is also rational because expected inflation is a correct forecast. However this policy has no unemployment gains just inflation losses of $\psi^2/8$. A better policy would be that obtained by implementing the no–loss strategy (2.29). Since the incentive to cheat on strategy (2.29) is predictable it can be argued that it can never be achieved. If a government tries to play (2.29) but agents expect cheating to occur then we have a fourth solution:

(iv) Expected cheating solution

$$x_t = 0\,;\; x_{t|t-1} = \psi/2\,;\; y_t = \bar{y} + \psi^2/2\,;\; S = -\psi^2/4. \tag{2.32}$$

This strategy actually yields unemployment losses of $\psi^2/4$ and is clearly sub–optimal compared with the time–consistent rational strategy (2.31). Hence there is an incentive for the government to play (2.31) to avoid (2.32) whilst (2.29) and (2.30) are unobtainable. If we assume that (2.28) represents the agents welfare function as well as the governments then we have outlined a standard prisoners dilemma problem. Both agents and government would prefer the outcome under (2.29) but the actual outcome is (2.31).

There is now an extensive branch of the literature examining conditions under which the no–loss solution can in fact be achieved and evaluation of the losses involved for the various solutions using different specifications. Barro and Gordon (1983a,b) extended the time–horizon of the Kydland–Prescott model and used a multi–period loss function. They then investigated the possibility of a policy–maker establishing credibility or reputation so that agents believed announcements of "no–loss" strategies. This approach to gaining credibility requires the specification of a learning mechanism and implies losses whilst agents are learning as hinted by solution (2.34). Barro and Gordon also consider the temptation to cheat and suggest punishment strategies to ensure that cheating is too costly (e.g agents revert to non–zero inflationary expectations). Development of these approaches has led to substantial consideration of how agents form expectations of policy variables. This is a different problem from that of forward expectations addressed earlier since we now have, in general, no specific model to generate the expectations terms. Since there may be more than one possible scheme for setting policy we are in effect introducing uncertainty about the model form.

To analyse the various scenarios that might be of interest, the economics literature has drawn from and contributed to the theory of games (a comprehensive treatment of game theory is given, *inter alia*, by Basar and Olsder (1982)). The general solutions for dynamic games in linear, forward expectation, continuous time models are given by Miller and Salmon (1985) and Cohen and Michel (1988). Particular games have been proposed and solved by other authors. Backus and Driffill (1985a,b) introduce a loss function explicitly for the private sector and two possible types of government with attendant probabilities and this constitutes a game to be played between agents and governments. The optimal solution to this game can then be obtained as a function of those probabilities. Barro (1986) considers games in which the private sector is atomistic and hence cannot form a particular joint view. Only the government can then set strategies. Barro goes on to show that reputational forces in this case can sustain the no–loss policy at least for a while. The ability to sustain the no–loss policy is the main conclusion from the Barro/ Backus and Driffill papers.

Canzoneri (1985) analyzes the case of asymmetric information so that only governments know if they have cheated and shows that this can lead to periodic bouts of inflation. Vickers (1986) generalizes this approach to different types of government by varying their objective functions. Vickers argues that reputational forces sustaining the Barro (and Backus and Driffill) solutions may not be present for certain possible types of government. Driffill (1987) re–establishes the proposition by allowing many types of government and allowing exogenous noise. Levine and Holly (1987) survey these various models and give extensions to the general stochastic, dynamic model. Other aspects of the time inconsistency literature have been surveyed by Blackburn (1987).

2.9 Optimal Control and Policy Analysis

In considering the optimal control and possible time–inconsistency of a large–scale nonlinear model, the literature is considerably thinner. One approach suggested has been that of linearizing the model so that optimization techniques can be implemented on a reduced version (e.g. Gaines *et al.* 1987). This enables one to apply all the analytical solutions obtained for linear systems. Alternatively, we need to consider direct optimization of the nonlinear system under consistent and non–consistent expectations assumptions. Not only do we need numerical algorithms but these must be carefully constructed so as to give us the particular form of solution required. Holly and Zarrop (1983) propose a penalty function approach for optimizing an objective function subject to a nonlinear forward expectations model (also described in Levine and Holly, 1987, and Holly and Hughes Hallett, 1989). This paper has recently been criticised by Wohltmann and Krömer (1989) on the grounds that the definitions of time–consistency and time–inconsistency used by Holly and Zarrop are contrary to those used by Kydland and Prescott. Hall (1984, 1986) has suggested a method of obtaining a time–consistent expectations– consistent strategy which he applied to a special version of the NIESR model. These two approaches will be examined in depth in Chapter 7 and will be compared to the procedures used by Wallis *et al.* (1987, Ch.3) and that suggested by Fair (1984, pp385–386). The solutions produced by some of these algorithms do not appear to have been well understood.

Finally, the analysis provided by the time–inconsistency literature focuses attention on the expectations of the "exogenous" variables in the model. The implications for standard simulations have been addressed by Wallis *et al.* (1986, Ch.2) and by Cooper (1987). The issues are resolved by examining the sensitivity of the simulation results to different assumptions and the approach of Wallis *et al.* is extended in Chapter 5.

2.10 Summary and conclusions

In this chapter we have surveyed a wide range of literature covering large–scale macroeconomic models, simulation, optimal control, and forward expectations. The presence of forward expectations in models is seen to complicate all of the procedures for the use and analysis of large–scale models. For linear models, the literature generally contains methods for resolving the problems introduced by forward–expectations. For nonlinear models, the literature contains something of a thin patchwork of partial solutions to the various problems. In some cases, such as terminal conditions, the literature does not appear to be complete. In other cases, such as optimal control or stochastic simulation, there are doubts as to whether the available procedures generate the appropriate results.

Chapter 3

SOLUTION METHODS FOR NONLINEAR FORWARD EXPECTATIONS MODELS

In Chapter 2 we briefly considered four solution algorithms for solving forward expectations models. The analysis of Blanchard and Kahn (1980) depends on solving an eigenvalue problem for a system of linear difference equations, and exact analytical results can be obtained. For a nonlinear model the eigenvalue problem is not defined, hence we could only apply this approach to a linearisation of the nonlinear system. However, the linearization itself will only be an approximation to the system of interest and so the solution is ultimately not exact. Furthermore, the linearization is only valid in the neighbourhood of some base trajectory around which the model is linearised (Kuh, Neese and Hollinger, 1985) and this imposes limits on the use of the linearized form.

In this chapter, we examine the feasibility and relative efficiency of the other three algorithms considered in Chapter 2, all of which can be used to solve the nonlinear model directly. We begin by presenting a detailed analysis of solving equation systems by first–order iterative methods, which can be used as a basic tool in each of the three approaches. The first approach for solving forward expectations models is then a generalization of first–order iterative methods applied simultaneously to all time periods in the solution. The generalization allows us to take advantage of the dynamic structure of systems which contain forward expectations. This approach is developed from that of Fisher, Holly and Hughes Hallett (1985, 1986) and Fisher and Hughes Hallett (1988a). A family of algorithms are presented which encompass those of Fair and Taylor (1983) and Hall (1985). Versions of these algorithms are then tested on three large–scale nonlinear macroeconometric models of the U.K. economy.

As a second approach we examine the penalty function methods of Holly and Zarrop (1983). This method can be viewed as a derivative–based Newton's method version of the first–order methods. We apply the penalty function technique to the same three models and the relative costs of the two algorithms are compared.

Finally we consider the shooting methods proposed for economic models by Lipton *et al.* (1982) and Spencer (1985). A generalization of these methods is proposed for a wider class of models including those with forward expectations of more than one period ahead. Despite these generalizations, shooting methods do not appear to be appropriate for solving the models since they involve solving the model in an inherently dynamically unstable form. Some comparative analysis is given based on the implied equation orderings of the different algorithms.

3.1 First order iterative solution techniques

In this section we consider general first–order iterative techniques for solving equation systems. We begin with a linear representation and then generalize to the nonlinear computational forms. Consider the linear equation system

$$By = b, \tag{3.1}$$

where B is a known real matrix of order n with nonvanishing diagonal elements; y and b are real vectors containing the unknown and predetermined parts of each equation respectively. Stationary first–order iterative techniques are derived by various decompositions of B in the form P–Q where P is non–singular. We can then develop first–order iterations by writing:

$$\begin{aligned} (P-Q)y &= b, \\ Py &= Q\,y + b, \\ y &= P^{-1}Q\,y + P^{-1}b. \end{aligned} \tag{3.2}$$

This last equality allows us to obtain the following iteration:

$$y^{(s+1)} = Gy^{(s)} + c \qquad s=0,1,2,\ldots \tag{3.3}$$

with an arbitrary start $y^{(0)}$, where $G=P^{-1}Q$ and $c=P^{-1}b$ define the iteration matrix and forcing function respectively (Young, 1971). A solution to (3.3) is defined as convergence in the sequence $y^{(s)}$ such that $|(y_i^{(s+1)}-y_i^{(s)})/y_i^{(s)}| < \tau$ for each element y_i where τ is some given tolerance level. These methods are routinely used to construct the numerical solution to (3.1) (Hageman and Young (1981), Varga (1962)). They are computationally efficient if B is a large, sparse or ill–conditioned matrix. They are also widely used for solving nonlinear equation systems, in which case B represents the system's Jacobian matrix.

Different decompositions of B into P–Q give different first order methods. The most widely used decompositions are those which produce the Jacobi, Gauss–Seidel and Successive Over–relaxation (SOR) iterative methods; respectively

$$P = I\ ;\ P = (I-L)\ \text{ and } P = \frac{1}{\alpha}(I-\alpha L), \tag{3.4}$$

where B is normalized to have a unit main diagonal and L is a matrix of order n containing the lower triangle of B (below the main diagonal) and zeros elsewhere, and α is a relaxation parameter to be chosen. We also define $U=B-I-L$ to be the upper triangle.

The rate of convergence in (3.3) is often significantly increased by the one parameter extrapolation

$$y^{(s+1)} = \gamma(Gy^{(s)}+c) + (1-\gamma)y^{(s)} = Hy^{(s)} + \gamma c. \tag{3.5}$$

Two particular versions of (3.5) are routinely used; the Jacobi over–relaxation (JOR) method with $G=(I-B)$, and the Fast Gauss–Seidel (FGS) method based on $G=(I-\alpha L)^{-1}(\alpha U+[1-\alpha]I)$.

It is well known that (3.3) converges if and only if $\rho(G)<1$, where $\rho(\cdot)$ denotes spectral radius (defined as the modulus of the largest eigenvalue). Proof: Young (1971), Theorem 5.1. Similarly (3.5) converges for some $\gamma>0$ if the real parts of the eigenvalues of G are all less than unity. Proof: Hughes Hallett (1981).

If $\rho(H)<1$, the number of steps to convergence in (3.5) is approximately $\log \tau/\log \rho(H)$. The corresponding speed of convergence can be measured either as an asymptotic rate $(-\log \rho(H))$ or as an average rate $(1/s)(\log ||H^{(s)}||)$ for some norm. We therefore aim to minimize $\rho(H)$ to ensure convergence as well as to minimize the number of steps. Optimal values for γ have been derived and generalisations of (3.5) to second order methods and multi–parameter extrapolations have also been made. These extensions will not be needed here and are surveyed in Fisher and Hughes Hallett (1988b).

There are few general results on convergence conditions for SOR iterations and the only general result is that $\alpha>0$ exists such that SOR is convergent if (I–B) has all its roots less than unity in real part. Proof: Hughes Hallett (1986). Other results require special conditions on (3.1) (e.g. symmetric, positive definite) which are not appropriate for an econometric model whose only restrictions are that (I–B) is real and (usually) non–singular. Hence FGS has proved to be a more helpful device for accelerating convergence (Fisher and Hughes Hallett, 1987). SOR tends to be used only when other methods fail.

3.2 Nonlinear models

We now consider the nonlinear model

$$y_j = g_j(y, z), \quad j=1,\dots,n \tag{3.6}$$

where z is a vector of predetermined elements (exogenous and lagged endogenous variables) and $g_j(\cdot)$ is a general, real–valued, nonlinear function. The JOR and SOR iterations for solving (3.6) are now

$$y_j^{(s+1)} = \gamma g_j(y_1^{(s)},\dots,y_n^{(s)}) + (1-\gamma)y_j^{(s)} \tag{3.7a}$$

and

$$y_j^{(s+1)} = \alpha g_j(y_1^{(s+1)},\dots,y_{j-1}^{(s+1)},y_{j+1}^{(s)},\dots,y_n^{(s)}) + (1-\alpha)y_j^{(s)} \tag{3.7b}$$

Meanwhile the FGS method extrapolates the SOR iterates, i.e.

$$y_j^{(s+\frac{1}{2})} = \alpha g_j(y_1^{(s+\frac{1}{2})},\ldots,y_{j-1}^{(s+\frac{1}{2})},y_{j+1}^{(s)},\ldots,y_n^{(s)}) + (1-\alpha)y_j^{(s)}$$

with

$$y_j^{(s+1)} = \gamma y_j^{(s+\frac{1}{2})} + (1-\gamma)y_j^{(s)}. \tag{3.8}$$

These methods give us the nonstationary iteration:

$$y^{(s+1)} = G^{(s)}y^{(s)} + k \tag{3.9a}$$

$$y^{(s+1)} = H^{(s)}y^{(s)} + \gamma k \tag{3.9b}$$

respectively for equations (3.7) and (3.8) where k is a separable constant dependent only on pre–determined values and $G^{(s)}$, $H^{(s)}$ are the iteration matrices defined at iteration s. The convergence results now apply locally to $G^{(s)}$, $H^{(s)}$ as appropriate.

3.3 Forward expectations models

Consider a model with forward (rational) expectations terms:

$$B_0y_t = B_1y_{t-1} + A_1y_{t+1|t-1} + u_t \tag{3.10}$$

where $y_{t+j|s} = E(y_{t+j}|\Omega_s)$, for j>0, is the expectation of y_{t+j} conditional on the information available at the end of period s denoted Ω_s; and where u_t represents all exogenous variables and random disturbances. For consistent expectations we require the forward expectation terms $y_{t+1|t-1}$ to be the same as the next period's forecast value obtained by solving the model conditional on the same information set. Hence the solutions are linked forward in time and to solve the system (3.10) for each y_t conditional on some start period 0 requires each $y_{t+j|0}$, for j=1,2,...,T–t; and a terminal condition $y_{T+1|0}$. Stacking this system up over time implies, for a consistent expectations solution:

$$\begin{bmatrix} B_0 & -A_1 & & 0 \\ -B_1 & \ddots & \ddots & \\ & \ddots & \ddots & -A_1 \\ 0 & & -B_1 & B_0 \end{bmatrix} \begin{bmatrix} y_1 \\ \vdots \\ \\ y_T \end{bmatrix} = \begin{bmatrix} u_1 \\ \vdots \\ \\ u_T \end{bmatrix} + \begin{bmatrix} B_1 \\ 0 \\ \vdots \\ 0 \end{bmatrix} y_0 + \begin{bmatrix} 0 \\ \vdots \\ 0 \\ A_1 \end{bmatrix} y_{T+1} \qquad (3.11)$$

where u_t is taken from Ω_0. The information set therefore contains either projections or known values for all exogenous variables and disturbance terms. This block triangular form is similar to the band matrices used for optimisation by Theil (1964). The solution to equation system (3.11) is given by

$$Y = \bar{B}^{-1}\bar{b} \qquad (3.12)$$

where $\bar{B}$ is the block tridiagonal matrix on the left of equation (3.11); Y is the stacked vector of endogenous variables conditioned on Ω_0; and $\bar{b}$ is the stacked vector of terms on the right of (3.11). A necessary condition for the existence of a solution to equation (3.12) is that $\bar{B}^{-1}$ exists. This in turn requires that neither B_0 nor $B_1B_0^{-1}A_1B_0$ have unit roots.

The solution of a dynamic rational expectations model therefore has exactly the same form as that for a conventional equation system. Any of the standard methods for solving equation systems can be applied to equation (3.11) including derivative–based (e.g. Newton–Raphson) or first–order iterative methods. The differences are only that:

(i) the unknowns of different time periods have to be determined simultaneously rather than recursively and hence:

(ii) the matrix B has been replaced by the block tridiagonal matrix $\bar{B}$;

(iii) the variables in $\bar{b}$ are augmented by the terminal conditions y_{T+1}.

Therefore we can continue to use simple first order iterative techniques as a cheap way of constructing numerical solutions to rational expectations models.

The ordering of the elements in y_t generally has no special significance, but the equations in equation (3.11) are ordered by time periods. In conventional models, multi–period solutions are generated sequentially forwards because that exploits the block recursive structure of equation (3.11) when $A_1=0$. Consequently the only relevant decompositions of the system are those defined for B_0 and we solve for each time period as a sub–system. But when $A_1 \neq 0$ that block recursive structure is lost and to generate first–order methods for solving equation (3.12) we consider decompositions of $\bar{B}$ over the whole of the time horizon.

The aim of these methods is to see if we can split out of $\bar{B}$ a small number of terms, leaving behind a structure which is relatively simple to solve. A two–part iteration between those variables split out and the remainder may then speed up convergence. For conventional equation systems this type of splitting is usually found to be highly beneficial (see e.g. Don and Gallo, 1987; Hughes Hallett and Fisher, 1987). The two main possibilities for solving (3.11) are:

(a) Decompositions of $\bar{B}$ without regard to its block structure i.e. we could treat (3.11) as one large equation system and use the decompositions given at (3.4) defined for $\bar{B}$ rather than B_0. These decompositions define a family of simple first–order iterations on the stacked system (3.11). Hence any first–order solution technique can be applied in the form of those for non–forward–expectations models. This method is most suitable when there are a large number of expectations terms and lagged terms and hence no efficient splittings.

(b) A decomposition of $\bar{B}$ into those blocks above the main block diagonal i.e. the expectation terms, and those on or below the block diagonal. The lower triangle can then be solved as an ordinary dynamic solution, holding expectations fixed. Different decompositions of B_0 then generate JOR, SOR and FGS iterations as in the conventional case. The upper block triangle is then solved by equating the expectation terms to the solution from the lower block triangle. These two decompositions then define a two–part iterative scheme: an inner iteration loop which solves for the current and lagged variables of each period sequentially, and an outer iteration loop which updates the forward expectations terms. Convergence on the outer loop determines that a consistent expectations solution has been found. This scheme is likely to be efficient when there are relatively few expectations terms.

Other decompositions are also possible. One could separate out both the upper triangle and the lower triangle into the outer loop. Thus the inner loop would solve only for the current dated terms holding lagged values and expectations fixed. The outer loop then updates both lead and lagged terms. This approach would be suitable for multi–processor computers which could evaluate all the time–periods simultaneously. Alternatively one could split off the lower block triangle into the outer loop on its own i.e. type (a) applied to the transverse of $\bar{B}$. In this case the inner loop would solve backwards through time solving for current dated and expectations terms. The outer loop would then set lag values by the updating scheme. This scheme would work well if the upper block triangle was relatively dense compared with the lower block triangle. Finally, one could mix up the decompositions so as to make multi–part iterations passing through different sub–block combinations in different sequences.

In most large–scale models there are far more lag terms than forward expectations. Furthermore the initial conditions (y_0 in 3.11) are usually known whereas terminal values (y_{T+1}) are often generated by the iterative process (see Chapter 4). These considerations usually mean that only the first two types of decomposition are relevant. However it is possible to generate further algorithms based on these decompositions. We define a third approach as follows.

(c) Type (b) splittings in which the inner iterations are not taken to full convergence at each step. This can be achieved by setting a substantially weaker convergence criterion on the inner loop. Computational savings are made because effort is not wasted in getting a full conditional inner loop solution which is then going to be changed again by the values generated in the next outer loop step.

Apart from the introduction of this family of methods by Fisher, Holly and Hughes Hallett (1985, 1986), two special cases of these schemes have already appeared in the literature. The Fair–Taylor method (Fair and Taylor, 1983) is a type (b) scheme and the inner loop here corresponds to their Type I iterations and the outer loop to their Type II iterations. The method suggested by Hall (1985) is a type (a) scheme i.e. a single inner loop pass on the whole equation system. These two methods can also be viewed as extreme type (c) methods for a relatively strong or weak inner loop convergence criterion respectively.

In Section 3.6 we note that the penalty function method (Holly and Zarrop, 1983) is a type (b) with the outer loop step solved by a derivative–based method. The multiple shooting method (Lipton *et al.*, 1982) will be written as a re–ordered version of a type (b) scheme. All of these schemes may use first–order methods (or Newton's method) for solving the inner loop. The outer loop steps may also be extrapolated in the same way as equation (3.5). Only type (b) and (c) methods allow different extrapolations for the inner and outer loop steps.

3.4 The Iterative Schemes

In this section we give the computational forms for each of the three general schemes suggested above. We begin by using a linear analysis which enables us to comment on the convergence conditions. In practice the models of interest are nonlinear and we evaluate convergence properties by testing on a selection of models.

First–order methods applied to the stacked system (3.11) require the decomposition $\tilde{B}=\tilde{P}-\tilde{Q}$ and hence an iteration matrix $\tilde{G}=\tilde{P}^{-1}\tilde{Q}$, and forcing function $\tilde{c}=\tilde{P}^{-1}\tilde{b}$. In order to accommodate the two part iterations, types (b) and (c), it will be necessary to define decompositions of $\tilde{B}$ separately across equations within a time period and across time blocks. Let $B_0=I-L-U$ where L and U are the lower and upper triangles defined for B_0 as for B in Section 3.2. The blockwise decomposition of $\tilde{B}$ is given by $\tilde{B}=D^*-L^*-U^*$ where $D^*=I_T\otimes B_0$ and U and L are upper and lower block triangles:

$$L^* = \begin{vmatrix} 0 & & 0 \\ B_1 & \ddots & \\ 0 & B_1 & 0 \end{vmatrix} \quad \text{and} \quad U^* = \begin{vmatrix} 0 & C_1 & 0 \\ & \ddots & C_1 \\ 0 & & 0 \end{vmatrix}$$

Finally we need the corresponding decomposition $\bar{B}=\bar{D}-\bar{L}-\bar{U}$ where

$$\bar{L} = \begin{vmatrix} L & & 0 \\ B_1 & \ddots & \\ 0 & B_1 & L \end{vmatrix} \quad \text{and} \quad \bar{U} = \begin{vmatrix} U & C_1 & 0 \\ & \ddots & C_1 \\ 0 & & U \end{vmatrix}$$

and $\bar{D}=I_{nT}$. Now in line with our schemes of section 3.3 we have:

(i) Type (a) iterative methods

This family of system wide iterations is based on the last of these decompositions: $\bar{B}=\bar{D}-\bar{L}-\bar{U}$. Thus the basic Jacobi, Gauss–Seidel and SOR methods applied throughout the system have iteration matrices:

$$\bar{G} = \bar{L}+\bar{U},\ (I-\bar{L})^{-1}\bar{U}, \text{ or } (I-\alpha\bar{L})^{-1}(\alpha\bar{U}+[1-\alpha]I) \text{ respectively.} \tag{3.13}$$

The JOR or FGS extrapolations imply iteration matrices $\bar{H}=\gamma_0\bar{G}+(1-\gamma_0)I$.

The computational forms of Jacobi and SOR iterations are given by stacking up respectively:

$$y_t^{(s+1)} = \left[B_1 y_{t-1}^{(s)} + (L+U)y_t^{(s)} + A_1 y_{t+1}^{(s)} + u_t \right] \tag{3.14a}$$

$$y_t^{(s+1)} = \alpha\left[B_1 y_{t-1}^{(s+1)} + Ly_t^{(s+1)} + Uy_t^{(s)} + A_1 y_{t+1}^{(s)} + u_t \right] + (1-\alpha)y_t^{(s)}. \tag{3.14b}$$

over the solution period $t=1,\ldots,T$. But in order to take full advantage of the spareness of $\bar{B}$ (blockwise and within its sub–matrices), the iterations would actually be performed equation by equation in the model's (possibly nonlinear) structural form. For example, the SOR iteration for the linear model would be (with summations over j):

$$y_{it}^{(s+1)} = \alpha \left[\sum_{1}^{n} B_{1ij} y_{j,t-j}^{(s+1)} + \sum_{1}^{i-1} L_{ij} y_{jt}^{(s+1)} + \sum_{i+1}^{n} U_{ij} y_{jt}^{(s)} \right. \tag{3.15}$$

$$\left. + \sum_{1}^{n} A_{1ij} y_{j,t+1}^{(s)} + u_{it} \right] + (1-\alpha) y_{it}^{(s)}$$

for $i=1,\dots,n$ and each $t=1,\dots,T$ in turn for each iteration. Convergence depends on $\rho(\bar{G})<1$, where $\bar{G}$ is the final matrix in equation (3.13) ($\rho(\bar{H})<1$ for JOR, FGS).

(ii) Type (b) iterative methods

This family of two part iterations can use several different decompositions. The inner loop is based on solving the system for each period based on $B_0=I-L-U$ and obtaining exactly the same iterations as used in Section 3.1 i.e.:

$$y_t^{(k)} = Gy_t^{(k-1)} + P^{-1} \left[B_1 y_{t-1}^{(*)} + C_1 y_{t+1}^{(s)} + u_t \right] \tag{3.16a}$$

for $k=1,2,\dots$ for a particular period t. Matrices G and P are any matching pair generated by the decomposition of B_0 and $y_{t-1}^{(*)}$ is the value obtained as the previous period's solution. Equation (3.16a) gives the Type I iterations of Fair and Taylor. Alternatively the inner loop iterations might be system wide:

$$y_t^{(k)} = Gy_t^{(k-1)} + P^{-1} \left[B_1 y_{t-1}^{(k)} + A_1 y_{t+1}^{(s)} + u_t \right] \tag{3.16b}$$

for $t=1,\dots,T$ and then $k=1,\dots$. Another version would replace $y_{t-1}^{(k)}$ by $y_{t-1}^{(k-1)}$ in (3.16b). In all these versions $y_{t+1}^{(s)}$ represents the value of the expectations variables at the last outer loop step and it is held fixed during the inner loop steps. Given the dynamic structure of most models which have a large number of lag terms, it seems sensible to use the first of these options (3.16a) unless we have vector processing available on the computer in which case these alternative forms may be more efficient since they would allow the same calculation to be made simultaneously for all time periods.

Stacking over $t=1,\ldots,T$ the inner loop iteration matrices for the system as a whole are given by $G_{(1)}=I_T \bullet G$ (G applied in each period) for (3.16a); and $G_{(2)}=(I_{nT}-(I_T \bullet P^{-1})L^*)(I_T \bullet G)$ or $G_{(3)}=I_T \bullet G+(I_T \bullet P^{-1})L^*$ for the two versions of (3.16b) respectively. Inner loop convergence depends in all three cases on $\rho(G)<1$ since $\rho(G_{(i)})=\rho(G)$ for $i=1,2,3$.

After inner loop convergence in all periods, we obtain the next outer loop step (the Type II iteration of Fair and Taylor) by inserting the converged inner loop values throughout the system e.g.:

$$y^{(s+1)} = (I_{nT}-G_{(3)})^{-1}(I \bullet P^{-1})(U^* y^{(s)}+u) \tag{3.17}$$

which implies an outer loop iteration matrix of $\bar{G}=(I_T \bullet B_0-L^*)^{-1}U^*$ based on a blockwise decomposition of $\bar{B}$. Once again JOR or FGS extrapolations can be applied to the inner or the outer loop iterations, or to both. Taking (3.16a) for simplicity, G will then be replaced by $H=\gamma_0 G+(1-\gamma_0)I$ and $\bar{G}$ by $\bar{H}=\gamma\bar{G}+(1-\gamma)I$ where $\gamma_0 \neq \gamma$ is entirely possible. Convergence therefore requires both G and $\bar{G}$ (respectively H and $\bar{H}$) to have spectral radii less than unity. Finally the efficient computational form for FGS based on (3.16a) would be

$$\begin{aligned} y_{i,t}^{(k+1/2)} = \alpha \Bigg[& \sum_{j=1}^{n} B_{1ij} y_{j,t-1}^{(*)} + \sum_{j=1}^{i-1} L_{ij} y_{j,t}^{(k+1/2)} + \sum_{j=i+1}^{n} U_{ij} y_{j,t}^{(k-1)} \\ & + \sum_{j=1}^{n} A_{1ij} y_{j,t+1}^{(s)} + u_{it} \Bigg] + (1-\alpha) y_{i,t|0}^{(k-1)} . \end{aligned} \tag{3.18}$$

Then given $y_t^{(k+1/2)}$, we set $y_t^{(k+1)}=\gamma_0 y_t^{(k+1/2)}+(1-\gamma_0)y_t^{(k)}$ for the current inner loop value. Finally, when the inner loop has converged for all t to $y^{(*)}$, the computational form of the outer loop step is

$$y^{(s+1)} = \gamma y^{(*)} + (1-\gamma)y^{(s)}. \tag{3.19}$$

(iii) Type (c) methods, incomplete inner iterations

If the inner iterations of a type (b) method are terminated after p steps, then equation (3.16a) becomes

$$y_t^{(p)} = G^p y_t^{(s)} + \sum_{i=0}^{p-1} G^i [P^{-1}(B_1 y_{t-1}^{(*)} + A_1 y_{t+1}^{(s)} + u_t)] \tag{3.20}$$

Setting $y_t^{(s+1)}=y_t^{(p)}$, for $t=1,...,T$; implies an outer loop iteration matrix of $\bar{G}=[I_{nT}-(I_T\otimes \sum_1^p G^i P^{-1})L^*]^{-1}[I_T\otimes G^p+(I_T\otimes \sum_1^p G^i P^{-1})U^*]$. Convergence obviously requires $\rho(\bar{G})<1$, and FGS extrapolations can easily be introduced both into (3.20) and by redefining $y^{(s+1)}=\gamma y^{(p)}+(1-\gamma)y^{(s)}$. Note however that setting p=1 in this G automatically recreates the iteration matrices given at equation (3.13) for the corresponding definition of the inner loop matrix G since $G=P^{-1}Q$ and $B_0=P-Q$. Similarly setting $p=\infty$, when $\rho(G)<1$ also holds, reproduces the iteration matrix of (3.17) since $(I-G)^{-1}P^{-1}=B_0$. Thus both type (a) iterations and type (b) iterations (e.g. Fair and Taylor, 1983) may be treated as special cases of the more general type (c) method. The increasing level of generality is reflected in wider opportunities for convergence, and the faster convergence speeds offered by the extra parameters which control these iterations. That is to say the most efficient type (c) methods cannot be slower since they incorporate type (a) and type (b) methods as extreme parameter choices. For a given model and tolerance level τ, only two parameters, α and γ_0, can be set in type (a) methods. In type (b) methods three parameters, α, γ_0 and γ can be chosen, whilst type (c) methods exploit four: α, γ_0, γ and p (or equivalently, a weaker inner loop convergence test).

Convergence conditions for these iterative techniques depend on $\bar{G}$ for type (a) methods, G and $\bar{G}$ for type (b) methods and on $\bar{G}$ following (3.20) for type (c) methods. If a type (a) or a type (b) method satisfies this test, then a corresponding type (c) method is automatically convergent for some choice of p. Thus convergence is always possible by one parameter extrapolations if $\bar{G}$ (and G if complete inner iterations are required) is chosen to have roots with real parts less than unity. Since JOR extrapolations take $G=(I-B_0)$ and FGS extrapolations $G=(I-\alpha L)^{-1}(U+(1-\alpha)I)$; $0<\alpha<2$; type (a) methods will be convergent provided $(D^{*}-\bar{B})$ has roots with real parts less than unity ($D^{*-1}\bar{B}$ has all its roots in the left half–plane). Similarly type (b) methods will be convergent provided $D^{-1}B$ and $(I*B^{-1})\bar{B}$ both have their roots in the left half–plane. Finally, the optimal FGS extrapolation parameters γ_0^{*} and γ^{*}, can be computed for each G and/or $\bar{G}$ matrix.

Unfortunately these convergence conditions are not of great use for nonlinear models other than to demonstrate the possibility of convergence. In a linear system one might be willing to calculate the eigenvalues of the relevant iteration matrices. In a nonlinear system the (non–stationary) iteration matrices are not usually computed and their eigenvalues would anyway be dependent on the iteration path. Convergence properties must therefore be evaluated by numerical simulation. The results given in the following section show that robust and efficient schemes can be implemented for all three models considered.

3.5 Empirical results

A type (b) method (Fair and Taylor iterations) and our preferred type (c) solution technique have been tested on three nonlinear econometric models of the United Kingdom economy which contain rational expectations terms: the LPL model, LBS model and the NIESR model. The models differ in the number of expectational terms and in overall model size: 40 forward expectations terms in the NIESR model (which has 199 endogenous variables), 6 in LPL (34 endogenous variables) and 3 in LBS (1258 endogenous variables) and hence each represents a quite different test for the algorithms. Furthermore, the expectation terms represent different lead lengths. The LBS terms are one quarter ahead, NIESR up to 4 quarters and LPL up to 5 years ahead. In each case we assume that only one stable feasible solution exists. This property is then examined by assessing the model's sensitivity to its terminal conditions in Chapter 4.

The iteration processes then depend on five parameters: the inner and outer loop convergence criteria; the internal SOR acceleration parameter (α_0); the inner loop extrapolation parameter (γ_0); and the expectations (or outer loop) extrapolation parameter (γ_1). For each model we examine the solution cost for a calculation of a standard government expenditure multiplier.

It is an observed feature of first–order iterative methods applied to an arbitrary equation system that different variables converge at different speeds. It is quite common to find that model convergence depends on one or two particularly slow variables. These are typically flow variables or identities such as the current account balance or the PSBR. Methods for restricting the convergence of the inner loop can take advantage of this feature. It is especially convenient if the variables of which expectations are formed are amongst the first to solve. This means that restricting the inner loop will not lead to a substantially larger number of outer loop iterations, guaranteeing a more efficient solution. In such cases the best way to limit the inner loop is to limit the convergence test to just those variables of which expectations are formed. If appropriate, this approach means that we do not necessarily need to search for either the most efficient number of p (the maximum number of iterations allowed) or τ_I (the most efficient weak inner loop convergence criterion) although an optimal setting of τ_I will still be of interest. In all methods, once expectations have converged, a further round with a comprehensive tight convergence criterion is used to confirm that the model has converged to the required global tolerance.

(i) The Liverpool model

In the Liverpool model, the variables for which expectations are formed converge relatively quickly in each time period. Applying the inner loop convergence criterion to these variables alone is the most effective method of reducing the number of iterations on the inner loop. Our fastest search method starts by locating an efficient pairing of α and γ_0 using a fixed initial value for the expectations and the limited convergence test on the inner loop. Fisher and Hughes Hallett (1987) showed that optimal values for α and γ_0 are related and the optimal pairing can usually be found by a grid search in less than six steps. Having obtained these values, we then search for the optimal outer loop extrapolation parameter and check the resulting combination for efficiency with respect to each parameter, to find the fastest possible solution. A complete set of results is displayed in Table 3.1.

A Fair–Taylor type (b) solution, with iteration parameters optimized at α_0=0.65, γ_0=1.0, γ_1=0.6 and full convergence criterion τ_I=.0002 (.02%), solves in a total of 1324 iterations for the 14 periods when subjected to a government expenditure shock. Incomplete inner iterations (controlled by a convergence test of τ_I=.005 (0.5%) on those variables having forward expectations in the model) reduces that total to just 827 iterations with the same parameter values. Finally, optimizing all parameters together yields $(\alpha_0,\gamma_0,\gamma_1)$=(0.7,0.85,0.55) and reduces the total number of iterations to 776; that is less than half the Fair–Taylor algorithm in its original form and only three times more expensive than the cost of solving the model with fixed expectations.

(ii) The London Business School model

The LBS is a much larger quarterly model and is solved here over 36 periods – hence it is much more costly to solve than the LPL model. The same search procedure was used as for the LPL case. The Fair–Taylor type (b) solution has optimal parameter values of unity. With no extrapolations and complete inner iterations, this method then requires 10999 steps. On the other hand, using incomplete iterations, also without any extrapolations, only 4934 iterations are required; that is just 42% of the standard case. This is a greater gain than that reported in Fisher *et al* (1986) because the overall convergence criteria have been tightened to τ_I=.0002 (.02%). The harder the model is to solve, the greater the

TABLE 3.1: Survey of empirical results : first order methods

	Liverpool Annual model 14 periods	LBS Quarterly model 36 periods	NIESR Quarterly model 23 periods
Strategy	$\alpha_0,\gamma_0,\gamma_1$;Itn's [1] [2]	$\alpha_0,\gamma_0,\gamma_1$;Itn's	$\alpha_0,\gamma_0,\gamma_1$;Itn's
F–T, SOR	0.65,1.0,0.6;1324	1.0,1.0,1.0;10999	1.0,1.0,0.9;5262
F–T, FGS	0.65,0.95,0.6;1300	1.0,1.0,1.0;10999	1.0,1.0,0.9;5262
I.I.I. SOR	0.65,1.0,0.6;827	1.0,1.0,1.0;4934	1.0,1.0,0.9;1915
I.I.I. FGS	0.7,0.85,0.55;776	1.0,0.9,0.95;4615	1.0,0.95,0.9;1875
F.E.	0.7,0.85,–;283	1.0,0.9,–;999	1.0,1.0,–,215

Notes

[1] α_0 SOR damping parameter
γ_0 inner loop FGS extrapolation parameter
γ_1 outer loop extrapolation parameter

[2] Total number of iterations across all time periods

[3] F–T = Fair–Taylor
III = Incomplete Inner Iteration
FE = Fixed Expectations
SOR = Successive–Over–Relaxation
FGS = Fast–Gauss–Seidel

gains usually found from these methods. The FGS extrapolations which have no efficiency gain in the Fair–Taylor solution, do improve the incomplete iteration solution to the final figure of 4615. Finally it should be noted that the cost of solving the model for consistent expectations is only some 4.5 times that for the fixed expectation case.

(iii) The National Institute Model (Comparison with Hall method)

The NIESR model has certain features which distinguish its convergence properties from the other models. In particular, the exchange rate equation contains a unit root in the dynamics of its structural form and the variable is, in itself, one of the last to converge in the iterative process. This unit root (which is an imposed condition on the lag and expectation terms) leads to a near–unit system root and causes the model to be heavily intertemporally dependent in both directions. The practical impact of this is that a large number of outer expectation loops are required in order to solve the model. The Fair–Taylor type (b) solution needs a total of 5262 iterations and FGS yields no improvement although optimizing the outer loop parameter $\lambda = 0.9$ is highly beneficial.

Because the exchange rate is one of the last variables to converge, an incomplete iteration scheme based on a restricted convergence test is not effective. This version of the model originally used the Hall (1985) type (a) method whereby only a single iteration is performed each period. For our incomplete iteration scheme we optimised with respect to a fixed number of inner loop iterations (such as the method used by Fisher *et al.*, 1986, for the LBS model). In this case we found the ideal number to be 2 inner loop iterations combined with one complete preliminary solution and the customary complete solution once expectations have converged. Whereas the Hall method required a total of 3795 iterations, our most efficient scheme needed a total of only 1875. As in the LBS case, FGS extrapolations yielded some improvement with the incomplete iterations that was not possible with either the Fair–Taylor case or the single iteration scheme.

As a result of this exercise some conclusions can be made with respect to the

Hall type (a) method. Firstly, the fixed small number of iterations ought to be efficient in the face of either a large number of expectational terms or near–unit roots. However, it is clear that a single iteration is unlikely to be the most efficient choice. Secondly, the single iteration case requires a convergence test between successive passes on the entire stacked system, and there is no separate outer loop updating expectations, just the usual Jacobi or FGS acceleration. Hence the convergence test on the expectations becomes coincident with the test for the endogenous variables. The single iteration case therefore requires expectations to converge simultaneously to the same tolerance as the underlying variables. In the two–part iterations, the outer loop is a single iteration updating the expectations and the outer convergence test is therefore a test for consistency of expectations. The tolerance level for the outer test is usually set to be weaker than the inner loop tolerance level. The single iteration case does not allow such a distinction and thus removes a degree of freedom in controlling the algorithm. Finally, the one iteration case has in practice proven much more susceptible to problems of invalid arithmetic and the SOR parameter α_0 needs to be reduced to 0.85 in order to generate a solution. Since α_0 is an inefficient form of damping but no other is available, this contributes to the higher costs.

3.6 Penalty function methods (Newton's method)

Penalty function methods were originally suggested for rational expectations models in the context of simultaneous optimal control and consistent expectations solutions by Holly and Zarrop (1979, 1983). That context is examined in Chapter 7. In this section we concentrate on the use of the penalty function to solve for consistent expectations. We show that the penalty function method can be interpreted as a type (b), two–part iterative scheme. The inner loop solves the model conditional on fixed expectation values as before, but the outer loop, which updates expectations, is then replaced by a derivative–based method for calculating the ("optimal") expectations values. In this approach it is convenient to distinguish clearly between the iteration values of the model variables (y_{t+1}) and the expectation terms, denoted (y^e_{t+1}) and we delete any references to the conditioning information to simplify notation. Both terms are expectations conditional on the same information set but they are only coincident at the final solution. In order to obtain the outer loop values, we set up a cost function to be minimized. For a simple quadratic example, we have the cost function evaluated at iteration s:

$$J^{(s)} = \sum_{t=1}^{T} \sum_{j=1}^{n} w_{jy}(\dot{y}^{(s)}_{j,t+1} - y^{e\,(s)}_{j,t+1})^2 + \sum_{t=1}^{T} \sum_{j=1}^{n} w_{jx}(y^{e\,(s)}_{j,t+1} - y^{e\,(s-1)}_{j,t+1})^2, \tag{3.21}$$

where the w_{jy} and w_{jx} are weights to be chosen for the n inconsistency terms and the n expectations terms respectively. These latter terms can be used as a form of damping factor but the w_{jx} weights are normally set to be very small so that there is no cost attributed to these terms at the final solution. More general functions than (3.21) allow the weights to vary across time periods and to place a cost on cross–product terms.

The cost function (3.21) can be submitted to any standard optimal control or function minimization package. In the United Kingdom a common algorithm used is the quasi–Newton scheme of Rustem and Zarrop (1979, 1981) and that is used in this thesis. However, in minimizing a scalar–valued function the convergence test needs careful attention. If consistent expectations is intended to be a binding constraint on the solution of the macroeconomic model, then the function J has an attainable minimum value of zero. Convergence based on the (absolute or percentage) change in J is not sufficient to attain consistent expectations, period–by–period. We therefore re–introduce the standard convergence test for consistent expectations i.e.

$$|(\dot{y}_{i,t+1} - y^e_{i,t+1})/y^e_{i,t+1}| < \tau_0, \forall\, t.$$

The function J is minimised by calculating an "optimal" set of values for the expectations. Holding expectations fixed, we calculate the partial derivative matrix N with elements $\partial(y_{j,t+1}-y^e_{j,t+1})/\partial(y^e_{i,s})$ from the econometric model. The outer loop then calculates approximately optimal values by assuming the model is a linear relation between $(Y-Y^e)=N.Y^e$ and solving the first order conditions for minimizing the cost function (3.21) subject to this linear model. The inner loop then re-solves the model using these approximately optimal values of the expectations. If the model is linear and the weights on the expectation terms (w_{jx}) are small, the resulting solution will be consistent. If not, the optimization is repeated until no improvement is achieved. The process can therefore be viewed as a type (b) iterative scheme in which the outer loop step is a derivative-based iteration. The penalty function is not exactly equivalent to Newton's method but for zero instrument costs (w_{jx}) and constant target costs (w_{jy}) the first order conditions for the minimisation coincide with the Newton algorithm. Writing the inconsistency terms as a function of the expectation terms only we have $g(y^e_{t+1})=0$ (substituting out all the other endogenous terms) and by stacking up the system over time we obtain $G(Y^e)=0$. From equation (2.8) we can obtain Newton's method iteration s: $Y^e(s+1)=Y^e(s)-H^{-1}G(Y^e(s))$ where H is the matrix of partial derivatives of G with respect to Y^e. This iteration is consistent with the differential of the stacked objective function i.e. $\Delta J=-W_Y H.\Delta Y^e$. For diagonal weighting matrices and weights of unity we can obtain $\Delta Y^e=-H^{-1}\Delta J$ which is equivalent to the Newton iteration. (Hence one can also see how the weights may act as a step length parameter in the Newton iteration.)

Furthermore we can interpret the penalty function approach to forward expectations models as analogous to a sparse-system Newton method applied to the stacked system (3.11). In these latter applications (e.g. Don and Gallo, 1987) above diagonal elements in the simultaneous block ($\bar{B}$ in 3.11) are called feedback or loop variables. These feedback variables are identified and solved separately using a Newton algorithm. In this case the system is sparse in its block structure and the above diagonal blocks contain only expectation terms. However, one could set up the algorithm to include the feedback variables from B_0 as well as the expectation terms.

Given that the penalty function is analogous to Newton's method, we would expect it to require fewer, but more expensive, outer loop steps than the first order iteration. For conventional models such comparisons usually favour the latter (Hughes Hallett and Fisher, 1987) but the sparse system nature of the algorithm may make it competitive if the number of expectation terms is small.

The penalty function method has fixed costs arising from one base solution and n derivative evaluations (n dynamic simulations with fixed expectations). However, the derivatives could be approximated and/or saved for use in repeated solutions. One possible approximation is that implied by the first order iteration:

$$\partial(y_{j,t+1} - y^e_{j,t+1})/\partial y^e_{i,s+1} = -1,\ j=i, t=s;\ 0 \text{ otherwise.}$$

Storage of the derivative and Hessian matrices (if required) can also be a problem when a large number of expectation terms exist. In the results presented below we ignore storage costs and the costs of obtaining the optimal solutions (which are of an order of magnitude less than the iteration costs). We calculate the derivatives numerically but we distinguish separately between the fixed costs and those of the repeated solutions during the descent path.

Finally, the choice of weights can, in principle, affect the speed of convergence since the optimal values at each step depend on those weights (see Hughes Hallett and Rees, 1983, Ch.7 for a formal summary of results using quadratic optimization procedures). In practice the weights are set to ensure that consistency is achieved and the convergence speed can not be improved by altering those weights. This arises because the w_{jx} are set very small so that at the optimum there is zero weight attached to the instruments. The targets are therefore all simultaneously perfectly obtainable and there is no implied trade–off either between targets and instruments or between targets. However, the weights can affect whether a particular set of optimal values produce a decrease or an increase in J *ex post*. This is a possibility because of the nonlinearities involved. In these circumstances a step factor is used to try and find a descent path (equivalent to increasing all the instrument weights by a scale factor) and the relative weights within the targets are then important. A combination of sensible weights and suitable damping factors is usually sufficient to eliminate any non–convergence.

The algorithms used here are those of Rustem and Zarrop (1979, 1981) with amendments to the convergence test as indicated above. The weights (q_{iy}) are chosen to be neutral between variables, reflecting only their scale and deviation from consistency. The weights (w_{jx}) are similarly set but with a scale factor of 0.01: thus they have negligible impact on J.

3.7 Comparative costs of the penalty function method

The results are shown in Table 3.2 for the same three models and input shocks used in Section 3.5. Considering the fixed costs first, these reflect the result that for a model of fixed size they would increase linearly with the number of expectation terms. These costs might be predicted by taking (n+1) times the fixed expectation costs shown in Table 3.1. In fact the actual costs lie between 65–73% of that figure. This is because the impulse shocks associated with the derivative evaluations have less impact than the step change in government expenditure for which we wish to solve the model. The NIESR model has fixed costs which exceed those of an unoptimized Fair–Taylor solution (6401>5262). The LPL model has fixed costs similar to an optimised Fair–Taylor solution (1300). The LBS model has fixed costs which are substantially less than the fastest iterative solution (2777<4615). Only on this model would one therefore contemplate penalty function methods which evaluate the derivatives every repeated solution.

Considering the iteration costs alone, it is clear that penalty function methods are not efficient in achieving a solution which satisfies the period–by–period convergence criterion. Each of the models has iteration costs substantially in excess of the Fair–Taylor scheme, which is the least efficient first–order method. The loss increases with the number of expectation terms. This is in contrast to the usual result that first–order schemes become relatively less efficient as the system becomes more dense but is entirely consistent with sparse system results which require a relatively low number of feedback variables for Newton to be more

TABLE 3.2: Penalty function costs

Model	Number of derivative evaluations	Derivative costs	Number of optimisations	It'ions[2]	Total costs
NIESR (23 quarters)	40	6401	82	24220	30621
LPL (14 years)	6	1293	8	2002	3295
LBS (36 quarters)	3	2777	16	15978	18755

Notes

[1] Derivative costs entail one base solution and one derivative evaluation for each expectation term.

[2] The iteration count does not include the cost of calculating the new optimal instrument values, only the costs of re–solving the model.

[3] Model solutions use optimal values for γ_0.

efficient (Hughes Hallett and Fisher, 1987). That is to say, since the Newton scheme is not being applied to every variable both the penalty function and first–order methods become more costly as the number of expectations terms increases. Extrapolating these results, it is possible that a model with just one expectation term would be more efficiently solved by a penalty function approach and this is supported by the conclusions of Westaway and Whittaker (1986).

This empirical result also reflects the nonlinearity of the models and the approximations of the algorithm. Newton algorithms for solving a general nonlinear problem are known to possess at least quadratic convergence rates under fairly weak conditions (see Hughes Hallett and Fisher, 1987) whereas first order methods possess only linear convergence rates. However the convergence rate of the Newton algorithm relies on the derivatives being re–calculated or re–approximated at every iteration. If the derivatives are calculated less frequently the convergence speed tends back to a linear rate (Ortega and Rheinboldt, 1979, p316). Furthermore the quasi–Newton algorithm used here involves approximations to the derivative and Hessian matrices and the loss arising from these approximations may be large when the model is nonlinear. The derivatives used in our calculations, are not re–evaluated between iterations and the assumption is made that derivatives calculated for the first period can be used to approximate the derivatives in later periods. The combination of these approximations is clearly reducing the convergence rate of the algorithm. More frequent derivative calculations cannot be contemplated because of the costs involved.

For the LBS and LPL models the descent trajectory was quite unproblematic. For the NIESR model an intermediate iteration produces an increase in J before consistent expectations has been achieved. In this model, global convergence depends almost entirely on convergence of the exchange rate and so a very high weight on that expectation term combined with a damping factor of 0.5 on the outer loop steps is sufficient for a solution to be found.

Sensitivity tests support the proposition that convergence speed is generally neutral with respect to the weights. It is therefore difficult to find techniques for speeding up the penalty function method. However, one can use extrapolation parameters on the outer loop and incomplete iterations for the inner loop just as with the first order methods. Indeed these could be adapted for any numerical optimization problem. These suggestions have not yet been examined, but even if they produced the same proportionate gains as for first order methods, penalty function methods would still be relatively inefficient.

3.8 Shooting techniques

The shooting technique is a general procedure for solving nonlinear difference equations (see. e.g. Keller, 1968 or Roberts and Shipman, 1972). The basic technique, described below, is known as single shooting since it solves the whole solution period with one iterative scheme. Multiple shooting is a generalization of this basic method which breaks the sample period into smaller sub–periods. This has been proposed by Lipton, Poterba, Sachs and Summers (1982) as a solution procedure for nonlinear economic models operating under forward consistent expectations. The methods can be interpreted as a finite–time, nonlinear model equivalent to the Blanchard and Kahn (1980) solutions for a linear model. In this section we extend the presentation of Lipton *et al.* to cover some particular formulations which are more representative of practical forecasting models. We then go on to extend the expectational framework to cover models with more than one lead and we propose extensions to the different types of terminal condition that can be used. Shooting techniques may not be feasible or practical on all models and conditions under which either of these situations arise are discussed in Section 3.9. In Section 3.10 we then present an *ad hoc* comparison of shooting methods with first–order iterative techniques.

(i) The method

The basic shooting method requires the system to be specified in the following nonlinear state space form:

$$x_{t+1} = F_t(x_t, z_t), \tag{3.22}$$

where, following the notation of Lipton *et al.*, x_t is an $n\times 1$ vector of endogenous variables, z_t is a $k\times 1$ vector of forcing (exogenous) variables and F_t is a real–valued vector of functions. The vector x_t is partitioned as

$$x_t = \begin{bmatrix} v_t \\ w_t \end{bmatrix}, \tag{3.23}$$

and we define boundary conditions $v_{T+1} = \bar{v}$, $w_1 = \bar{w}$ (note that the initial period here is chosen to be period 1 rather than 0 as used by Lipton *et al.*: this is consistent with our presentation of other algorithms). This partition separates out the (m) 'jump' or forward–looking variables (v_t) from the (n–m) backward–looking terms (w_t). This dichotomy allows us to include lag terms in the model by constructing elements of w_t and equating their leads to current dated variables in v_t. The shooting algorithm requires the user to form an initial guess for v_1 and to integrate F forward until T+1 is reached. If x_{T+1} satisfies the terminal condition the solution is found. If not then a Newton algorithm is used to update the initial condition v_1 as a function of the terminal condition error, iterating until the terminal conditions are satisfied.

By successive application of F_t, we can write the terminal values as a function of the initial condition:

$$x_{T+1} = H\left[\begin{bmatrix} v_1^* \\ \bar{w} \end{bmatrix}; z_1, \dots, z_{T+1}\right] = \begin{bmatrix} \bar{v} \\ w_{T+1} \end{bmatrix}, \tag{3.24}$$

where v_1^* is to be found. Each value of v_1 which solves equation (3.24) is a solution to the two–point boundary value problem. To generate the Newton algorithm we define the function H^m which maps the arguments of H into the m terminal values i.e. $\bar{v} = H^m(v_1, \bar{w}, Z)$. Since $\bar{w}$ and Z are given, we take a first–order Taylor series approximation around the solution $\hat{v}_1$:

$$H^m(v_1) \cong H^m(\hat{v}_1) + \left[\frac{\partial H^m}{\partial v_1}\right](v_1 - \hat{v}_1), \tag{3.25}$$

where the matrix of numerical derivatives is evaluated at $\hat{v}_1$. Taking v_1 as the solution v_1^* and noting that $H^m(v_1^*)=\bar{v}$, we can write

$$\bar{v} - v_{T+1}(\hat{v}_1) \cong \left[\frac{\partial H^m}{\partial v_1}\right](v_1^* - \hat{v}_1). \tag{3.26}$$

Finally, if the inverse of the derivative matrix exists, we can solve for v_1^*:

$$v_1^* \cong \hat{v}_1 - \left[\frac{\partial H^m}{\partial v_1}\right]^{-1}(v_{T+1}(\hat{v}_1) - \bar{v}). \tag{3.27}$$

If H^m is linear, the approximation is exact and equation (3.27) will solve for v_1^*. If the model is nonlinear we can use the approximation to generate the following Newton iteration:

$$v_1(s) = v_1(s-1) - \left[\frac{\partial H^m}{\partial v_1}\right]^{-1}(v_{T+1}(v_1(s-1)) - \bar{v}), \tag{3.28}$$

where the derivative matrix is evaluated at $\{v_1(s-1), \bar{w}, Z\}$.

Multiple shooting is a generalisation of the basic (single) shooting algorithm which breaks up the solution period into smaller sub–periods. In applications of single shooting, it was found (Lipton *et al.* p133) that:

> " ... simple Newton search typically fails in economic applications. Incorrect guesses of v_0 are magnified through time so that huge errors are recorded in v_T."

In such circumstances, Lipton *et al.* report that the Newton iterations simply break down, either due to a poor approximation of the first–order Taylor approximation, an ill–conditioned derivative matrix or the mapping F becoming impossible to evaluate at some time period. The theoretical explanation of this empirical finding does not appear to have been previously stated. A stable linear system solved in the form of equation (3.22) automatically becomes unstable in solution. The Blanchard and Kahn conditions for a unique stable path require that there be as many unstable roots as there are non–predetermined (i.e. expectational) variables. Hence solving a rational expectations model by expressing it in backward mode only, must generate a system which is inherently unstable. For example, consider the simple system:

$$x_{t|t-1} = 0.8\, x_{t+1|t-1} + \alpha\, z \tag{3.29a}$$

or

$$x_{t+1|t-1} = 1.25\,(x_{t|t-1} + \alpha\, z). \tag{3.29b}$$

This equation is stable as a function of the terminal condition x_{T+1}, i.e. it has a root larger than unity. However if we try to solve this system in the form of (3.29b), each iteration trajectory will be dynamically unstable since the lag coefficient is greater than unity. The algorithm is, in fact, searching over a space in which there is one stable solution and an infinite number of unstable solutions. Hence, interim solutions during the iterations are necessarily unstable. Other algorithms solve the predetermined part of the model conditional on the expectations variables. Usually the predetermined part delivers only stable roots and the problem of instability does not then arise in either the first–order or penalty function algorithms given the existence of a saddlepoint.

To overcome this problem with the shooting algorithm we sub–divide and stack the time period T to produce multiple shooting. By reducing the time interval to smaller periods the numerical problems become less inhibiting. The presentation given above generalises so that the vectors of initial conditions for each sub-period are stacked over time, and the initial conditions of all but the first sub-period are used to generate the target terminal conditions of the previous sub-period. Hence the sub-periods overlap by one period. The derivative matrices are defined linking the terminal condition of each sub-period to its initial condition and are similarly stacked (Lipton *et al.* p1332). The whole system is then solved simultaneously. An example of the precise formulation of a multiple shoot for the two sub-period case is given in Section 3.10. A method has been suggested by Spencer (1985) to make the formation of the sub-periods dependent on the observed terminal condition error but that "bounded shooting" procedure is not considered here.

(ii) Generalizing to a wider class of models

Lipton *et al.* suggest the following generalization to cover a wider class of models:

$$x_{t+1} = F_t(x_t, y_t, z_t), \tag{3.30a}$$
$$0 = G_t(x_t, y_t, z_t), \tag{3.30b}$$

where G represents part of the structural form of an econometric model, F the remainder and the (j×1) vector y represents variables whose expectations do not appear in the model.

If the conditions for the implicit function theorem hold then (3.30b) can be rewritten to give

$$x_{t+1} = F_t(x_t, B_t(x_t, z_t), z_t), \tag{3.31a}$$
$$y_t = B_t(x_t, z_t). \tag{3.31b}$$

In general the partition involved in (3.30) and the implied sub-block forms in (3.30a) and (3.31) are not analytically available for nonlinear macroeconomic systems and we solve the two parts of the structural form by numerical methods simultaneously.

Returning to our own notation, consider the nonlinear structural econometric model

$$f(y_{1,t}, y_{2,t}, y_{2,t+1|t-1}, z_t; \theta) = u_t, \tag{3.32}$$

where f is an $n \times 1$ vector of general nonlinear functions; $y_{1,t}$ is an $n_1 \times 1$ vector of endogenous variables; $y_{2,t}$, $y_{2,t+1|t-1}$ are $(n-n_1) \times 1$ vectors of the remaining endogenous variables and their one period ahead expectations; z_t is a $k \times 1$ vector of predetermined variables and θ is a vector of parameters; u_t is a $n \times 1$ vector of structural disturbance terms.

In order to obtain explicitly the form required for the shooting technique we require the following partition and normalization to exist (conditioning on period 0 and dropping the conditioning information notation for simplicity):

$$y_{2,t+1} = f_2(y_{1,t}\ y_{2,t}\ y_{2,t+1}, z_t; \theta), \tag{3.33a}$$
$$y_{1,t} = f_1(y_{1,t}\ y_{2,t}\ y_{2,t+1}, z_t; \theta). \tag{3.33b}$$

Each period is then solved conditional on $y_{2,t}$ and z_t ($y_{2,t}$ is fed through from the previous period). The block recursive feature of (3.31) has now been lost and the two equations of (3.33) are solved simultaneously by numerical methods such as Gauss–Seidel or Newton. Conditions in which the functions (3.33) are not explicitly available are dealt with in Section 3.9.

(iii) Generalizing the expectational framework

Consider the case where all expectations are of m periods ahead, m>1. The system will solve as before, but now m sets of initial and terminal conditions are required and the sub–periods of the multiple shoot will each overlap by m periods. Each period uses $y_{2,t}$ generated in period t–m and in turn generates $y_{2,t+m}$.

If expectations are of mixed leads or use several leads for each variable we take the longest lead on each variable and then set up the system for period t as follows where each vector $y_{i+1,t+i}$, defines the vector of variables with a maximum lead of i:

$$\begin{aligned}
y_{m+1,t+m} &= f_{m+1}(y_{1,t}\ y_{2,t}\ y_{3,t+1}\ y_{3,t}\ y_{3,t+1}\ y_{3,t+2} \cdots y_{m+1,t+m}, z_t; \theta),\\
y_{m,t+m-1} &= f_m\ (y_{1,t}\ y_{2,t}\ y_{2,t+1}\ y_{3,t}\ y_{3,t+1}\ y_{3,t+2} \cdots y_{m+1,t+m}, z_t; \theta),\\
&\ \ \vdots \qquad\qquad\qquad\qquad \vdots \qquad\qquad\qquad\qquad\qquad (3.34)\\
y_{2,t+1} &= f_2\ (y_{1,t}\ y_{2,t} \qquad\qquad\qquad\qquad \cdots y_{m+1,t+m}, z_t; \theta),\\
y_{1,t} &= f_1\ (y_{1,t}\ y_{2,t} \qquad\qquad\qquad\qquad \cdots y_{m+1,t+m}, z_t; \theta),
\end{aligned}$$

so that there are (m+1) sub–vectors y_j, j=1,m+1 and each sub–vector has maximum lead j–1. Each equation in each sub–vector has its maximum lead term as the dependent (or output) variable.

This system can now be solved numerically for period t (the dating is nominal) with all endogenous terms not dependent variables in (3.34) held constant. The interim future values not generated as expectations in period t are generated in previous time periods t–m to t–1. A mixture of initial values are thus required, each of which will be updated by a corresponding error in a terminal condition.

(iv) Generalizing the terminal conditions

Lipton *et al.* give a set of terminal conditions as fixed values:

$$y_{T+1} = y^*_{T+1}. \qquad (3.35)$$

They suggest that more general schemes can be treated by variable transformations. Such a transformation $V(\cdot)$ can be expressed as

$$y_{T+j} = y^*_{T+j} \equiv V(y^*_{T+j-1}, \cdots, y^*_{T+1}, y_T, y_{T-1}, \cdots\ ), \qquad (3.36)$$

which requires the terminal value y^*_{T+j} to take calculated numerical values instead of those supplied (e.g. $y^*_{T+j} = y_T$ or $y^*_{T+j} = y_T + j(y_T - y_{T-1})$). The terminal conditions for a sub–period k which has interval $y_{1(k)}, \dots, y_{T+1(k)}$ are given by:

$$y_{T+1(k)} = y^*_{1(k+1)}, \tag{3.37}$$

so that the final value of sub–period k equals the initial value of sub–period k+1 which, since the sub–periods overlap, is equivalent to ensuring that expectations are consistent across the sub–periods.

3.9 The feasibility of shooting techniques

(i) The non–existence of a solution

The problem of possible non–existence of a feasible solution can be demonstrated for the linear case. Consider the linear system without expectations

$$By_t = Cz_t + u_t. \tag{3.39}$$

A necessary condition in order to obtain a reduced form solution for the model is that B is nonsingular so that its inverse exists. By construction, the system is usually normalized so that B has unit elements down the main diagonal and it is highly unlikely that B will be singular since this requires very particular cross–equation restrictions. Now consider the linear system with expectations:

$$By_t = Cz_t + Ay_{t+1|t-1} + u_t. \tag{3.40}$$

The matrix A is usually singular since only a subset of the possible expectations terms are used. Let us assume that we can partition A into $\begin{bmatrix} A_1 \\ 0 \end{bmatrix}$ where A_1 is non–singular. Then for shooting we can solve equation (3.40) as:

$$-A_1 y_{2,t+1} = C_2 z_t - B_{11} y_{2t} - B_{12} y_{1t} + u_{2t} \tag{3.41}$$
$$B_{22} y_{1t} = C_1 z_t + B_{21} y_{2t} + u_{1t}.$$

However, (3.40) will not generally partition into this form simply by re–arranging. This can be demonstrated with a trivial example. Take the following two–equation system:

$$y_{1t} = a_1 y_{2t+1} + c_1 z_{1t} \tag{3.42a}$$
$$y_{2t} = c_2 z_{2t}. \tag{3.42b}$$

The system (3.42) can easily be solved by leading (3.42b) and substituting into equation (3.42a). A first–order type (b) iteration could solve (3.42) in one outer loop. However, as it stands, the system cannot be solved by the shooting technique. It is impossible to re–arrange this system so that y_{1t} and y_{2t+1} are the dependent variables. To overcome this problem we could manipulate the system by substituting out (3.42b). Alternatively we can lead equation (3.42b) so that we then have two equations with y_{1t} and y_{2t+1} as dependent variables. This procedure is proposed by Preston and Pagan (1982, Section 10.4) who consider such lead transformations in a "shuffle algorithm" designed to reduce the entire system to a first order difference equation. These procedures are not useful to us in this context because this kind of manipulation is generally intractable on nonlinear systems. If the shuffle algorithm terminates after the first iteration (i.e. before use of the lead operator) it simply produces a system of the form (3.41). Such a system would be "regular" by Preston and Pagan's Definition 10.2 and Theorem 10.4. Regularity is sufficient but not necessary for the system to be "solvable" (Preston and Pagan Definition 10.1).

There is no inherent property of linear rational expectation models which ensures that they can be written in the form of (3.41). Since the linear model is a particular case of the general nonlinear model, the same conclusion can be drawn. This result can be expressed in graph theoretic concepts by the fact that we have no *a priori* restriction to ensure that $y_{1,t}$, $y_{2,t+1}$ form a feasible output set (Steward, 1962) for (3.40) or (3.41) whereas, $y_{1,t}$ $y_{2,t}$ will normally form such a set for reasons given above (i.e. B^{-1} exists). Even if we obtain (3.41) we cannot then be sure that $\begin{bmatrix} -A_1 \\ B_{22} \end{bmatrix}$ is invertible for all models.

(ii) Reformulating the nonlinear system

In a large nonlinear econometric model the equations are usually formulated in a way which reflects their estimated structural form. Expectational terms will invariably appear as explanatory variables not dependent variables. Hence, finding the implicit functions (3.33) may not be an analytically tractable problem even when a linearised form of the model does not suffer from the problems noted above. This problem can be overcome by re–writing (3.33a) in its original structural form

$$0 = f_3(y_{1,t}\ y_{2,t}\ y_{2,t+1}, z_t; \theta), \tag{3.38}$$

and solving (3.38) and (3.33b) numerically for y_{1t} and $y_{2,t+1}$.

Unfortunately this precludes the use of single parameter Gauss–Seidel iterative methods since no simple updating relation is available between the error in an equation of (3.38) and any variable in $y_{2,t+1}$. Newton–based methods which calculate descent directions can solve equation (3.38) given sensible start values. The preclusion of Gauss–Seidel would imply substantial cost increases in most large, sparse systems even if one were able to patch together a Gauss–Seidel system for solving (3.33b) with a Newton method for solving (3.38).

(iii) An example of a model for which shooting is infeasible

One model which appears not to be suitable for this technique is the Liverpool model as published in Minford *et al.* (1984). If one takes the 27 equations there defined, plus two missing identities, then there is a sub–block of eight equations containing nine current dated endogenous variables. For first–order iterative techniques or penalty function methods this sub–system forms part of the whole without any problems since the output set of the equation system in each time period (i.e. the variables for which it solves) are simply the current dated endogenous terms. However, if we attempt to re–arrange the system so that the expectational terms (the longest leads on four variables) are in the output set (i.e. are solved for) and their corresponding current dated terms are not, then this sub–block of eight equations now contains only seven output variables. Any system which contains such a sub–block is unable to produce a solution, even as part of a larger system, unless there is a redundant equation in the sub–block and a redundant variable elsewhere (Steward, 1962). Thus it would appear that this version of the Liverpool model cannot be solved by these shooting techniques without some degree of prior manipulation.

3.10 Comparative analysis of the shooting method

In the shooting procedure each pass through the equations produces consistent expectations but the terminal conditions do not hold until solution is achieved. In first–order iterative algorithms the model is solved for fixed expectations which are then updated in proportion to their inconsistency. Thus the terminal conditions always hold but expectations are not consistent until convergence is achieved.

Although the same set of structural equations are being solved there is an important numerical difference in the approach. The shooting method also has an inner and outer loop but the outer loop is the terminal condition/initial value relation and the inner loop solves for a different nominated output set each period. In the first–order iterative (FOI) algorithm a distinction is made between the expectations terms (denoted y^e_{t+1}) and the current iteration value y^*_{t+1}. A linear relation between the two is used for the outer loop at iteration (s):

$$y^e_{t+1}(s) = (1-\gamma)y^e_{t+1}(s-1) + \gamma y^*_{t+1}(s-1), \tag{3.43}$$

for some damping factor γ.

In the shooting technique the Newton algorithm yields the following relation between y_1 and $(y_{T+1}-y^*_{T+1})$:

$$y_1(s) = y_1(s-1) - \left[\frac{\partial H}{\partial y_1}\right]^{-1} (y_{T+1} - y^*_{T+1}), \tag{3.44}$$

where H is the implicit function which generates y_{T+1} as a function of the initial conditions and exogenous variables. We can re–write equation (3.44) as the non–stationary iteration:

$$y_1(s) = y_1(s-1) - V(s)\,(y_{T+1} - y^*_{T+1}). \tag{3.45}$$

The need to evaluate and then invert $\left[\frac{\partial H}{\partial y_1}\right]$ is a considerable drawback since partial derivatives must be calculated numerically and matrix inversion is costly if large numbers of expectational terms are involved. In the multiple shoot we have (3.45) stacked over time for sub–intervals. This means calculating up to T matrices and then inverting each, which is an enormous numerical burden which exists because of the need to map from $y_{T+1(k)}$ into $y_{1(k)}$ for each sub–period k.

To gain further insight in comparison of this technique we can examine linear systems and look at the equation orderings implied. For simplicity suppose the model takes the form $B_0y_t=Ay_{t+1}+B_1y_{t-1}+Cz_t+u_t$. The lag terms do not normally appear explicitly in the shooting algorithm, but they can be incorporated without undue complexity.

(i) Single shooting

At iteration (s) for the single shoot we have the outer loop:

$$\begin{aligned} y_1(s) &= y_1(s-1) - V(s)(y^e_{T+1} - y^*_{T+1}) \text{ or} \\ W(s)y_1 &= -V(s)y^e_{T+1} + V(s)y^*_{T+1}. \end{aligned} \tag{3.46a}$$

The inner loop system, stacked over time is given by:

$$\begin{bmatrix} -A & 0 & \dots & & 0 \\ B_0 & -A & 0 & \dots & 0 \\ -B_1 & B_0 & -A & & \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & B_0 & -A & 0 \\ 0 & \dots & 0\ -B_1 & B_0 & -A \end{bmatrix} \begin{bmatrix} y_2 \\ y_3 \\ \\ \vdots \\ y_T \\ y_{T+1} \end{bmatrix} = \begin{bmatrix} B_1 y_0 - B_0 y_1(s-1) \\ B_1 y_1(s-1) \\ 0 \\ \vdots \\ \\ 0 \end{bmatrix} + \begin{bmatrix} Cz_1 \\ Cz_2 \\ \\ \vdots \\ \\ Cz_T \end{bmatrix} + \begin{bmatrix} u_1 \\ \\ \\ \vdots \\ \\ u_T \end{bmatrix}. \tag{3.46b}$$

At each outer loop iteration y_0 is given but we must calculate y_1 and then solve (3.40b) in the order shown.

(ii) First–order iterative techniques

The FOI algorithm has the following outer loop steps (e.g. type (b) or Fair and Taylor, 1983):

$$y_{t+1}(s) = y_{t+1}, \tag{3.47a}$$

and the following inner loop system stacked over time:

$$\begin{bmatrix} B_0 & 0 & \dots & & 0 \\ -B_1 & B_0 & & & \vdots \\ 0 & \ddots & \ddots & & \\ \vdots & \ddots & & B_0 & 0 \\ 0 & \dots & 0 & -B_1 & B_0 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ \\ y_T \end{bmatrix} = \begin{bmatrix} B_1 y_0 \\ 0 \\ \vdots \\ \\ 0 \end{bmatrix} + \begin{bmatrix} Ay_2(s-1) \\ Ay_3(s-1) \\ \vdots \\ \\ Ay^*_{T+1} \end{bmatrix} + \begin{bmatrix} Cz_1 \\ Cz_2 \\ \vdots \\ \\ Cz_T \end{bmatrix} + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ \\ u_T \end{bmatrix}. \tag{3.47b}$$

(iii) Multiple shooting, two sub–periods

Let us assume that the interval is broken at point $1<K<T$ so that $K+1$ is the terminal date of the first sub–period and the initial date of the second. The iteration updating the initial values of the two sub–periods is

$$y_1(s) = y_1(s-1) - V_1(s)(y_{K+1} - \overset{*}{y}_{K+1}) \text{ or}$$

$$W_1(s)y_1 = -V_1(s)y_{K+1} + V_1(s)\overset{*}{y}_{K+1} \tag{3.48a}$$

and similarly

$$W_2(s)\overset{*}{y}_{K+1} = -V_2(s)y_{T+1} + V_2(s)\overset{*}{y}_{T+1}. \tag{3.48b}$$

Hence the terminal conditions for the first sub–period are generated by the initial value for the second, thus implicitly linking the very first initial condition (y_1) right through to the final terminal value ($\overset{*}{y}_{T+1}$). The inner loop system stacked over time is then:

$$\begin{bmatrix} -A & 0 & & \cdots & & & & 0 \\ B_0 & -A & & & & & & \\ -B_1 & B_0 & -A & & & & & \\ 0 & & -B_1 & B_0 & -A & & & \\ & & & -B_1 & 0 & -A & & \vdots \\ \vdots & & & & 0 & B_0 & -A & \\ & & & & & \ddots & \ddots & 0 \\ 0 & \cdots & & & & -B_1 & B_0 & -A \end{bmatrix} \begin{bmatrix} y_2 \\ y_3 \\ \vdots \\ y_{K+1} \\ y_{K+2} \\ y_{K+3} \\ \vdots \\ y_{T+1} \end{bmatrix}$$

$$= \begin{bmatrix} B_1 y_0 - B_0 y_1(s-1) \\ B_1 y_1(s-1) \\ 0 \\ \vdots \\ 0 \\ -B_0 y_{K+1}(s-1) \\ B_1 y_{K+1}(s-1) \\ 0 \end{bmatrix} + \begin{bmatrix} Cz_1 \\ Cz_2 \\ \\ \vdots \\ \\ Cz_T \end{bmatrix} + \begin{bmatrix} u_1 \\ \\ \\ \vdots \\ \\ u_T \end{bmatrix}. \tag{3.48c}$$

In the second sub–period, y_{K+2} is dependent on the initial condition y_{K+1} generated as a function of the terminal condition error at period T+1 not on the terminal value of the first period. That value of y_{K+1} is used to estimate the terminal condition error for the first sub–period and thereby to update y_1.

(iv) Comparisons

In the stacked forms for the inner loop given above we have effectively split out the variables which are determined on the outer loop and placed them on the right–hand side. Hence, in each case we are left with a block recursive system which can be solved in a number of ways. In particular, the three schemes proposed for type (b) algorithms in Section 3.4 can equally be applied to the shooting systems. Furthermore, a Newton method could alternatively be used on each of these three schemes. The presentation of iteration matrices given in Section 3.4 applies equally well here, the only difference being that the decompositions for shooting matrices are based on a re–ordered system.

In order to maximize convergence speed we wish to minimize the spectral radius of the iteration matrices for the inner and outer loops. Given that the decompositions are different the iteration matrices will not coincide. For example, using the block recursive scheme to solve system (3.40) period–by–period gives an iteration matrix applied to each period based on a decomposition of A, whereas in the FOI method it is based on a decomposition of B_0. We have no way of comparing these matrices for a general system. The same is true for the outer loop iteration matrices.

The only comparison that can be drawn is based on the results of ordering algorithms. In Gabay *et al.* (1980), and Don and Gallo (1987) it is proposed that the most efficient way to order a model's equations for solution is based on identifying a feedback set. The feedback variables are those which, when removed from the system allow the rest to be ordered in lower triangular form. The minimal feedback set is the smallest set of variables which meets this criterion. A minimal set may not be unique and locating a minimal set is computationally expensive; algorithms for this purpose are given, *inter alia*, by Don and Gallo. Having located a feedback set we place these variables at the end of the ordering and order the rest of the system to be lower triangular.

In the context of consistent expectations models we are proposing to solve the system by splitting out those elements which leave the rest as block lower triangular. Those variables which are solved on the outer loop are the feedback elements in each scheme. Ordering schemes normally aim to minimise the number of feedback variables but there is no guarantee that such orderings are actually more efficient, since efficiency can only be gained by reducing the spectral radius of the iteration matrix. On the basis of published results (e.g. Don and Gallo (1987); Gabay *et al.* (1980)), efficiency gains do seem to be achieved by this approach. In comparing the FOI and shooting algorithms, we note that the latter has a feedback set defined by the initial condition (a maximum n variables) and invokes dependencies on the feedback set in line with the number of non–zero entries in B_0 and B_1. The FOI algorithm tears out Ay_{t+1} every period and hence has a maximum nT variables in the feedback set. Since $1<T$, a single shooting algorithm, if feasible, may produce a more efficient ordering than an FOI. However, the number of non–zero entries in A is usually quite small relative to the number of non–zero entries in B_0 and B_1 and so the result is unclear. If multiple shooting is used the number of torn elements in the shooting algorithm goes up towards a maximum of nT variables. Hence multiple shooting implies a clear reduction in efficiency relative to single shooting and where A is substantially less dense than B_0 and B_1, relative to FOI as well.

In a comparison of outer loop iteration costs the shooting algorithm is a clear loser. The FOI methods simply update the expectation terms to be consistent through a series of identities – a relatively costless exercise given the usual small number of expectation terms. However, the shooting algorithm needs to generate the derivative matrix $\partial H/\partial y_1$ and invert that matrix at each outer loop iteration. For multiple shooting the costs are an order of n larger (the precise formulation of the derivative matrix in this case is given by Lipton *et al.* p1332). The costs of this outer loop and the inherent numerical instability problems considered above make it unlikely that the shooting algorithm will be more efficient unless the number of expectational terms is very small.

3.11 Summary

In this Chapter we have considered three methods for solving nonlinear forward consistent expectations models. First, a variety of first–order iterative techniques have been proposed and particular schemes applied. It has been found that differing models may benefit from different versions of these schemes. Second, penalty function methods have been evaluated but have been found to be less efficient than first–order methods for solving the same problem. It has been shown that the penalty function approach can be interpreted as Newton's method applied to a sparse system. Given this interpretation, the results appear to be well founded. Third, shooting techniques have been surveyed and some extensions given. Shooting methods are shown to be fundamentally unreliable for models that satisfy the usual stability conditions and difficult to apply to the general, nonlinear dynamic structures found in macroeconometric models.

Chapter 4

TERMINAL CONDITIONS, UNIQUENESS AND STABILITY

In this chapter we aim to establish the appropriate choice of terminal conditions in nonlinear forward–consistent expectations models and the conditions under which alternative choices are appropriate. In the process we attempt to clarify the understanding and use of such conditions and establish the link between terminal conditions and the stability and uniqueness conditions.

In Section 4.1 we derive the general saddlepoint conditions for a unique stable solution from a simple presentation of the model in its structural form. In Section 4.2 we then examine the need for terminal conditions to select the saddlepoint path in a finite horizon solution. We propose various choices of terminal condition and assess their implications for accuracy in approximating the saddlepoint solution. This analysis is carried out using a combination of linear model analysis and numerical simulation of demonstration examples.

In Section 4.3 we examine the consequences for the convergence of iterative solution algorithms of alternative choices of terminal condition. We pay particular attention to models which do not satisfy the saddlepoint criteria. In Section 4.4 we then test alternative terminal conditions by their application to three large–scale nonlinear models. For the quarterly models we propose adjustments to the terminal conditions to account for the observed seasonality in the solution path. Section 4.5 contains a summary of our results and conclusions.

4.1 Uniqueness and stability conditions

In the following discussion we derive the stability and uniqueness conditions for a linear rational expectations model. The presentation has the benefit of examining a general linear model in its structural form. We begin by reviewing the stability conditions for non–expectations models. Consider the structural form of the general, dynamic linear model presented in Chapter 2 (equation 2.2):

$$B(L)y_t = -C(L)x_t + u_t. \tag{4.1}$$

We recall that B(L) and C(L) are matrices of polynomials in the lag operator such that

$$B(L) = B_0 + B_1L + \dots + B_pL^p \tag{4.2a}$$
$$C(L) = C_0 + C_1L + \dots + C_qL^q. \tag{4.2b}$$

In this chapter we examine the stability and uniqueness of the dynamic solution to systems of equations such as (4.1) leaving a discussion of alternative model forms and solution modes to Chapter 6. The dynamic reduced form solution to equation (4.1) is given by:

$$y_t = -B(L)^{-1}C(L)x_t + B(L)^{-1}u_t. \tag{4.3}$$

Equation (4.3) is referred to by Salmon and Wallis (1982) as the final form and this can be re-expressed as:

$$y_t = -\frac{b(L)}{|B(L)|}C(L)x_t + \frac{b(L)}{|B(L)|}u_t \tag{4.4}$$

where b(L) is the adjoint matrix comprising finite-degree lag polynomials and $|B(L)|$ is the determinantal polynomial. The basic dynamic solution is obtained by solving equation (4.1) conditional on x_t and u_t, $t=1,\dots,T$ and a set of initial values y_s, $s=1-p,\dots,0$; x_s, $s=1-q,\dots,0$. For the system to produce a stable (stationary) solution for y_t, the matrix B(L) must be invertible such as to produce a distributed lag on x_t in equation (4.4) which is non-explosive. Hence, the stability condition is that the determinantal polynomial $|B(L)|$, must have roots inside the unit circle.

This stability condition can be derived by noting that:

$$1/|B(\xi)| = 1/[\prod_{i=1}^{np}(1-\lambda_i\xi)] = \sum_{i=1}^{np}[A_i/(1-\lambda_i\xi)], \tag{4.5}$$

where λ_i, $i=1,\ldots,np$; are the roots of $|B(\xi)|$ and ξ is an arbitrary argument. If $|\lambda_i|<1.0\ \forall\ i$, then geometric expansion of each term on the right–hand side of equation (4.5) yields a polynomial of infinite length with a declining series of lag coefficients on x_t, i.e.

$$1/(1-\lambda\xi) = (1 + \lambda\xi + \lambda^2\xi^2 + \ldots). \tag{4.6a}$$

Each component of the solution for y_t is then a finite–valued function of lagged values of x_t and hence the complete solution for y_t is non–explosive.

If all the roots of the polynomial $|B(\xi)|$ lie outside the unit circle, the geometric expansions used above generate polynomials with series of increasing lag coefficients on x_t. As these would become infinite in value, the solution for y_t becomes explosive and the expansions are not properly defined. However an alternative expansion can be obtained by writing

$$\begin{aligned} 1/(1-\lambda\xi) &= -\lambda^{-1}\xi^{-1}/(1-\lambda^{-1}\xi^{-1}) \\ &= -[(1/\lambda)\xi^{-1} + (1/\lambda)^2\xi^{-2} + (1/\lambda)^3\xi^{-3} + \ldots]. \end{aligned} \tag{4.6b}$$

In this case we have a series of infinite distributed *leads* (since at this stage, before we consider expectations formation, we define $L^{-1}x_t=x_{t+1}$) with declining coefficients, and a finite–valued solution for y_t can only be defined conditional on the *future* path of x_t. Such a model solution does not coincide with conventional notions of stability (or causality) since attempting to solve (4.1) (in its original structural form, but with roots all greater than unity) given initial conditions for y_s, $s=1-p,\ldots,0$, would yield an explosive trajectory for y_t.

We now consider the following forward expectations model:

$$B_0 y_t = -A_1(F) y_{t+1|t-1} - C_0 x_t + u_t, \qquad (4.7)$$

where, as before, $y_{t+1|t-1}=E(y_{t+1}|\Omega_{t-1})$ is the expectation of y_{t+1} formed using the information set Ω_{t-1} containing information available at the end of period t−1 (including projections or known values of exogenous variables for periods after t−1). Equation (4.7) is a generalization of equation (2.19) where we now allow for lead terms of up to k periods ahead. The matrix polynomial $A_1(F)$ in the forward operator generally implies the particular definition that $Fy_{t+1|t-1}=y_{t+2|t-1}$, $F^2 y_{t+1|t-1}=y_{t+3|t-1}$, etc. as in Sargan (1983). In the consistent expectations dynamic solution we solve for $y_{s+k|s-1}=\hat{y}_{s+k}$ where $\hat{y}_{s+k}$ is the model solution in period s+k.

The consistent expectations solution to equation (4.7) can then be expressed as:

$$\begin{aligned} A(F)\hat{y}_t &= -C_0 x_t + u_t \\ |A(F)|\hat{y}_t &= -a(F)C_0 x_t + a(F)u_t \end{aligned} \qquad (4.8)$$

where $A(F)=B_0+A_1(F)$; $a(F)$ is its adjoint matrix and $|A(F)|$ its determinantal polynomial. Under consistent expectations we define the forward operator applied to a solution value to be equivalent to the inverse of the lag operator ($F=L^{-1}$, hence $F\hat{y}_t = \hat{y}_{t+1}$). If the roots of $|A(F)|$ lie within the unit circle, we can define a stable infinite distributed *lead* on x_t. If the roots are outside the unit circle, we can define an infinite distributed *lag* as noted above. In order to get a unique stable solution to equation (4.7) we will need the former to hold for the following reasons. If we solve equation (4.7) over a finite sample period, t=1,...,T, we require values for the expectations of y_{T+i}, i=1,...,k. These values are for periods outside the solution interval and thus must be supplied as a vector of terminal conditions. Different choices of terminal values will lead to different solution paths. In our discussion of solution algorithms in Chapter 3, these terminal conditions were treated as fixed values. If the stable solution is defined as a function of lagged x_t, then the y_{T+i} may be freely chosen without altering the stability of the solution path. There is then an infinite number of stable solution trajectories. If the stable solution is defined as a function of future x_t, knowledge of the future path of x_t after period T can be used to tie down a unique value of y_{T+i} (example in Section 4.2), and hence a unique stable solution. We therefore require $|A(F)|$ to have all its roots within the unit circle in order to generate a *unique* stable solution.

The general model is the mixed case with both leads and lags

$$B(L)y_t = -C(L)x_t - A_1(F)y_{t+1|t-1} + u_t \tag{4.9a}$$
$$D(L,F)y_t = -C(L)x_t + u_t \tag{4.9b}$$
$$|D(L,F)|y_t = -d(L,F)C(L)x_t + d(L,F)u_t \tag{4.9c}$$

where $D(L,F)=B(L)+A_1(F)$ with adjoint matrix $d(L,F)$ and determinantal polynomial $|D(L,F)|$.

To derive the system roots we factorise the matrix polynomial $D(L,F)$. Assuming that there are no zero roots, we can generally write $D(\xi,\xi^{-1})$ as

$$D(\xi,\xi^{-1}) = D_1(\xi)\, W\, D_2(\xi^{-1}) \tag{4.10}$$

where $D_1(\xi)$ is the same order as $B(\xi)$; $D_2(\xi^{-1})$ is the same order as $A_1(\xi^{-1})$; D_{10}, D_{20} are identity matrices and W is a non-singular matrix. Now the roots of the determinantal polynomial $|D(\xi,\xi^{-1})|$ are given by the roots of $|D_1(\xi)|$ and $|D_2(\xi^{-1})|$. We can rewrite the latter as $|\xi^{-k}D_2^*(\xi)|$ where $|D_2^*(\cdot)|$ has the same coefficients as $|D_2(\cdot)|$ but in reverse order. Expanding the inverse of each of these polynomials separately we can see that we require the roots of $|D_1(\xi)|$ and $|D_2^*(\xi)|$ to lie within the unit circle to avoid explosive solutions. If the roots of $|D_2^*(\xi)|$ are within the unit circle then the roots of $|D_2(\xi^{-1})|$ must lie outside since the roots of one are the inverses of the roots of the other. We therefore require that there be as many roots of $|D(\xi,\xi^{-1})|$ outside the unit circle as there are forward expectation terms and the remainder to be inside the unit circle. This condition then defines both infinite distributed lead and lag polynomials on x_t which tie down the right number of terminal conditions yet preserve the stability of the solution. This is the same saddlepoint condition derived, *inter alia*, by Blanchard and Kahn (1980) and discussed in Section 2.6.

Note that to obtain a saddlepoint condition it is sufficient and necessary to show that the roots of $|D_1(\xi)|$ are inside the unit circle and those of $|D_2(\xi^{-1})|$ are outside. However this does not automatically generate any such conditions on $|B(L)|$ or $|A_1(L)|$. However it will often be the case that a model is constructed so that its backward looking part B(L) is stable and that its forward looking part introduces the unstable roots. This is particularly likely to be true when forward expectations are grafted on to an existing model. In the United Kingdom, the NIESR, LBS and Her Majesty's Treasury model have all developed forward expectations models from backward–looking models (see e.g. Hall and Henry 1985a). As a counter–example, the LPL model was conceived, designed and built to specifically incorporate forward expectations (Minford *et al.*, 1984).

The derivation of the saddlepoint conditions is given by several other authors as discussed in Chapter 2. The exposition developed here has two purposes. We have examined the problem from the viewpoint of obtaining a dynamic solution to the structural form without appealing to any particular solution algorithm or transforming to any special model form. In a nonlinear system we cannot calculate eigenvalues or roots to assess uniqueness and stability. However our analysis shows how it is that the terminal conditions select the solution path in the finite horizon case and that the appropriate choice is that which yields a unique stable solution. This result carries over to the nonlinear system. Secondly, by laying out the model in a general structural form, we can follow directly on to the alternative model forms and solution modes discussed in Chapter 6. Before moving on to examine terminal conditions choice we note two further points.

When dealing with a system of linear equations we can calculate single equation roots to determine the properties of single–equation solutions and the roots of the characteristic polynomial to determine the properties of the system. In general, one cannot use the roots of individual equations to infer the properties of the dynamic system, nor could one attribute system roots to individual equations. However, in principle, the system sets of roots are functions of the coefficients of the system. One can estimate the sensitivity of roots due to changes in coefficient values or the elimination of particular lags. It is often found that individual roots can in practice be associated with particular coefficients (Kuh, Neese and Hollinger, 1983, develop such sensitivity analysis using conventional models).

In a nonlinear system the roots are not defined, except for a linearisation which is, in general, state dependent. By considering a linearisation we may obtain information regarding local stability from a first order approximation to the system. However, this form of analysis has an inherent weakness. It is conventional to linearise around a stable base trajectory. If this is not done, it is possible for the base itself to be responsible for generating unstable roots (see the example in Kuh *et al*). If a stable trajectory can be obtained for the linearisation procedure then the model must have at least one stable solution! It can therefore be difficult to untangle the stability properties of a system using the linearisation approach. Hence the (in)stability of a model is usually assessed first by numerical simulation.

4.2 Terminal conditions in the linear model

To investigate possible ways of setting terminal conditions we return to the example of a price equation as used in Muth's (1961) paper but using forward expectations:

$$p_t = \alpha p_{t+1|t-1} + u_t. \tag{4.11}$$

The forcing variable u_t may represent a known process or a disturbance term. This equation can also be derived from an inverted money demand equation in the new–classical macro–model and, as such, is used by Gourieroux *et al* (1982) as their example equation. We assume throughout that $|\alpha|<1$ and hence a unique stable solution exists since the implied root satisfies the saddlepoint conditions. We recall from Chapter 2 that the solution of equation (4.11) (ignoring conditioning notation) is given by:

$$p_t = \sum_{i=0}^{\infty} \alpha^i u_{t+i}. \tag{4.12}$$

Now, we assume that u_t achieves a steady state equilibrium value of u^* in period T. The solution for period T and after is then:

$$p_T = (1-\alpha)^{-1}u^* = p_{T+i}, \quad i>0. \tag{4.13}$$

For periods prior to T we have the solution:

$$p_t = \sum_{i=t}^{T-1} \alpha^{i-t}u_i + \alpha^{T-t}(1-\alpha)^{-1}u^*, \ t=1,\ldots,T-1. \tag{4.14}$$

Equation (4.13) can provide a terminal value for solving the model numerically over periods 1,...,T . We can write equation (4.14) as:

$$p_t = \sum_{i=t}^{T-1} \alpha^{i-1}u_i + \alpha^{T-t}p_T. \tag{4.15}$$

Thus setting $p_T=(1-\alpha)^{-1}u^*$, the numerical solution to equation (4.15) coincides with the analytical solution given by equation (4.14).

Similar analysis has led some authors to a treatment of the consistent expectations solution as a two-point boundary value problem in which the initial and terminal values are treated as transversality conditions (see for example, Holly and Beenstock, 1980; Minford *et al.*, 1979, 1980). They suggest the imposition of terminal conditions which characterise the equilibrium properties of the model. For the nonlinear model used by Minford *et al.* such conditions are obtained by the analysis of a small linear system on which the nonlinear model is based.

An alternative procedure is to set the terminal value by some kind of automatic rule. We shall consider three choices of terminal condition, each of which may potentially select a different solution. Let $\hat{p}_t$ denote the solution of the model when a terminal condition is imposed.

(i) Fixed value $\hat{p}_T = \bar{p}_T$

Here $\bar{p}_T$ is set from off–model analysis. For example, if p_t is measured as the difference of price from some fundamental equilibrium, $\bar{p}_T$ may be set to zero. This choice of terminal value may then lead to a difference in the value of $\hat{p}_T$ from the analytical solution. This error is given by the relation:

$$p_T - \hat{p}_T = e_T = \alpha[(1-\alpha)^{-1}u^* - \bar{p}_T]. \tag{4.16}$$

If, and only if, $\bar{p}_T$ is chosen to be the equilibrium value then $e_t = 0.0$ for all t. Otherwise, in earlier periods we obtain an error of:

$$e_t = \alpha^{T-t+1} e_T. \tag{4.17}$$

This relation between e_t and e_T holds for all choices of terminal value in this model. Hence the effect of an error in the terminal condition on the solution of interest should go to zero as T increases if $|\alpha| < 1.0$ and e_T does not explode at a rate faster than $1/\alpha$.

(ii) Constant level $\hat{p}_{T+1} = \hat{p}_T$

This condition implies:

$$\hat{p}_T = \alpha \hat{p}_T + u_T = (1-\alpha)^{-1} u_T. \tag{4.18}$$

Therefore if $u_T = u^*$ there is no error produced by this terminal condition in this context. Hence we can use a constant level rule in place of the analytical solution value.

(iii) Constant growth $\hat{p}_{T+1} = \hat{p}_T\hat{p}_T/\hat{p}_{T-1}$

This condition implies

$$p_T = \alpha p_{T+1} + u_T = \alpha (p_T^2/p_{T-1}) + u_T,$$
$$p_{T-1} = \alpha p_T + u_{T-1},$$
and hence
$$p_T = \alpha (p_T^2/(\alpha p_T+u_{T-1})) + u_T,$$
$$p_T(\alpha p_T+u_{T-1}) - \alpha p_T^2 = (\alpha p_T+u_{T-1})u_T,$$
$$p_T u_{T-1} - \alpha p_T u_T = u_T u_{T-1},$$
$$p_T(1-\alpha u_T/u_{T-1}) = u_T. \tag{4.19}$$

Now if $u_T=u_{T-1}=u^*$ this solution collapses to give $p_T=(1-\alpha)^{-1}u^*$, $\alpha\neq1$. Hence a constant growth rule also uniquely solves for the same value as the analytical solution. We will show that the advantage of such conditions is that they are also appropriate when the forcing variable u and hence p has not reached an equilibrium steady state.

Let us now assume that u follows an autoregressive process

$$u_t = \gamma_0 + \gamma_1 u_{t-1}. \tag{4.20}$$

The combined system (4.20) and (4.11) has two roots of α and γ_1. Equation (4.20) ensures that the system is recursive and hence we require that $|\alpha|,|\gamma_1|<1$ for a saddlepoint solution. We can then obtain the following analytical results:

$$u_t = \gamma_1^t u_0 + \sum_{i=1}^{t} \gamma_1^{i-1}\gamma_0 = \gamma_1^t u_0 + \gamma_0(1-\gamma_1)^{-1}(1-\gamma_1^t) \text{ for } \gamma_1\neq1.0, \tag{4.21a}$$
$$u_\infty = \gamma_0(1-\gamma_1)^{-1}, \quad \text{for } |\gamma_1|<1.0. \tag{4.21b}$$

$$
\begin{aligned}
p_t &= \sum_{i=0}^{\infty} \alpha^i [\gamma_1^{t+i} u_0 + \gamma_0(1-\gamma_1)^{-1}(1-\gamma_1^{t+i})], \\
&= u_0 \gamma_1^t \sum_{i=0}^{\infty} \alpha^i \gamma_1^i + \gamma_0(1-\gamma_1)^{-1} \sum_{i=0}^{\infty} \alpha^i - \gamma_0(1-\gamma_1)^{-1} \gamma_1^t \sum_{i=0}^{\infty} \alpha^i \gamma_1^i, \\
&= u_0 \gamma_1^t (1-\alpha\gamma_1)^{-1} + \gamma_0(1-\gamma_1)^{-1}(1-\alpha)^{-1} \quad u_0 \gamma_0 (1-\gamma_1)^{-1} \gamma_1^t (1-\alpha\gamma_1)^{-1}, \\
&\quad \text{for } |\alpha|<1.0,\ |\alpha\gamma_1|<1.0,\ \gamma_1 \neq 1.0,
\end{aligned}
$$

and hence

$$
p_t = \gamma_1^t (1-\alpha\gamma_1)^{-1} [u_0 - \gamma_0(1-\gamma_1)^{-1}] + \gamma_0(1-\gamma_1)^{-1}(1-\alpha)^{-1}. \tag{4.22}
$$

Note that we do not require $|\gamma_1|<1.0$ in the derivation of this solution for p_t. If $|\gamma_1|<1.0$ then p_t converges to a steady state solution of:

$$
p^* = \gamma_0(1-\gamma_1)^{-1}(1-\alpha)^{-1}. \tag{4.23}
$$

If $|\gamma_1|>1.0$ then both roots are greater than unity and this contravenes the saddlepoint condition. When $|\gamma_1|>1.0$, the exogenous variables are growing exponentially. In large scale models used for forecasting non–stationary time series, exponential or trend growth in the exogenous variables is quite usual. However, we can always obtain a unique trajectory for the pre–determined variable u_t given u_0 and then, using equation (4.22), a solution for p_t with a unique steady–state growth path as long as $|\alpha|, |\alpha\gamma_1|<1.0$. In this case the steady–state growth rate is obtained as:

$$
p_t/p_{t-1} \to \gamma_1 \text{ as } t \to \infty. \tag{4.24}
$$

If $|\alpha\gamma_1|>1.0$ then the solution for p_t is not finite valued and we rule out this possibility.

We now proceed to evaluate different choices of terminal condition in the two cases:

(a) $|\alpha|, |\gamma_1| < 1.0$,

(b) $|\alpha|, |\alpha\gamma_1| < 1.0, \quad |\gamma_1| > 1.0$.

In order for the solution of interest to be insensitive to changes in the time horizon for sufficiently long solution periods, we require that the terminal period error grow no faster than a rate of less than $1/\alpha$ as $T \to \infty$. Ideally we would like the terminal period error to be small for any particular value of T and, rather than growing, to vanish at as fast a rate as possible. The absolute size of the error and its rate of reduction are therefore the grounds for comparing different choices.

Case (a): the saddlepoint solution

(i) Equilibrium value conditions $\tilde{p}_{T+1} = p^*$.

For this choice we use the long–run solution to the model as the terminal value in period T. This choice is in the spirit of the proposals of Minford *et al.* (1979, 1980).

$$\tilde{p}_{T+1} = (1-\alpha)^{-1}u_* = (1-\alpha)^{-1}(1-\gamma_1)^{-1}\gamma_0$$

hence

$$\tilde{p}_T = \alpha(1-\alpha)^{-1}(1-\gamma_1)^{-1}\gamma_0 + u_T. \tag{4.25}$$

The error is

$$p_T - \tilde{p}_T = e_T = \alpha(p_{T+1} - \tilde{p}_{T+1}) \text{ and hence}$$

$$e_T = \alpha(1-\alpha\gamma_1)^{-1}[u_0 - \gamma_0(1-\gamma_1)^{-1}]\gamma_1^{T+1}. \tag{4.26}$$

Therefore as $T \to \infty$; $e_T \to 0.0$ and so equilibrium terminal conditions possess the required property that $e_1 \to 0.0$.

(ii) Constant level conditions $\tilde{p}_{T+1} = \tilde{p}_T$.

In this case

$$\bar{p}_T = (1-\alpha)^{-1}u_T = (1-\alpha)^{-1}[\gamma_1^T u_0 + \gamma_0(1-\gamma_1)^{-1}(1-\gamma_1^T)]. \qquad (4.27)$$

The error is then:

$$\begin{aligned}(p_T - \bar{p}_T) &= e_T = \alpha(p_{T+1} - \bar{p}_{T+1}) \\ e_T &= \alpha\,\gamma_1^T[u_0 - \gamma_0(1-\gamma_1)^{-1}][\gamma_1(1-\alpha\gamma_1)^{-1} - (1-\alpha)^{-1}]. \qquad (4.28)\end{aligned}$$

Therefore, as $T \to \infty$; $e_T \to 0$ and so constant level conditions also have the required property.

(iii) Constant growth conditions $\bar{p}_{T+1} = \bar{p}_T^2/\bar{p}_{T-1}$

Here we have

$$\bar{p}_T = \alpha\,(\bar{p}_T^2/\bar{p}_{T-1}) + u_T = (1 - \alpha u_T/u_{T-1})^{-1}u_T. \qquad (4.29)$$

It is possible to solve this equation for a closed form expression as a function of u_0 by substituting in for u_{T-1}, u_T. However, this does not appear to be a useful exercise. From the expression given above we can see that as $u_T \to u_\infty = \gamma_0(1-\gamma_1)^{-1}$ and $u_T/u_{T-1} \to 1.0$, then $\bar{p}_T \to p_\infty$ and so the error vanishes.

In comparing the three types of condition we find all three possess the property that the terminal period error vanishes as $T \to \infty$ and therefore all three are potentially valid in producing approximations to the saddlepoint solution. Relative effectiveness then depends on the size of the error for any particular value of T. Comparing the formulae for the equilibrium and constant growth conditions we find that the relative error depends on the choices of α, γ_0 and γ_1. We therefore undertake a grid search to examine the errors numerically for all three conditions in order to derive some general conclusions about which is likely to produce the minimum error. A sample of the results is shown in Table 4.1. From this exercise we find the following results:

TABLE 4.1: The accuracy of alternative terminal conditions: the saddlepoint case.

Table shows error in terminal period value as a percentage of the analytical solution.

Model Coeffs.	Terminal Condition	Time horizon 5	10	20	50	100
α=0.95	Equilibrium	0.55	0.04	0.00	–	
γ_1=0.6	Levels	–4.15	–0.32	0.00	–	
	Growth	166.86	4.51	0.03	0.00	–
α=0.8	Equilibrium	11.32	2.47	0.14	0.00	–
γ_1=0.8	Levels	–8.49	–1.85	–0.10	0.00	–
	Growth	19.36	2.74	0.14	0.00	–
α=0.6	Equilibrium	216.26	112.37	46.37	7.30	0.53
γ_1=0.95	Levels	–16.22	–8.43	–3.48	–0.55	–0.04
	Growth	12.69	3.11	0.74	0.08	0.01

Notes

Model:
$$\hat{p}_t = \alpha \hat{p}_{t+1} + u_t$$
$$u_t = \gamma_0 + \gamma_1 u_{t-1}, \quad \gamma_0 = u_0 = 10.0$$

Equilibrium terminal conditions use $\hat{p}_{T+1} = (1-\alpha)^{-1} u^*$.

Levels terminal conditions use $\hat{p}_{T+1} = \hat{p}_T$

Growth terminal conditions use $\hat{p}_{T+1} = \hat{p}_T^2 / \hat{p}_{T-1}$.

(I) Equilibrium terminal values are the most effective in minimising terminal period error only when α is near unity and γ_1 is away from unity.

(II) Constant growth conditions are relatively effective only when γ_1 is near unity and α is away from unity.

(III) Constant level conditions are generally the most effective other than the specific cases noted above.

On the basis of these results we conclude that where a model tends to a steady state, constant level terminal conditions are likely to be the best choice. If the model jumps very quickly to the steady state then equilibrium values, if they can be pre–calculated, are a better choice. If the model is sluggish in moving to equilibrium then constant growth conditions become the best choice. These special cases occur when the solution trajectory is dominated either by the smallest unstable root or the largest stable root respectively.

<u>Case (b) The steady–state growth solution $|\gamma_1|>1.0$, $|\alpha\gamma_1|<1.0$</u>

In this case, p_T will be tending to a steady state growth path of γ_1 as shown above by equation (4.24). We do not therefore envisage applying constant level or equilibrium value conditions. Two relevant alternatives are:

(i) applying the equilibrium growth rate,

(ii) applying a constant growth condition.

(i) Equilibrium growth rate $\dot{p}_{T+1} = \gamma_1 \dot{p}_T$

Hence

$$\dot{p}_T = (1-\alpha\gamma_1)^{-1}u_T = (1-\alpha\gamma_1)^{-1} \gamma_1^T[u_o - \gamma_0(1-\gamma_1)^{-1}] + \gamma_0(1-\gamma_1)^{-1}. \quad (4.30)$$

The error is then:

$$\begin{aligned} p_T - \overset{*}{p}_T &= e_T = \alpha(p_{T+1} - \overset{*}{p}_{T+1}) \\ e_T &= \alpha\,\gamma_0(1-\gamma_1)^{-1}\,[(1-\alpha)^{-1} - \gamma_1(1-\alpha\gamma_1)^{-1}]. \end{aligned} \tag{4.31}$$

Therefore e_T is constant for all T. As $T \to \infty$; $p_1 \to 0$ since we recall that $p_1 = \alpha^T e_T$. Hence using the equilibrium growth rate delivers the required insensitivity property.

(ii) Constant growth terminal conditions $\overset{*}{p}_{T+1} = \overset{*}{p}{}_T^2/\overset{*}{p}_{T-1}$

Constant growth terminal conditions yield the same error formula as in the saddlepoint case (b). No useful expression emerges beyond the solution:

$$\overset{*}{p}_T = (1-\alpha\, u_T/u_{T-1})^{-1}u_T. \tag{4.32}$$

Numerical simulations have been undertaken for a variety of parameter values. A selection of the results is shown in Table 4.2. These results show that the error under constant growth assumptions also tends to a constant and thus delivers the insensitivity property. However the long–run error is different from the constant error induced by using the equilibrium growth rate. The simulations further reveal that constant growth conditions perform relatively well when γ_1 is close to unity i.e. the condition $|\alpha\gamma_1|$ is easily satisfied. As $|\alpha\gamma_1| \to 1.0$ equilibrium conditions become relatively more accurate.

Finally it should be noted that constant growth conditions can produce very large errors when the number of time periods is very small. This can be traced to the term in $(1-\alpha\, u_T/u_{T-1})^{-1}$ in equation (4.32). If α and all values of u_t, t=0,... are positive then p_t should also be positive. However it is possible that $(u_T/u_{T-1})>(1/\alpha)$ for short time horizons. This would force p_T to be negative! This result is confirmed by numerical simulation. Analytically it can be derived by considering the growth rate of u_t:

$$u_t \,/\, u_{t-1} = [\gamma_1^t u_0 + \gamma_0(1-\gamma_1)^{-1}(1-\gamma_1^t)] \,/\, [\gamma_1^{t-1}u_0 + \gamma_0(1-\gamma_1)^{-1}(1-\gamma_1^{t-1})]. \tag{4.33}$$

TABLE 4.2: The accuracy of alternative terminal conditions: the steady–state growth case.

Table shows error in terminal period value as a percentage of the analytical solution.

		Time horizon				
Model Coeffs.	Terminal Condition	5	10	20	50	100
$\alpha=0.75$	Equilibrium	–23.20	–8.53	–1.32	–0.01	0.00
$\gamma_1=1.2$	Constant gr	–9885.96	28.63	2.81	0.01	0.00
$\alpha=0.75$	Equilibrium	–35.19	–17.43	–7.75	–1.34	–0.11
$\gamma_1=1.05$	Constant gr	92.09	14.77	3.28	0.35	0.03

Notes

Model:

$$p^*_t = \alpha\, p^*_{t+1} + u_t$$

$$u_t = \gamma_0 + \gamma_1 u_{t-1}, \quad \gamma_0 = u_0 = 10.0.$$

Equilibrium terminal conditions use $p^*_{T+1} = \gamma_1\, p^*_T$.

Constant growth terminal conditions use $p^*_{T+1} = p^{*2}_T / p^*_{T-1}$.

As $t \to \infty$ this growth rate tends to γ_1 if $|\gamma_1|>1.0$ and hence if $|\alpha\gamma_1|<1.0$ the term $|\alpha u_T/u_{T-1}|<1.0$. However, for small T and small values of u_0, the growth rate will be close to $(1+\gamma_1)$ – enough to make $\dot{p}_T$ go negative. The numerical simulations reveal a discontinuity in the terminal period error as a function of increasing T after which point the performance of the condition rapidly improves. If $|\gamma_1|<1.0$ the growth rate tends to 1.0 but the problem for short time horizons remains. Hence we need to be careful about using constant growth terminal conditions when the time horizon is short or the exogenous variables have not settled to a growth rate close to their equilibrium. This conclusion helps to explain why constant growth conditions perform less well as $|\alpha\gamma_1| \to 1.0$ since under these circumstances this short run problem is more likely to emerge.

To summarize the results so far we find that constant growth conditions perform reasonably well for systems which produce a steady–state growth rate or which exhibit sluggish adjustment to equilibrium. The time horizon must be reasonably long and the unstable roots not too close to unity. In systems producing a steady–state level equilibrium, then constant level conditions are likely to be more robust. In both cases, if the smallest unstable root is too close to unity then the system will be very sensitive. In such cases a long time horizon is required and any information concerning the equilibrium solution will help to reduce sensitivity. In a model of mixed growth and level equilibria for different variables then we obviously need a mixed set of terminal conditions. If the nature of the model solution is not known then a first guess may be made by examining the order of integration (Granger, 1981) of the individual variables over an historical period. We might select growth conditions for I(1) (or non–stationary) variables and level conditions for I(0) (or stationary) variables.

4.3 The implications of terminal conditions for model solution

We now consider the implications of alternative terminal conditions for solving a model when the system roots do not possess the required stability properties. In an empirical setting we may be unaware whether the model that has been constructed is stable or not. Sensitivity tests must therefore be used not only to validate the choice of terminal condition but also to establish the stability of the model. This procedure is especially relevant for nonlinear systems for which we cannot obtain dynamic roots.

If $|\alpha|>1$ in our demonstration model then there is no saddlepoint solution. Instead, there is an infinite number of stable solutions and the analytical solutions (4.13, 4.19) are not defined (since α^i, is explosive). However the use of a fixed terminal value will select one of the infinite number of stable paths. Consider the two–period solution using either a fixed terminal value ($p_3=\bar{p}_3$) or a constant level terminal condition ($p_3=p_2$):

$$\begin{bmatrix} 1 & -\alpha & 0 \\ 0 & 1 & -\alpha \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ \bar{p}_3 \end{bmatrix} \tag{4.34a}$$

$$\begin{bmatrix} 1 & -\alpha & 0 \\ 0 & 1 & -\alpha \\ 0 & -1 & 1 \end{bmatrix} \begin{bmatrix} p_1 \\ p_2 \\ p_3 \end{bmatrix} = \begin{bmatrix} u_1 \\ u_2 \\ 0 \end{bmatrix} \tag{4.34b}$$

We consider solving this system by a first–order iterative technique. The three schemes for achieving consistent expectations outlined in Chapter 3 all coincide when there is only one equation. The Jacobi iteration matrices are respectively:

$$\begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & \alpha \\ 0 & 0 & 0 \end{bmatrix} \text{ and } \begin{bmatrix} 0 & \alpha & 0 \\ 0 & 0 & \alpha \\ 0 & 1 & 0 \end{bmatrix}. \tag{4.35}$$

The spectral radii of the two systems are 0 and $\sqrt{\alpha}$ respectively. The stability condition is $|\alpha|<1.0$ in which case $\sqrt{\alpha}<1.0$. Hence the convergence conditions of the first–order iterative technique using constant level terminal conditions coincide with the those for the existence of a saddlepoint. There is in fact a solution with a constant level at the final period defined by $p_2=(1-\alpha)u_2$, $p_1=\alpha(1-\alpha)u_2+u_1$ but this solution is not uniquely stable and the algorithm will not converge to these values. On the other hand, if a fixed terminal value is used then convergence is always achieved. This result generalizes to simultaneous systems and longer time periods. If fixed values are used one may therefore select a solution to this model even when there is more than one stable solution and this non–uniqueness will not be apparent. Hence the use of such conditions may be misleading. This result has an analogy in the solution of a linear backward–looking model. Given a fixed initial condition it always possible to solve such a model for a finite time horizon even when it is unstable.

We would like to able to generalise these results to a system of mixed leads and lags and to constant growth terminal conditions. However, there seems to be no such generalisation which can be derived analytically for an arbitrary, finite time horizon. Even for two–period solutions, the roots of the system and the spectral radius of the iteration matrices become highly complicated. Furthermore, the use of constant growth terminal conditions introduces a nonlinearity which precludes the use of linear algebra and hence we need to resort to numerical simulation experience.

We use a slightly more general version of our demonstration model which includes a lag term as well as a lead:

$$p_t = \alpha p_{t+1} + \beta p_{t-1} + u_t - u_{t-1} \tag{4.36}$$

Our results in this section are independent of the precise form in which u enters the equation. This equation is similar to the reduced form of Muth's 1961 small macro–model (equation 4.12, p325). It is also similar to the exchange rate equation in some vintages of the NIESR model. We solve different parameterisations of this system for the range of terminal conditions, for different time horizons and for different bases.

(i) Fixed terminal values

Using arbitrary, fixed–value conditions and a Fair–Taylor solution algorithm, this model solves for all choices of $0<\alpha,\beta<1$ including those combinations which do not satisfy the saddlepoint condition. However, those models involving parameter choices which do not satisfy the saddlepoint conditions become relatively more difficult to solve as the time horizon is extended. Furthermore, the solution itself is sensitive to the time horizon in these non–saddlepoint cases.

(ii) Constant level conditions

Using constant level terminal conditions, parameter combinations leading to both roots being greater than unity (or both less than unity) result in explosive iterative paths and hence fail to solve. Hence we can surmise that a general nonlinear system is likely to fail to solve due to numerical problems in similar circumstances. However, for short horizons (e.g. less than ten periods) a solution is produced for models which contain one or two unit roots ($\alpha+\beta=1$). These solutions, with one exception discussed below, are sensitive to the simulation horizon and therefore non–unique.

When one root is unity and the other greater than unity (e.g. $\alpha=0.4$, $\beta=0.6$) the constant level terminal condition produces a solution which appears to be stable and unique. This special case arises because the initial condition is fixed and hence the unit root, which is present instead of a stable root, does not cause the solution to fail and there is a unique solution for a given initial condition. This solution is then attainable and unique for any time horizon.

(iii) Constant growth conditions

Using constant growth terminal conditions, a solution is achieved in only two cases: when a saddlepoint property holds in the model (e.g. $\alpha=\beta=0.4$) or in the special case noted above when there is one unstable root and one unit root.

On the basis of these experiments, summarised in Table 4.3, constant growth terminal conditions are more consistent than constant level conditions in revealing certain non–saddlepoint cases. For long time horizons, either choice of terminal condition tends to non–convergence in these unit root models. These conclusions reflect the fact that in very short horizons the terminal condition has an undue influence on the spectral radius of the iteration matrix and can reduce it to less than unity for some non–saddlepoint models. Constant growth conditions do not appear to suffer from this problem in the same way as constant level conditions.

As a result of these findings, the unit root model appears to be an interesting special case. If there exists a unit root instead of an unstable root (e.g. a single equation random walk specified as a forward difference) then there is no stable solution but no explosive solutions either: there are an infinite number of possible solutions depending on the choice of terminal value. In this case we can only map a trajectory from a given initial condition to a given terminal value: constant growth and (ultimately) constant level conditions both fail to solve. The terminal condition must therefore be a fixed value and this will then define a unique solution path. This conclusion should not be surprising since even a lag model with a random walk implies that we cannot predict a unique long–run value of a variable independent of the initial conditions. However, because the terminal condition is arbitrary numerically, it does not have to be arbitrary economically. We can still use *a priori* reasoning to choose a likely path and specify a terminal value.

4.4 Empirical results: Applications to large–scale models

In order to examine stability by simulation methods, we solve three nonlinear models using different solution periods and terminal conditions to assess the sensitivity of the solution. If a model is stable, the solution period sufficiently long and the terminal condition appropriate, then the solution over the period of interest will not be affected by increasing the sample length. This principle is built into the solution method of Fair and Taylor (1983) as their Type III iterations in which the model is continually re–solved over a solution path which is gradually increased

TABLE 4.3: Summary of convergence results for alternative terminal values

Model : $p_t = \alpha p_{t+1} + \beta p_{t-1} + u_t - u_{t-1}$

	First root		
Second root	$\lambda_1<1$	$\lambda_1=1$	$\lambda_1>1$
$\lambda_2<1$	F[1] solves[2] L,G fail	F,L solve G fails	F,L,G solve
$\lambda_2=1$	F,L solve[3] G fails	F,L solve[5] G fails	F,L,G solve
$\lambda_2>1$	F,L,G solve[4]	F,L,G solve[6]	F solves[7] L,G fail

Notes

1 F – Fixed terminal value i.e. $p_{T+1}=p^*$
L – Levels terminal condition i.e. $p_{T+1}=p_T$
G – Growth terminal condition i.e. $p_{T+1}=p_T p_T/p_{T-1}$

2 e.g. $\alpha = 0.7$, $\beta = 0.6$, $|\lambda_1| = |\lambda_2| = 0.925$

3 e.g. $\alpha = 0.4$, $\beta = 0.6$, $\lambda_1 = 0.66$, $\lambda_2 = 1.0$

4 e.g. $\alpha = 0.4$, $\beta = 0.4$, $\lambda_1 = 0.5$, $\lambda_2 = 2.0$

5 e.g. $\alpha = 0.5$, $\beta = 0.5$, $\lambda_1 = 1.0$, $\lambda_2 = 1.0$

6 e.g. $\alpha = 0.6$, $\beta = 0.4$, $\lambda_1 = 1.0$, $\lambda_2 = 1.5$

7 e.g. $\alpha = 0.6$, $\beta = 0.7$, $|\lambda_1| = |\lambda_2| = 1.08$

until no sensitivity is observed. That process (known as the Extended Path algorithm) is not used generally by models of the United Kingdom economy for every solution, partly because of its obvious cost implications. It can also distract attention from the choice of terminal condition. Fair and Taylor simply set the terminal value from a base (control) solution. If the model is unstable then Fair and Taylor's Type III iterations should not converge. Our own sensitivity analysis effectively repeats the Extended Path procedure on a one–off basis and examines every solution. We also consider variations in the terminal conditions to see whether different types of condition reveal different properties. Once the stability properties of the model are established, the user just needs to keep a watchful eye on the simulations, as discussed in (iv) below.

(i) Liverpool model

By default, the Liverpool (LPL) model sets the values of the terminal conditions in accordance with equilibrium conditions derived from a theoretically derived linear representation of the model. Although discussion of the general problem and the general nature of these conditions usually appears in presentations of the model (e.g. Minford *et al*, 1984), the actual conditions employed have been published only by Wallis *et al* (1985). Three points are worth noting about the LPL model structure and its terminal conditions. Firstly, no dependent variable in the structural form of this model has its own expectation as an explanatory variable. This renders single equation analysis of the dynamics of even less relevance than usual. Secondly, it should be noted that the terminal conditions do not constrain the solution to pass smoothly through the terminal point or imply that the solution would remain at the terminal values after the terminal date if the time horizon were extended. Hence equilibrium terminal conditions do not actually constrain the model solution to reach a stable equilibrium. For example, although capacity utilisation (defined as the deviation of output from trend) is constrained to zero in period T+1, it is freely determined by the model in period T. Finally, how long the model takes to reach equilibrium is an open question and the model is generally

TABLE 4.4: Expectational variables and terminal conditions: LPL model

Expectational variables	Equation in which expectation appears
Capacity utilization	Non–durable consumption
Real debt interest	Equilibrium government spending
Real exchange rate	
1 year ahead	Real short interest rate
5 years ahead	Real long interest rate
Inflation	
1 year ahead	Nominal short interest rate
Average 5 years ahead	Nominal long interest rate

Terminal conditions
(T denotes the last time period of the solution)

Capacity utilization	Measured as deviation from trend output and projected at zero
Real debt interest	Projected by its value in period T times the growth of equilibrium GDP at T
Real exchange rate	Projected 5 periods after T at a constant value given by multiplying the value of the equilibrium level of the exchange rate at T by its growth rate at T.
Inflation	Projected 5 periods at constant value given by the value of the exogenous PSBR/GDP ratio at period T

observed to be rather sluggish.

Four variables have expectations which appear in the model: capacity utilization (deviation from trend), inflation, real debt interest and the real exchange rate. Two of these, inflation and the exchange rate, appear twice with different lead lengths, in order to determine both short and long rates of interest (both real and nominal). The terminal conditions for inflation and capacity utilization are set by exogenous variables, whereas the other two terminal conditions are generated from the paths of 'long–run equilibrium' variables which are determined endogenously. These conditions are detailed in Table 4.4. It is not clear, however, that they specifically represent the long–run solution of the empirical version of the model rather than that of the theoretical model on which they are based. For example, in the empirical model the equilibrium rate of inflation is equal to the long–run PSBR/GDP ratio (which is exogenous and yields the terminal inflation value) only if the rate of growth of real money demand is zero. Since the latter is determined by an equation which contains a time trend and the level of GDP (which has a non–zero equilibrium trend), this seems unlikely to occur.

Using the Autumn 1984 version of the model, we solve the model reducing the time horizon by annual increments. The same conclusions obtain whether the period is shortened or lengthened. Solution paths for real debt interest and the real exchange rate are shown in Figure 4.1. Similar response patterns occur in all the other endogenous variable trajectories.

Two main conclusions emerge from this exercise. First, there does seem to be a unique long–run path to which the early parts of the solution provide a good approximation. As the solution period is increased, this becomes insensitive to the terminal date. Second, the last five periods of each solution show marked and consistent deviations from this path. This divergence begins in the period at which the terminal conditions have a direct effect. This produces a *prima facie* case that the equilibrium terminal conditions represent a poor approximation to the long–run stable solution path. A model user would have little confidence in the results during this divergent period (Liverpool themselves only report forecast solutions for five

Figure 4.1 LPL model, 1984: alternative solution periods

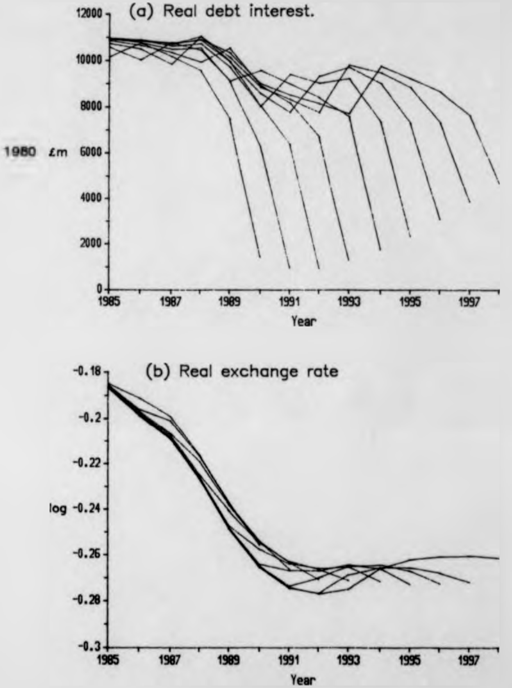
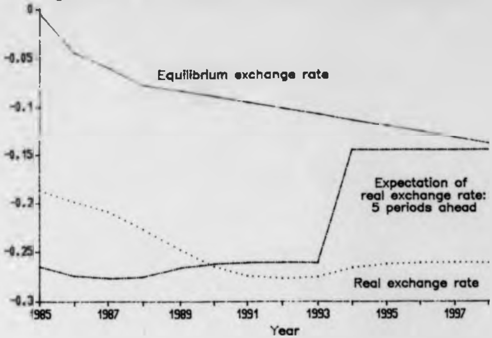


Figure 4.2 LPL model: exchange rate trajectories



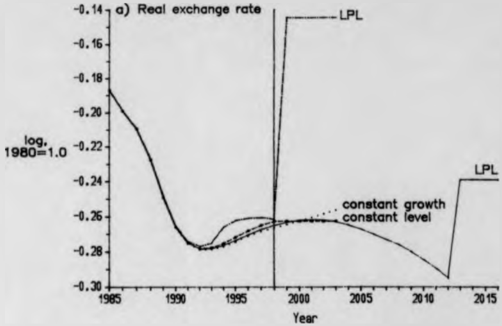
years ahead).

The reason for the poor performance of the equilibrium conditions is given by the behaviour of the real exchange rate. Whereas the terminal condition for the real exchange rate is derived from the equilibrium real exchange rate equation, the actual real exchange rate shows a persistent divergence from the equilibrium level, this being greater than the latter value by a factor of two. These paths are shown in Figure 4.2 for a 14 year horizon. The implication of this gap between actual and equilibrium values is that the expected real exchange rate 'jumps' in year 9 (1994), leading to the dynamic behaviour of endogenous variables over the last five years of the solution noted earlier. The gap between the actual and equilibrium exchange rate can be traced back to the value of the first–order serial correlation coefficient (0.9) in the exchange rate equation. Hence the model is too sluggish to achieve equilibrium in fourteen years.

In Figure 4.3 we show comparative results for the application of different types of terminal condition to the model. Several solutions over a short period are compared with a longer run solution in which the LPL equilibrium conditions were applied. It is apparent that the LPL terminal conditions applied to the shorter forecast, do not perform well, exhibiting a substantial divergence for up to six years prior to the terminal date. Terminal conditions based on constant growth rates or constant levels do much better, with negligible distortion. Examining all variables, it becomes apparent that some are better served by constant growth rates and some, like the exchange rate, by a constant level depending on the long–run trajectories of the conditioning variables.

In the Autumn 1985 version of the model, corrections are made to the equilibrium terminal conditions in such a way as to prevent the real exchange rate jumping to the equilibrium level. Instead it jumps to the same rate of growth as the equilibrium exchange rate. The resulting sensitivity analysis, for real debt interest and the real exchange rate, is shown in Figure 4.4. Clearly the sensitivity of this solution path has been almost entirely removed and this result is reflected in other variables. However, the nature of the equilibrium terminal conditions on other

Figure 4.3 LPL model: alternative terminal conditions



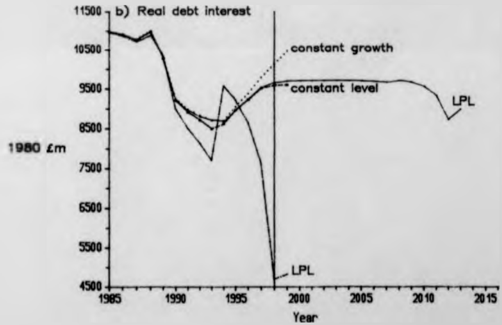
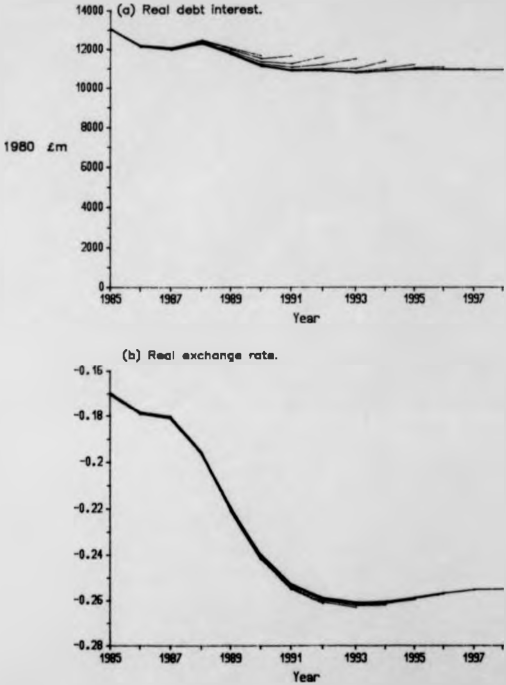


Figure 4.4 LPL model, 1985: alternative solution periods



variables raises further potential problems in simulation analysis. The sensitivity may reappear for different simulations and this is discussed in (iv) below.

(ii) LBS model

The LBS model contains three expectation terms in its financial sector. The price of gilts has a constant level terminal condition, the prices of equities and overseas assets both have constant growth rate conditions. This information is summarized in Table 4.5. In Figure 4.5a we present some sensitivity analysis focusing on the price of gilts which is representative of all three variables.

The first obvious conclusion is that the sensitivity has a seasonal pattern i.e. the trajectory depends on which quarter of the year the simulation takes as an end period. The price of gilts clearly has a seasonal pattern over its entire base solution. Failure to predict this seasonal component leads to drastic over and undershooting depending on which quarter of the year the solution finishes in. This effect is transmitted through the whole model so that even GDP has a seasonal pattern.

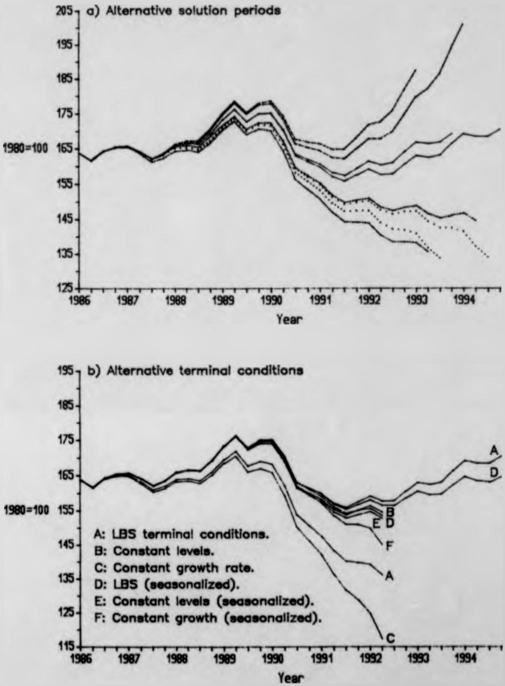
The LBS model is largely constructed using seasonally adjusted or theoretically non–seasonal data (e.g. the financial sector data are not adjusted but should not be seasonal). The seasonal pattern in the base solution is therefore somewhat surprising. In general, seasonality can be induced in a model by the inputs or the equation dynamics. In this model, the presence of seasonality can be traced back to the residual adjustments, exogenous variables, seasonal dummies initial conditions for lagged endogenous variables, and fourth–order equation dynamics. Of particular importance is the uprating of tax allowances and local authority rates in line with the annual inflation rate in each budget quarter. Strictly, such uprating ought to be seasonally adjusted before input into a model forecasting seasonally adjusted data.

If a saddlepoint path exhibits seasonality then the terminal conditions must be adjusted to take account of the fact. The LBS model does seem to possess a saddlepoint path as reflected in the lack of sensitivity of the first half of the solution. In a quarterly model, possible adjustments to the constant level and

TABLE 4.5: Expectational variables and terminal conditions: LBS model

Expectational variables	Terminal conditions
Gilt price	Constant level, quarter on quarter
Equity price	Constant rate of growth, quarter on quarter
Overseas assets price	Constant rate of growth, quarter on quarter

Figure 4.5 LBS model: price of gilts



constant growth conditions could therefore take the form $y_{T+j} = y_{T+j-4}$ and $y_{T+j}/y_{T+j-4} = y_{T+j-4}/y_{T+j-8}$ (or $y_{T+j}/y_{T+j-k} = y_{T+j-4}/y_{T+j-k-4}$) respectively. These conditions are expressed in annual terms and so will accurately reproduce the seasonal component, adjusting the terminal value accordingly and eliminating the seasonality in the sensitivity results.

In Figure 4.5b we show some alternative solutions to the LBS model for a short time horizon in which a variety of terminal conditions are applied. We apply three types of condition: the LBS mixture, constant levels, and constant growth rates (results labelled A, B and C respectively). We then apply the correction for seasonality to each in turn (labelled D, E, F). Terminal conditions in which seasonality is taken account of clearly perform better than those in which it is not. Of the latter, the constant levels version appears to do remarkably well but this reflects the seasonal pattern at the chosen terminal date and a turning point in the series which fortuitously occurs at that point. This turning point is caused by the fact that the base is a genuine *ex ante* forecast and has been adjusted carefully for the period of interest. Hence the residual adjustments are much fewer and much smoother after the first five years.

(iii) NIESR model

The NIESR model contains 40 expectations terms in 11 variables with leads of up to four periods ahead. These variables and their terminal conditions are detailed in Table 4.6. By far the most important of these variables is the real exchange rate. The other 10 variables all use constant growth rate terminal conditions. In this version (model 8), the NIESR model has an imposed unit root in the real exchange rate equation which can be written:

$$\ln \rho_t = 0.102 \ln \rho_{t-1} + (1-0.102) \ln \rho_{t+1|t-1} + 0.009\, \Delta\, d, \qquad (4.25)$$

where ρ is the real effective exchange rate and d is the real UK:Rest of the World interest rate differential. Since d is exogenous and no other variable enters the

TABLE 4.6: Expectational variables and terminal conditions: NIESR model

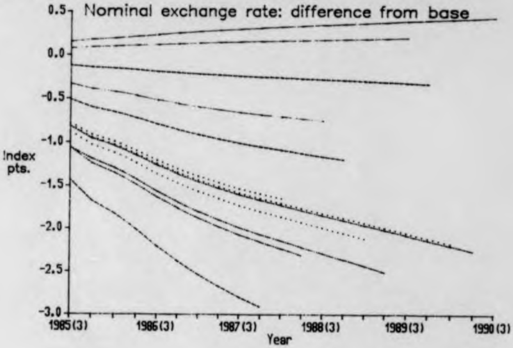
Expectational variables
All have leads of up to four quarters ahead: Output index (GDP) Output index (manufacturing) Output index (other) Output index (mainly public) Consumer price index Wholesale price of manufactures Average earnings Personal disposable income Interest rate on local authority debt Nominal effective exchange rate Rate of employer's national insurance contributions
Terminal conditions
All use a constant rate of growth, quarter on quarter, except for the nominal effective exchange rate, which is calculated from : $\log(\epsilon\, p/p^f)_{T+1} = \log(\epsilon\, p/p^f)_0 + 1.75\,[\log(X/M)_T - \log(X/M)_0]$ where ϵ is the nominal exchange rate, p is the wholesale price of manufactures, p^f is the foreign wholesale price (implicitly projected at a constant growth rate) X is total exports, M is total imports, T is the last solution period and 0 the last historical period.

equation, the single equation roots deliver the system roots of 1.0 and 0.114 . The system therefore has a unit root where it should have a root outside the unit circle to meet the saddlepoint condition. The model therefore has no unique stable solution and (consistent with the results of Section 4.4) fails to solve for constant growth terminal conditions. Any fixed terminal value supplied will, however, generate a solution to the model. Sensitivity analysis is not necessary to come to this conclusion. The NIESR model imposes a terminal condition which links the exchange rate to its initial condition (y_0) and any long–run change in the current trade balance (see Table 4.6). This condition ensures that the exchange rate moves in the long run to (partly) clear the current trade balance.

In examining the sensitivity of the model we are handicapped by a large residual adjustment to the exchange rate in the final period of the forecast, which ensures that the exchange rate is roughly constant over the forecast. In changing the terminal date we usually assume implicitly that the exogenous inputs are on a stable trajectory over the horizon covered by the different end–points. This forecasting residual denies that assumption and introduces sensitivity to the time horizon. Removing the residual causes the model to fail to solve for the full solution period. We therefore remove the residual and shorten the horizon until the model solves successfully. One period shorter is sufficient and this new solution serves as a base from which we consider deviations.

Figure 4.6 demonstrates the sensitivity of the exchange rate solution to changes in the terminal date. Similar patterns are repeated for all other endogenous variables due to the common characteristic equation. Three results stand out. First, the sensitivity of the solution is not confined to the final few periods but affects even the first period. This is symptomatic of our demonstration examples but not of the LPL and LBS models. The result for the NIESR model reflects the pervasive influence of the exchange rate equation, coupled with the unit root. A second result is the overall degree of the sensitivity: up to 3 points on the nominal exchange rate, 0.5% on GDP and up to 5 points on the price level. Finally, the results are clearly seasonal as we found for the LBS model. As in the LBS case, the

Figure 4.6 NIESR model: alternative solution periods



NIESR model is constructed largely with seasonally adjusted data and the seasonality reflects the exogenous inputs and equation dynamics. The most important factor is again the annual uprating of tax allowances in the budget quarter.

Because the unit root allows any choice of terminal value, no alternatives are tried here. We can rule out conditions such as a constant growth rate for the real exchange rate although a constant level condition should yield an arbitrary solution for short time horizons. With a unit root, the terminal condition must embody sufficient exogenous information to tie down the terminal value.

Later vintages of the NIESR model (e.g. Model 9, Autumn 1986) incorporate two corrections. The exchange rate equation is altered to include an endogenous current account term thus altering the system root of unity. The terminal condition for the exchange rate has therefore changed and a constant growth assumption is now employed. However, probably due to a near–unit root remaining, the time horizon for the current model is required to be very long (up from 23 periods to 40) and an element of an equilibrium condition remains in that the exchange rate is required move so as to eliminate any change from the base level of net overseas assets.

The terminal conditions are also changed in that the average of the quarter–on–quarter growth rates over the previous year is used as the quarter–on–quarter growth rate at the terminal date. This change successfully removes most of the seasonality in the sensitivity analysis but not all. The quarter–on–quarter growth rate terminal condition is not the original source of the seasonality in the model but acts as a transmission mechanism. The correction is only intended to remove the seasonality in the sensitivity results not the underlying seasonality in the models dynamic solution. Using the average growth rate does not accurately reflect the differing quarter–on–quarter growth rates in different quarters of the year whereas our proposals made above do capture this variation.

(iv) Simulation responses

In simulations with these models, any sensitivity with respect to the terminal date affects both the base and perturbed solutions and the issue is whether these effects cancel out. It is also possible that different terminal conditions yield different responses and this possibility also needs to be examined.

Using the LPL model with an amended terminal condition for the real exchange rate, there is no sensitivity to the terminal date in the simulation response. The main question for this model is therefore whether the continuing use of some equilibrium values for other terminal conditions affects the simulation properties. In particular the terminal value of inflation is set by an exogenous variable proxying the equilibrium PSBR/GDP ratio (see Table 4.4). We consider three types of simulation, retaining the LPL terminal conditions. Firstly, we have simulations which involve a change in this exogenous variable and hence automatically adjust the terminal value. Secondly we examine simulations which do not adjust the terminal value explicitly but do not imply a long–run change in the inflation rate. Finally we consider simulations in which there is an implied long–run change in the inflation rate but for which the exogenous terminal value remains constant.

(a) A standard government expenditure shock simulation, under balanced financing rules is introduced by changing the exogenous PSBR/GDP ratio and hence the terminal values are directly re–calculated. Consequently we observe no sensitivity of the simulation properties to changes in the solution horizon. Furthermore, we observe no major differences when the terminal condition is altered to a constant growth rate or constant level. For this simulation the explicit adjustment to the terminal value appears to be accurate.

(b) The same conclusion is reached for any simulations which do not imply a change in the long–run inflation rate. Typically this includes all temporary shocks, tax–financed government expenditure changes or a shock to world trade. In all of these simulations there is no substantial long–run effect on real money growth which is the only endogenous component in the inflation equation.

(c) Equilibrium conditions are found to affect the results for any shock which implies a rise in the long–run inflation rate and which does not involve a direct change to the terminal condition e.g. a permanent government expenditure increase financed by monetary growth. In these cases, the equilibrium terminal conditions force the inflation rate to return to base at the final period, thus distorting the response of the model. The interpretation of such a simulation result is that government expenditure and monetary growth return to base in the final period. Even for this interpretation, the result is distorted because the equilibrium is being imposed immediately upon cessation of the shock.

If we wish to examine a shock for which the inflation rate is allowed to rise in the long–run, we can allow the price level to be determined by constant growth terminal conditions. This implies that the equilibrium can be characterised by any stable inflation rate rather than a particular value of inflation. Sensitivity analysis then reveals that a permanent increase in government expenditure, financed by monetary growth is neither distorted, nor sensitive to the time horizon.

In some simulations even constant growth terminal conditions are insufficient to guarantee that the solution corresponds to that of the desired shock. A permanent reduction in the VAT rate which is allowed to permanently affect the PSBR, which in turn is financed by balanced monetary growth, implies an accelerating inflation rate in the LPL model. In such a simulation the exogenous PSBR/GDP ratio is explicitly changed whereas government expenditure and tax rates remain constant. The terminal condition is thus explicitly changed, but the implication of this change is that inflation attains a stable level in the final period. Constant growth conditions on the price level yield a very similar solution in which inflationary growth is truncated at the terminal date. Constant growth conditions on inflation (rather than the price level) do not yield a solution. This latter result delivers the correct conclusion that the model is being used in such a way that it becomes globally unstable. It should not therefore yield any solution purporting to be stable. A model user must be aware of, and watch out for any circumstances in which the terminal condition appears to be inconsistent with the dynamic trajectory of the simulation response. Ultimately simulations must be restricted to those for which a saddlepoint solution exists.

This particular example reflects the fact that the simulation involves more than a simple shock. If we simply add a small perturbation to an exogenous variable, the stability properties of the model should be preserved — if it is stable in the base, it is almost always stable in a small neighbourhood. If, however, the exogenous variable is put on an unstable trajectory, or used as an instrument to target an endogenous variable, the stability properties may alter. In the example given above, the VAT rate is increased and the PSBR is allowed to vary. This is achieved by targeting the level of government expenditure, otherwise endogenous, using the exogenous PSBR/GDP ratio as an instrument. This prevents government expenditure adjusting to return the PSBR to its base level. This additional rule, for this shock, makes the model unstable.

In the LBS model the seasonality of the solution discussed above leads to a modest degree of sensitivity to the terminal date and to the terminal condition. Amending the terminal conditions to take account of seasonality successfully removes this sensitivity. The conditions employed then appear to be valid for all interesting simulations (e.g. all those reported in Fisher *et al.*, 1988, 1989).

In the unit root version of the NIESR model the terminal condition is arbitrary and must be selected using off–model information. Hence different simulation properties could be revealed by different assumptions. For example, assuming a long–run response of the exchange rate to the trade balance gives quite different multiplier properties. However the effects of seasonality on the simulation results can again be removed by appropriate amendments as discussed above. In later vintages of the model, with the unit–root problem resolved, constant growth conditions appear to be appropriate for most of the variables but a long time horizon appears to be necessary to remove sensitivity. Given our results above that the performance of constant growth conditions deteriorates as the smallest unstable root tends to unity, and that the required time horizon therefore increases, it may be that equilibrium–type terminal conditions would be more appropriate for the exchange rate in the NIESR model.

4.5 Summary and Conclusions

In this Chapter we first derive the standard saddlepoint condition for a unique stable solution in a forward expectations model and discuss how terminal conditions are used to select such a solution in a finite horizon problem. The choice of terminal value has received scant attention in the literature and the issue does not seem to have been resolved. Minford *et al.* (1979, 1980) suggest the use of conditions which characterize their model's known equilibrium, whereas Wallis *et al.* (1985, 1986) recommend the use of rules base on constant growth rates or levels. Either choice is dismissed by some authors as arbitrary (e.g. Blake *et al.*, 1989). In this chapter we examine the degree of approximation involved in employing different types of terminal condition on a simple demonstration model and show that, in general, the terminal condition should reflect the long–run trajectory of the variables concerned: either constant level or constant growth rate conditions. In models with unstable roots near unity, the effectiveness of the terminal condition in approximating the saddlepoint is improved if the long–run solution or long–run growth rate is known since these provide better terminal values than those constructed from the solution itself.

We assess the implications of different choices of terminal condition for the numerical solution of models with particular emphasis on the non–saddlepoint case. In general we conclude that constant growth conditions only allow convergence when a saddlepoint solution exists, whereas fixed value conditions allow convergence in all cases. The latter are therefore more robust but their use may conceal the instability of the model.

We test out various terminal conditions on three large–scale nonlinear forecasting models. We find that the results of our analysis are confirmed, with constant growth terminal conditions working well on models with well–established saddlepoint solutions but breaking down for a unit–root model. Ultimately, the choice of terminal condition is dependent on the model structure. The most important conclusion is therefore that sensitivity testing is vital both to establish the stability properties of a model and to validate the terminal condition and time horizon employed.

Chapter 5

EXPERIMENTAL DESIGN AND STOCHASTIC SIMULATION

In this chapter, we begin by examining the consequences of different assumptions concerning input shocks in forward expectations models. The alternative assumptions concern whether the shocks are anticipated or unanticipated; temporary or permanent. This then leads us to a proposed method of stochastic simulation for such models. A series of stochastic shocks will generally be a sequence of unanticipated temporary shocks although other possibilities are not ruled out. Our proposed stochastic simulation method differs from two previous proposals in the literature. The reasons for these differences are critically assessed.

Numerical results for large–scale models using different assumptions about input shocks are presented in Wallis *et al.* (1986, Section 2.4). In Section 5.1, we develop a comprehensive analysis of how anticipated and unanticipated shocks can be introduced into a general dynamic model. This leads to a slightly wider range of possibilities than those considered by Wallis *et al.*. In Section 5.2 we extend the numerical results of Wallis *et al.* on the same large–scale models. In Section 5.3 we analyse the introduction of temporary and permanent shocks and combine this with the possibility of unanticipated changes to the duration of the shock. Again, this leads to a wider range of possibilities than those examined in Wallis *et al.*; numerical results are presented in Section 5.4.

In Section 5.5 we develop methods for stochastic simulation of large–scale forward expectations models. The method presented here is different from that used by Hall and Henry (1985a,b) and that suggested by Fair (1984, pp383–384). Those methods are shown to require extreme assumptions about the input shocks. In Section 5.6 we use stochastic simulation to evaluate the variability of output and the price level in the face of stochastic shocks. The experiments are repeated under alternative financing rules for the PSBR, so updating an earlier study by Fisher, Wallis and Whitley (1985) which employed stochastic simulation on models without forward expectations. We can therefore examine whether changes to the models, and the introduction of forward expectations in particular, have implications for those previous results. Section 5.7 contains a summary and concluding comments.

5.1 Anticipated and unanticipated shocks

Consider the following two demonstration models, the first with forward expectations the second with a lagged dependent variable:

$$y_t = \alpha y_{t+1|t-1} + u_t, \tag{5.1}$$
$$y_t = \beta y_{t-1} + u_t, \tag{5.2}$$

where the u_t term is the source of all input shocks. We then introduce an impulse shock δ to u in period s and examine the solution for y_t, t=1,...,T. In the lag model (5.2) the solution is unambiguous: there is no effect on y_t, t=1,...,s–1; y_s is increased by δ; y_{s+i} is increased by $\beta^i\delta$, i=1,...,T–s. Hence in this particular model, all the effects occur at or after the introduction of the shock.

The solution to the model (5.1) for consistent forward expectations is derived in Chapter 4 and is given by equation (4.15) which we re–state as:

$$y_t = \sum_{i=t}^{T-1} \alpha^{i-1} u_i + \bar{y}_T. \tag{5.3}$$

However, the values u_i, i=t+1,...,T–1 are those used to form $y_{t+1|t-1}$ and are therefore more properly regarded as the *expected* values at the end of period t–1 and usually only u_t at most is an observed value in equation (5.3). The effects of a particular shock are then dependent on an assumption concerning how expectations of the forcing variable u_t are formed. If the shock δ is anticipated at the end of period 0 (s>0), y_t, t=1,...,s will increase by an amount of $\alpha^{s-t}\delta$, and y_t, t=s+1,...,T will not change from base. Hence, in this particular forward expectations model, all the effects of an anticipated shock occur at or before the date of its introduction.

We next assume that δ becomes anticipated only at the end of period s–k. Let us re–write equation (5.3) as:

$$y_t = u_t + \sum_{i=t+1}^{T-1} \alpha^i u_{i|t-1} + \bar{y}_T. \tag{5.4}$$

First of all, consider the case of k=1 so that the shock is unanticipated prior to period t; then $u_{s|t-1}$, t=1,...,s–1 is unaffected by δ and hence so is y_t, t=1,...,s–1. The first period to be affected is y_s which increases by δ and then y_{s+i}, i=1,...,T–s is unchanged. Hence the only reaction to δ occurs in period s. Between the two extreme cases (k=1, k=s), we could assume that δ is anticipated in period s–k, 1<k<s. In these cases there will be a response of $\alpha^{s-t}\delta$ beginning in period s–k but the response in earlier periods will be zero as will be the response of y_{s+i}, i=1,...,T–s.

We therefore see that different assumptions concerning how far in advance the input shocks are anticipated generate different responses from a forward looking model. These assumptions determine, in just one respect, the expectations formation concerning the exogenous variables and stochastic disturbances rather than the endogenous variables. By definition, the exogenous variables and disturbance terms are not explained by the model structure. Therefore we can only assess the impact of alternative expectations assumptions using sensitivity analysis.

A general dynamic model possesses both lag and forward expectation terms, and the results provide a generalization of our demonstration examples. Let us consider a more general form of model:

$$B(L)y_t = A_1(F)y_{t+1|t-1} + C(L)x_t + u_t. \tag{5.5}$$

The terms in equation (5.5) are defined as for equation (4.9). We assume that there is a dynamic consistent expectations base solution y_t^b, t=1,...,T and that the residual terms have been set to appropriate values (for example, sample–period residuals, zero or the forecast adjustments). We then wish to introduce a shock, w_s, at period s, so that $x^*_s = x_s + w_s$. We divide up the possibilities into two alternative assumptions we can make concerning the date at which w_s becomes known: either w_s is anticipated at the end of period s–k, $1 \leq k \leq s$, i.e. before/at the date of the shock; or w_s is not anticipated until the end of period t=s–k ($0 \leq k \leq -T+s$) i.e some date after the shock is introduced.

In the first case ($1 \leq k \leq s$), the solution for periods t=1,...,s–k remains unchanged from base. We therefore re–calculate the solution from period s–k+1 and solve under consistent expectations for y_t, t=s–k+1,...,T, taking into account the shock w_s. We link the base solution for the horizon 1,...,s–k with this perturbed solution to give a complete solution over the horizon 1,...,T. The main question of experimental design in such a calculation is the value of k, i.e. for how many periods in advance is the shock w_s correctly anticipated. If k=1 then the shock is anticipated only as it occurs and we call this an unanticipated shock. In this first general case, the only difference in experimental design between the anticipated and unanticipated shocks is the start of the perturbed simulation horizon.

If $s-T \leq k \leq 0$ then we move to the case in which the shock is anticipated only at some point after the end of period s and is unanticipated even in the period in which it first directly affects y. In our simpler demonstration example (5.3) we assumed that the shock was always anticipated at least by the end of period t–1. Relaxing that assumption and setting k=0 for that model would produce a response equal to the shock in period s (since y_s depends directly on the actual value of u_s not its expectation) and zero for all other periods. Hence, assuming that k=0 is equivalent to assuming that k=1 for equation (5.4). This reflects the lack of lag terms in that model (hence δ does not affect $y_{s+1|s-1}$, nor the model solution for y_{s+i}, i=1,...,T–s). In the more general model (5.5) the solution reflects the presence of lagged terms; hence the response is non–zero after period s. Furthermore, the expectations formed at the beginning of period s depend on whether w_s is anticipated or not since the forward expectations $y_{s+i|s-1}$ are affected by anticipated changes in w_s.

In this second general case ($s-T \leq k \leq 0$) we will only consider $k=0$ for simplicity, so that w_s is always known by the end of the period in which it is introduced. The results for $k=0$ generalize easily to cover $s-T \leq k < 0$. We then observe that w_s affects y_s but not $y_{s+i|s-1}$, $i>0$. Here we need two solutions in addition to the base simulation. We use the base to determine $y_{s+1|s-1}$. We then solve for y_s given w_s and the base value of $y_{s+1|s-1}$. Finally we solve for y_t, $t=s+1,\ldots,T$; for consistent expectations given w_s, y_s. This is then a fully unanticipated shock. This second kind of shock was not considered in the exercise of Wallis *et al.* (1986).

We can extend this analysis to a series of shocks either interdependent or independent. A series of interdependent shocks is a sequence w_p, $p=s,\ldots,T$ which can be anticipated after the first shock in the sequence is realised. Hence w_{s+j} becomes known at the same time as w_s for all $j>0$. In case one ($k>0$) we simply solve for y_t, $t=s-k+1,\ldots,T$ given all the w_p. In case two ($k \leq 0$) the solutions for y_t, $t=1,\ldots,s-1$ and for $y_{s+i|s-1}$ ($i>0$) are the same as for a single shock. Only the solution for y_t, $t=s+1,\ldots,T$ is different and that solution is now given w_p, $p=s,\ldots,T$ and not just w_s.

To extend to a series of independent shocks (which therefore cannot be anticipated in advance), $w_s, w_{s+1},\ldots,w_{T-\ell}$ ($\ell>0$ for reasons which follow) then we have to repeat the calculations in each case for each new unanticipated shock. For every period in which a new piece of information, w_s, becomes anticipated, we need to calculate one new dynamic solution for each time period s with all shocks w_{s+j} set to zero (their expected value). In the second type of unanticipated shock we need two solutions for each period. We first need to evaluate the expectations from a dynamic consistent expectations solution given the lagged solution values from the previous period and no new shock. We then evaluate the impact of the shock from a single period solution given those expectations. The final shock is in period $T-\ell$ since we must keep some extra periods unshocked at the end of the horizon of interest to generate the expectations for the final period. For the interdependent sequence the shocks can occur right up to period T as long as that sequence possesses a stable trajectory (although one would still maintain T at a date beyond the horizon of interest).

In this section we have analysed the introduction of anticipated and unanticipated shocks in a general linear model. In the case of a single period shock we require either one or two solutions in addition to the base. It will be apparent that this analysis also generates the basic approach to deal with stochastic simulation in Section 5.3.

5.2 Unanticipated and anticipated shocks: empirical results

In this section we give an empirical illustration of the difference in effect between anticipated and unanticipated shocks. As noted above, conventional backward–looking models react no differently to anticipated and unanticipated shocks whereas if the shock is anticipated with forward–consistent expectations we expect to observe a change in behaviour prior to the actual shock. A common example in the U.K. economy is that if agents expect excise duties to be raised in the annual budget, then we observe a pre–budget spending boom on alcohol, tobacco, petrol etc.. In the following exercises we assume that the anticipated shocks do actually then occur. However it is entirely possible that agents may form expectations on the basis of a change in policy which does not then occur. The failure to change policy then becomes a particular type of unanticipated shock and this possibility is examined further in Sections 5.3 and 5.4.

If a model meets the stability criteria given in Section 4.1 then it has a unique long–run solution. We then expect to observe the same long–run effects for anticipated shocks as for unanticipated shocks, since the same steady–state growth path should be attained in both cases.

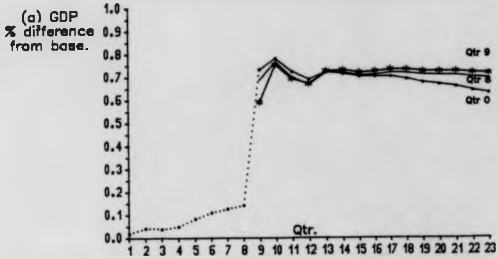
In general, models do not allow for different reactions to anticipated/ unanticipated shocks through the specification of individual equations, and differences emerge only from complete model solutions. A counter example is the LPL model which incorporates the reaction of economic agents to unanticipated inflation (defined as $y_1-y_{1|0}$) through a term in the real wage equation (with a negative coefficient). Under consistent expectations this term is identically zero for all except the first period where the lagged expectation term ($y_{1|0}$) is an initial condition formed outside of the solution period (at the beginning of period 0) and hence is not required to be consistent. The unanticipated inflation term can be set either to zero (the anticipated case) or can be calculated using the initial condition (the unanticipated case). In order to isolate the impact of this term we report a second set of results from the LPL model examining the impact of unanticipated inflation only.

(i) LBS and NIESR models

The shock to both the LBS and NIESR models is a step increase in real general government expenditure of £386m per quarter, assuming fixed nominal interest rates (implied money finance) and is introduced in quarter 9. We carry out three experiments. First we assume that the shock is anticipated at the start of the time horizon (quarter 0). Second we assume that it becomes known at the end of period 8 so that it is anticipated for the period in which it takes effect. Third, we assume that the shock becomes known at the end of period 9 and so it is completely unanticipated. The results for the response of GDP and the nominal exchange rate are graphed in Figures 5.1 (NIESR) and 5.2 (LBS).

For both models the effect of an unanticipated relative to an anticipated shock is a slightly smaller response of GDP in the period in which the shock is introduced. If the shock is not anticipated until after it occurs (quarter 9), the initial impact is lowest but the highest peak effect is observed later in the solution horizon. For GDP, differences persist over the whole of the solution. In both models, the effect of the unanticipated shocks starts slightly smaller and ends slightly larger, with the fully unanticipated simulation finishing with the largest effect of the three. The difference between the two types of unanticipated shock is not large but both have a persistently larger effect in the long–run than the anticipated shock (although the rate of change over time is the same). The

Figure 5.1 Comparison of anticipated and unanticipated shocks: NIESR model.



Note: shock occurs in Qtr 9 and becomes anticipated at the end of the period indicated.

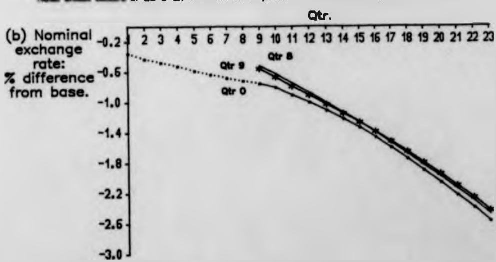
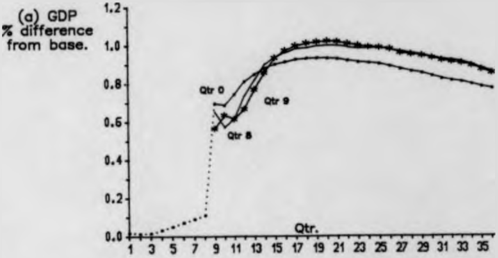
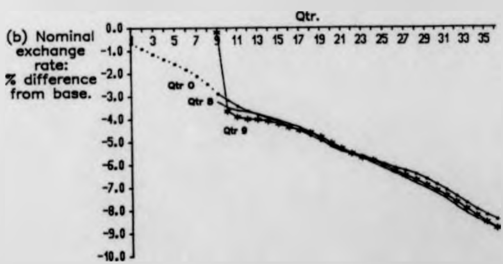


Figure 5.2 Comparison of anticipated and unanticipated shocks: LBS model.



Note: shock occurs in Qtr 9 and becomes anticipated at the end of the period indicated.



anticipated shocks support the proposition that some effects may be felt in advance of the imposition of the shock.

The pre–impact effects are much more important for the exchange rate than for GDP. The pre–shock effects on the exchange rate show an initial jump down in the first period followed by a steady depreciation and there is no noticeable change in its trajectory in the period when the shock is actually introduced.

It is noticeable that the simulations in both models do not appear to be tending to the same long–run solution for GDP whereas the exchange rate paths are very similar. In both models there may be some distortion due to the sensitivity of the solutions to the terminal condition (see Chapter 4). The only major difference between the two models is that the LBS model is more responsive overall, particularly for the exchange rate.

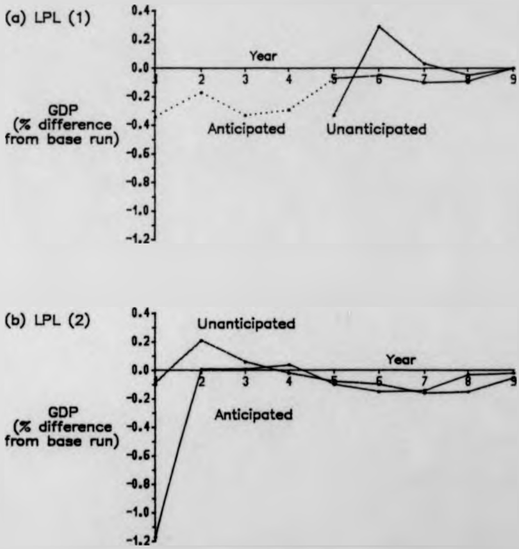
(ii) LPL model

In the comparable experiment for the LPL model, illustrated by Figure 5.3a, we again impart a step impulse to government expenditure under an assumption of money finance. In this case the shock takes effect in period (year) 5 and is either fully anticipated or anticipated only at the end of period 4.

There are clear signs that both anticipated and unanticipated solutions tend towards the same long–run solution, with the difference largely in terms of dynamics. Unlike the quarterly models, the unanticipated effects are stronger than in the anticipated case in the period in which the shock is introduced, but this ranking is reversed after two periods. The pre–impact effects are also far more significant for this model and it is a quite striking result that GDP falls for several years before the government expenditure shock is introduced. This reflects the model property that anticipated future inflation has a negative impact on current GDP.

A second set of LPL simulations are shown in Figure 5.3b to reflect the unanticipated inflation term in the wage equation as discussed above. The increase in government expenditure occurs in the first period for both the anticipated and

Figure 5.3 Comparison of anticipated and unanticipated shocks: LPL model.



unanticipated cases and any difference in the solution is entirely due to the inclusion of this term. This leads to a marked difference in the results for the period in which the shock occurs, but this is rapidly eliminated so that the two simulations coincide after about 5 years. The unanticipated shock causes real wages to be lower in the first period and hence output is higher than in the anticipated case.

The results suggest, therefore, that the choice of unanticipated/anticipated assumption does matter for the models although principally in terms of their short–run dynamic response. In the case of the LPL model, similar implications can be drawn from both the anticipated and unanticipated variants after 5–6 periods of the simulation, but the two quarterly models are more sluggish and the relatively short solution periods produce some medium–term discrepancy. Taking account the fact that the LPL model is annual and the other two are quarterly, then this difference is probably insignificant. The LPL responses also demonstrate that the effects can be extremely powerful when the unanticipated shock is directly formulated as part of a behavioural equation.

The shocks illustrated for all three models are to real general government expenditure. There is a simultaneous monetary impulse in the quarterly models due to the fixed nominal interest rate assumption. The LPL model has endogenous interest rates by construction and so the monetary shock is introduced simultaneously as part of the financing assumption. As a possible alternative, we find that shocks directly to nominal interest rates in the quarterly models yield similar conclusions except that there is only a small long–run response of GDP for any of the simulations.

5.3 Permanent and temporary shocks (policy reversal)

The second basic component of experimental design is the difference between permanent and temporary policy changes. A temporary policy change may reflect policy reversal in which case the nature of that reversal may be treated as anticipated or unanticipated. In this section we concentrate on the permanent/temporary distinction as in Wallis *et al.* (1986), but also consider the question of unanticipated policy reversal/continuation.

Conventional backward looking models react no differently to permanent and temporary shocks during the period in which they are in force. If we assume that such a model meets its stability conditions as described in Section 4.1 then it possesses a long–run steady–state solution. A permanent policy change will generally take the solution to the model on to a new steady–state trajectory. A temporary policy change will take the model solution on to the same path whilst it is in force. Once the temporary shock is removed, the model will tend back towards the original steady state, at a rate of adjustment depending on the intrinsic dynamics. The difference between a permanent and temporary policy change on a conventional model is therefore restricted to the period following the removal of the temporary change.

The long–run difference between temporary and permanent shocks is also a feature in models with forward expectations variables. In addition, they may react differently to a shock during the period in which it is in force, depending on whether it is anticipated to be permanent or temporary. If the temporary nature of a policy is anticipated, differences in reaction occur before the policy removal. Only if a shock is anticipated to be permanent but is actually temporary is the difference between temporary and permanent shocks restricted to the period following the removal of the shock. In our empirical results which follow, we generalise the results of Wallis *et al.* (1986) to include the case of unanticipated policy reversal and unanticipated policy continuation. For this purpose we assume that the policy is limited to the application or not of a shock to a policy variable.

As in the anticipated/unanticipated distinction, the models do not usually distinguish different reactions to permanent and temporary stimuli through the specification of individual equations. Although a consumption function might rest on a permanent income theory, for example, separate identification of permanent and transitory components of income seldom occurs. If we consider the forward expectations model (5.1) and its solution (5.3) we can immediately see how the solution for y_t depends on the expected values of u_s, $s=t,\ldots,T$ (and any implied change in y_T). The effect of a temporary shock is to truncate the influence of future changes in u_t and thus to diminish the effect on y_t in this simple model (for positive shocks, and given the implied coefficient of unity on u_t). We note that for finite horizon simulations, the temporary shock should end sufficiently far in advance of the terminal date that the terminal condition causes no distortion to the results.

5.4 Temporary and permanent shocks: empirical results

The experiment to be used is a shock of £386m to real general government expenditure, financed by increased monetary growth (fixed nominal interest rates for the quarterly models) as in Section 5.2. For each model the time horizon of the temporary shock is taken as a reasonable proportion of the available sample. For the LBS this is five years and for the LPL model, six. For the NIESR model, the available time horizon is relatively short and so only three years are used. Both permanent and temporary shocks are unanticipated on introduction. As variant solutions we simulate unanticipated changes in the duration of the shock. These are referred to as unanticipated policy reversal and unanticipated policy continuation. In these cases we assume that expectations immediately switch to a correct prediction of the new regime when the change is made.

In the LPL model, nominal monetary growth is exogenous and balanced– or money–financed increases in government expenditure must be accompanied by an explicit increase in monetary growth. The model identifies two separate exogenous monetary growth variables one denoted "temporary" and the other "permanent". Since either may be used to finance a government expenditure change, this prompts the question of how long such a change can be financed by "temporary" monetary growth – to which there seems to be no definitive answer. The "permanent" counterpart of monetary growth in the LPL model is calibrated to give balanced–finance changes in the PSBR and automatically adjusts the terminal value of inflation. When using temporary monetary growth for long periods the terminal condition must be relaxed (e.g. using constant growth rates) so as to find its own level corresponding to the stable solution. It is then a numerical issue as to whether the model is stable under a "permanent" shock to "temporary" monetary growth.

The response of GDP and the nominal exchange rate for the quarterly models are graphed in Figures 5.4 (NIESR) and 5.5 (LBS). We observe that the simulation responses of both models are in general smaller when the shock is temporary than when it is permanent, as forecast from our demonstration model, although the differences are not that substantial for the LBS model. For both models, after the end of the temporary shock, the GDP trajectory returns close to its base level whereas the nominal exchange rate returns to a trajectory parallel to its base. The temporary shock exhibits a permanent effect on the price level and hence on all nominal variables. The permanent shock appears to cause a permanent shift to the level of GDP and a constantly depreciating nominal exchange rate.

The unanticipated policy reversal/continuation simulations are identical to their permanent/temporary companions until the change in policy takes effect. On both quarterly models the trajectories then switch suddenly. The two temporary shocks attain very similar trajectories to each other within a few periods of the regime change, as do the two permanent shocks. In these models the main effect of the temporary/permanent distinction is the effect on the short–run of agent's expectations about future policy being incorrect.

Under "temporary" money finance in the LPL model, illustrated by Figure 5.6a, an unsustained shock generates larger output effects but weaker price effects than its sustained counterpart. The reversal of the ranking of the GDP responses compared with the quarterly model reflects the strong negative effect of future inflation from the permanent shock on current output, which more than offsets the effects of increased demand from the increased government expenditure. In the unsustained case the inflation effects are much weaker. After the end of the temporary shock the level of output returns rapidly to its base level. The permanent shock also has no long–run effect on the level of GDP. Hence the responses of the LPL model support the propositions of the new classical model under both temporary and permanent shocks.

Figure 5.4 Comparison of temporary and permanent shocks: NIESR model.

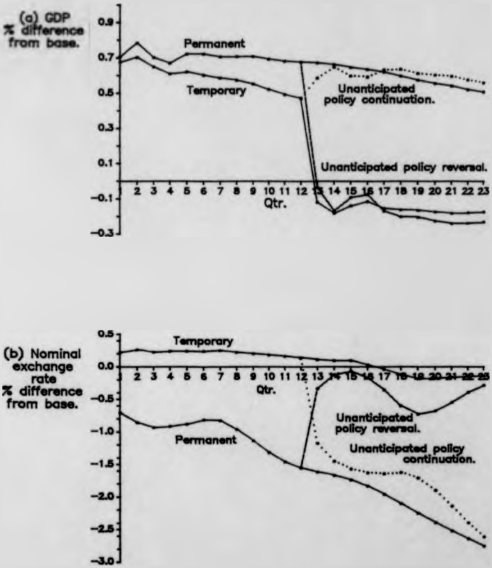
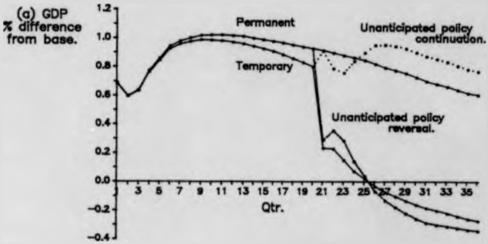
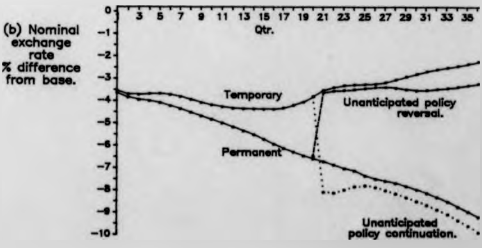


Figure 5.5 Comparison of temporary and permanent shocks: LBS model.





Once the terminal conditions have been correctly set, there is no difference in the GDP responses depending on whether permanent or temporary monetary growth is used to finance the PSBR. Figure 5.6b illustrates the use of the temporary component for a five year and fourteen year horizon. The only differences from Figure 5.6a are that both the temporary and permanent shocks generate a smaller GDP response (0.25% on average rather than 0.35%) and this reflects slightly higher inflation rates.

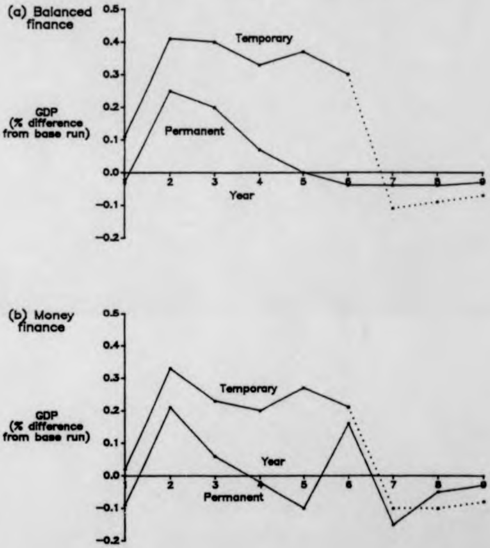
These conclusions from the LPL model suggest that a government expenditure increase will only have a substantial impact on the real economy if it is perceived to be temporary, as in the case of a pre–election boom. Such examples occur in practice with the stop–go policies historically associated with general elections in the U.K.. Governments typically relax fiscal and/or monetary policy in the period immediately before a general election. If such a relaxation was thought to be a pre–election boom, then the LPL model predicts that it would work: a higher rate of growth, lower unemployment and some inflationary increase for the duration. If it was thought to have been a permanent change in policy then there would be no boom. The conclusion from the LPL model is that for a pre–election boom to work, agents must know that it will be reversed post–election.

In conclusion we do find some differences between temporary and permanent shocks. In the quarterly models the temporary shocks tend to have weaker output and price level effects than their permanent counterparts, but the differences for the various shocks are not that great. In contrast the LPL model generates larger output effects for an unsustained shock. Here the higher price effects generated by a sustained shock adversely affect output. All the models suggest that an unsustained money financed expansion of spending merely raises the price level, but a sustained increase generates a higher inflation rate.

5.5 Stochastic simulation

As discussed in Chapter 2, stochastic simulation generates an empirical distribution of solutions for the endogenous variables whereas deterministic

Figure 5.6 Comparison of temporary and permanent shocks: LPL model.



simulation generates only point estimates. The empirical distribution is used for two main purposes. The vector of first moments estimated from the distribution gives unbiased estimates of the conditional expectations of the endogenous variables whereas the deterministic solution is a biased estimate. Higher order moments of the distribution are also of interest, either in their own right or in order to generate confidence intervals around forecasts.

In forward expectations models, the issue is complicated by the presence of explicit conditional expectations terms as explanatory variables and by the effect of alternative information assumptions in constructing values for these variables. If the information set includes the stochastic properties of the model then, for full rationality, every deterministic simulation should not have model consistent expectations but rather expectations which are unbiased estimates of the conditional expectation.

Stochastic simulation methods can be derived directly from the discussion in Section 5.1 of a series of independent, unanticipated shocks. Consider a vector of random disturbances, v_t, t=1,...,T, each of which enters the information set in an unanticipated fashion, i.e. v_s becomes known either at the end of period s or at the end of period s–1. We assume therefore that either agents can recognise such shocks instantly or with a one–period lag. The appropriate solution method is then the dynamic solution mode for a series of independent shocks. As each element of v enters the information set, a new solution must be formed, beginning in period s and given lagged values generated by the solution to the previous time period. This solution must then solve for enough periods to correctly generate the expectations $y_{s+1|s-1}$. A stochastic simulation follows by drawing a large number of pseudo–random disturbances and re–solving the model dynamically for each draw.

There remains the distinction between whether v_s is anticipated at the end of period s–1 or known only at the end of period s. If the latter assumption is made, two solutions are required for each random shock, first with v_s=0 to obtain $y_{s+1|s-1}$ and second given this expected value and v_s. This substantially increases the cost and complexity of a dynamic stochastic simulation.

Each simulation is a dynamic solution in which agents are revising expectations as each piece of information becomes known. Hence the v_t, t=1,...,s–1, are treated as known in period s and each solution for y_s is given lagged solution values for y_t, t=1,...,s–1. If we repeat this exercise for R pseudo–random draws of v_t, t=1,...,T–ℓ (retaining ℓ unshocked periods to satisfy the stability conditions) we can obtain a distribution of solutions for y_t, t=1,...,T–ℓ. Each observation of this distribution can be interpreted as a possible alternative out–turn of the economy and its moments can be interpreted as the mean out–turn, standard deviation of the out–turn, etc.

In order to generate the vector of pseudo–random disturbances, we need to consider carefully the calculation and distribution of the realised structural disturbances. Whichever of the methods discussed in Chapter 2 that are used to generate pseudo–random disturbances, one needs to form estimates of the realised disturbances over some historical period. In the parametric methods (i.e. all other than the Mariano and Brown approach) we assume that the realised disturbances are generated from a normal distribution. The estimates of the disturbances are the single–equation residuals from the structural equations and are therefore dependent on the unobserved historical values of expectations variables. Since these values are unknown, we must consider ways of approximating the expectations in a single equation context such that the structural residuals, and hence our pseudo–random draws, have the desired distribution. We propose to use the actual historical outcome as a proxy for the expectation (i.e. a perfect foresight assumption) and briefly consider the resulting distribution.

Under a rational expectations assumption, agents' expectations should be unbiased estimates of the actual outcome with no deterministic error component: such an error could be exploited systematically to improve the forecast. Hence the residuals generated by assuming perfect foresight should have zero mean error. That they also have the correct variance is less obvious since the residuals generated in this way incorporate the error in the expectation. For example, taking equation (5.1) and treating u_t as a random disturbance term we have:

$$y_t = \alpha y_{t+1|t-1} + v_t, \qquad v_t \sim N(0, \sigma_v^2) \tag{5.6a}$$

$$y_{t+1|t-1} = y_{t+1} + \epsilon_t, \qquad \epsilon_t \sim N(0, \sigma_\epsilon^2) \tag{5.6b}$$

$$y_t = \alpha y_{t+1} + w_t, \qquad w_t = v_t + \alpha \epsilon_t \tag{5.6c}$$

if $E(v_t \epsilon_t) = 0$, then

$$w_t \sim N(0, \sigma_v^2 + \alpha^2 \sigma_\epsilon^2) \tag{5.6d}$$

If (5.6b) is a correct representation of expectations formation, then equation (5.6d) gives the correct variance when the model is solved for consistent expectations. This follows since the consistent expectation terms in model solution are generated by identities which set $y_{t+1|t-1} = y_{t+1}$ neglecting the error term implied by equation (5.6b). Generating residuals using a perfect foresight assumption compensates for this omission by combining both error terms in the input disturbances. If an explicit expectations mechanism, and hence expectations values were available then the two disturbances would be separately identified each with its own variance, and the error in the expectations formation equation would be entered separately.

The model's equations are often estimated using McCallum's (1976) instrumental variable/2 stage least squares (IV/2SLS) approach in which the expectation is approximated by the historical outcome and then instrumented. The residuals from any 2SLS regression have a normal distribution only when the actual explanatory variable values are substituted back into the final regression. In our case the variables being instrumented are the forward expectations terms whose "actual" values are the one–period ahead historical values. Hence the use of a perfect foresight assumption in calculating single equation residuals should be consistent with the distributional assumptions used in estimation.

The residuals from these equations will, in general, be serially correlated since the disturbance v_{t+1} will be partly responsible for the error in the expectation of y_{t+1} i.e. v_{t+1} and ϵ_t are correlated. In his extension of the McCallum procedure, Wickens (1982, p64) claims that the residuals are not serially correlated but this follows only as a result of particular informational assumptions. If serial correlation is found to be present, it can be included in the pseudo-random disturbances for dynamic simulation by an amended form of the McCarthy (1972) procedure given by Schink (1971).

Other proposed methods of stochastic simulation.

Two other methods of stochastic simulation for forward-expectations models have been suggested in the literature; by Hall and Henry (1985a,b) and Fair (1984, pp383–384). The Fair method is only a suggestion and does not appear to have ever been applied. This scheme involves the following steps.

(i) Given draws of u_t^r, $t=1,\ldots,T-\ell$, $r=1,\ldots,R$, solve the model R times for consistent expectations, each solution being given the values of v_t^r over the entire horizon.

(ii) Average the solutions for y_t, $t=1,\ldots,T-\ell$ and use these to construct the expectation terms.

(iii) Subject the model to another stochastic simulation holding expectations fixed at the values calculated at step (ii). This final solution is calculated one period at a time, presumably with either average or actual values for the lags – this is not made clear by Fair.

The properties of this method do not appear to have been fully analysed. It can only be applied for a static stochastic simulation because step (iii) is for one period only – it cannot be extended to the dynamic case because expectations are held fixed at their average value and cannot respond to the random shocks. Once a shock has been observed and y_t has changed, we would expect $y_{t+2|t}$ to change. Fair confirms the static nature of his method when he explicitly states that the final solution only requires random draws for the period of interest. Furthermore, it is not clear why the average expectations from the first stochastic simulation are rational. Each solution is constructed as if each v_s is fully anticipated in period zero. This seems to be an arbitrary and extreme assumption.

Hall and Henry explicitly refer to their use of stochastic simulation as a method of obtaining "rational" rather than consistent expectations. By this they explain that they mean unbiased estimates i.e. which correct for the bias arising from the deterministic simulation of a nonlinear model. They thus attempt to differentiate between their approach and that for obtaining the full distribution for the endogenous variables in a forward–consistent–expectations model. There seem to be no justifiable grounds for such a distinction. The unbiased expectation that they seek is an unbiased estimate of the mathematical conditional expectation of the endogenous variables. Such an estimate is provided by estimating the first moment from a sample of possible out–turns of the economy. Our stochastic simulation procedure outlined above automatically provides "unbiased" expectations conditional on the model and the exogenous variable values.

The algorithm used by Hall and Henry is not actually described in either of their 1985 papers nor their 1987 survey of their work. The precise details of the algorithm reported here are based on a verbal description provided by Stephen Hall. The method is distinct from our preferred technique and that of Fair; it proceeds as follows:

(i) Solve the model dynamically for y_t, $t=1,\dots,T$ given $v_t=0$ under consistent expectations for $(y_{t+1|t-1})$ from the solution.

(ii) Holding the expectations fixed, solve for y_t^r, $t=1,\dots,T$, $r=1,\dots,R$ using R successive drawings of pseudo–random disturbances v_t, $t=1,\dots,T-\ell$.

(iii) Estimate the average values $\bar{y}_t$ and obtain $(y_{t+1|t-1}-\bar{y}_{t+1})$ as an estimate of the average error in the expectations.

(iv) Re–set $y_{t+1|t}=\bar{y}_{t+1}$ and perform another stochastic simulation; iterate until the expectations do not change.

The convergence properties of this algorithm are not evaluated by Hall and Henry but it would appear that if $y_{t+1|t-1}$ has converged to an unchanging value, then it must also have converged to $\bar{y}_{t+1}$. The distribution produced by step (iv) of this algorithm does not constitute an estimate of the distribution of possible out–turns for the economy. The expectation terms are held fixed at $\bar{y}_{t+1}$ and hence the variance of y_t does not reflect any variance component due to the expectations terms. The algorithm thus implicitly assumes that all shocks remain unanticipated even after they have occurred. We can make a variety of assumptions about when a disturbance becomes known, but to assume that it remains unknown after it has affected y_t seems to be an extreme assumption. In practice, once a shock has occurred, even if previously unanticipated, expectations of all future periods will be revised. At period t=0, agents know that they will revise their expectations in the future as shocks unfold. If our stochastic simulation does not reflect this then it cannot generate a true estimate of the distribution of the endogenous variables. Hence the Hall and Henry method makes an arbitrary assumption at the other extreme to that of Fair.

5.6 Empirical results: stochastic simulation

In Fisher, Wallis and Whitley (1985, hereafter FWW) we considered the effects of differing assumptions concerning the financing rules for implied PSBR changes. In the experiments reported in Sections 5.2 and 5.4 we assume that such changes are financed by the endogenous movements of the money stock (exogenous changes in the LPL model only). The models are not generally set up in such a way that bond financing assumptions can be automatically applied. In all three models, bond financing is imposed by holding the money stock fixed. This can be achieved by varying bond stocks in the LBS model, nominal interest rates in the NIESR model and by not changing the money stock in the LPL model. The two possible variations are more properly described as an endogenous money stock or money targeting but we retain the labels and money and bond financing respectively.

FWW report two types of simulations to show the effect of the alternative assumptions. The first method used is simply deterministic simulation over a fixed time horison to examine the different dynamic trajectories of output and the price level. The theoretical literature on this topic is founded on the work of Christ (1968) and Blinder and Solow (1973). More recently, Whittaker *et al.* (1986) used a theoretical model with a range of plausible parameter values and conclude that fixed money stock rules are less stable than fixed bond stock rules.

In general, FWW find some empirical support for the results of Whittaker *et al*. The results from dynamic simulations are, however, somewhat inconclusive due to the difficulties of extracting long–run properties from large–scale nonlinear models (for one possible resolution of this problem see Deleau *et al*, 1989). Whittaker *et al* also find that alternative expectations formation mechanisms do not affect their conclusions but this issue could not be addressed on the model vintages used by FWW.

The second method used by FWW is to undertake stochastic simulations under the alternative financing rules so as to examine the changes in the distributions of output and prices. Thus one can examine the *variance* of the endogenous variables as well as their dynamic *variability*. The theoretical literature (Currie, (1976, 1978); Rau, (1985)) again suggests that using money rather than bonds to accommodate PSBR changes leads to greater stability – in this case a lower variance of output.

The models used by FWW in the stochastic simulations are the LBS and NIESR models before the introduction of forward expectations. In this section we therefore extend their results using the next generation of those models, including forward expectations. Not only does this allow us to evaluate the issue of alternative financing rules but also the effects of introducing forward expectations.

In Tables 5.1, 5.2 and 5.3 we give results of stochastic simulation experiments consistent with those of FWW but with solution of the models under consistent expectations, as outlined above in Section 5.5. We assume that the shocks are anticipated at the beginning of the period in which they are introduced. We include the LPL model as well as the NIESR and LBS models in these experiments.

In order to generate the pseudo–random draws we use the McCarthy (1972) procedure and thus make an assumption that they have the same variance–covariance matrix as the model's single–equation errors and that they are normally distributed. The LPL model coefficients are mostly calibrated for forecasting rather than being freely estimated and the observed single equation residuals may not possess the properties of a normal distribution We could resort to non–parametric methods such as those of Mariano and Brown (1984) but those methods require a reasonable sample size to work with and we are limited to eight available on this model (a small sample size will also adversely affect the properties of the residuals generated by the McCarthy method). In all three models we restrict attention to a single period stochastic simulation. The 250 replications used then require 250 dynamic simulations equivalent to 250 consistent expectations solutions — this is already a substantially greater cost than a *dynamic* stochastic simulation on a conventional model. The extension to the dynamic case is not only more costly but unlikely to be fruitful in this context since the dynamic solutions to the models are not very robust dynamically under a bond financing assumption — which provides further support for the possible dynamic instability under fixed nominal money stock rules.

The random number generator used here is not truncated and will therefore occasionally draw observations in the extreme tail of the normal distribution for which the nonlinear models cannot solve (e.g. if it means unemployment becoming negative). The scale of the input shocks for the quarterly models is thus reduced to minimize the chances of extreme draws and thereby to ensure that the distribution does not have significantly truncated tails. This is especially important given that we are interested primarily in the second moment rather than the first. We then consider only the relative magnitude of the degree of variability and not the absolute magnitudes (see notes to tables for further information on the experimental design).

(i) NIESR model

Taking the NIESR results presented in Table 5.1, we show appropriately scaled results from FWW in brackets alongside the results for the forward expectations models. We see that under bond finance the standard deviation of GDP rises significantly compared with money finance. In this model, a fixed money stock is dynamically unstable under forward consistent expectations and so the money stock has only been fixed in the first period of the eight quarter solution

TABLE 5.1: Stochastic simulation of the NIESR model[4]

Standard deviations[3]

Experiment[1]	GDP[3]		Price level[3]	
Money finance	59	(59)	0.234	(0.17)
Bond finance[2]	75	(73.5)	0.226	(0.20)

Notes

1 Solution period is eight quarters, 1985(3)–1987(2); stochastic shocks applied to first quarter only (hence these are one–step–ahead results); 250 replications using McCarthy technique, no antithetic variates. The standard error of the standard deviation is (approximately) $\sigma/\sqrt{2n}$.

2 Bond finance assumption is maintained only in the first quarter using the exogenous nominal interest rate to target £M_3.

3 Input disturbances are scaled down to one tenth of historical disturbances in order to prevent numerical failures and thus to avoid truncating the input distribution. Output units are £m and index pts for GDP and the price level respectively but should then be re–scaled by a factor of 10.

4 NIESR model is version 8, Autumn 1985.

5 Figures in brackets are scaled down values calculated from the results of Fisher, Wallis and Whitley (1986) for an earlier non–forward–expectations version of the model.

horizon. The instability reflects weak interest rate effects on the money stock, strong interest effects elsewhere (especially on the exchange rate which gives the most important forward looking channel) and weak money influences on the model. To evaluate the significance of the increase in the standard deviation we calculate its standard error which is approximately $\sigma/\sqrt{2n}$ or in this case $74/\sqrt{500} = 3$. Hence the difference of 15 represents approximately 5 standard errors. In contrast, the standard deviation of the price level has fallen but not significantly.

These results give similar conclusions to those of FWW on the earlier, backward–looking version of the NIESR model (Model 7 as opposed to Model 8). The main change in the results is that the variability of the price level declines under bond finance when expectations are consistent whereas under the earlier treatment, it increased. In fact the variability of the price level under bond finance has not greatly changed and the result rather reflects an increase in variability under money finance. Output (GDP) has an increase in variability under bond finance, compared with money finance, which is almost identical on the two model vintages. This increase in variability is much more marked if real interest rates are used as an instrument to target the money stock, as opposed to nominal interest rates. In principle, there are many different policy assumptions which could be evaluated using this approach. Obviously the results reflect not only the introduction of equations with forward–consistent expectations terms, but also the changes in other equations between the two versions of the model, which means that the comparison should be treated with some degree of caution.

(ii) LBS model

Results for the LBS model are shown in Table 5.2. In this model the incorporation of forward expectations has greatly increased the model's sensitivity to shocks (see Wallis *et al*, 1985, 1986) and this is reflected in an increase in variability of more than double for GDP compared with FWW. Again, the main qualitative difference is that price level variability under bond finance is now about three standard errors less than under money finance whereas previously rough parity

TABLE 5.2: Stochastic simulation of the LBS model[4]

Standard deviations[5]

Experiment[1]	GDP[3]		Price level[3]	
Money finance	126	(40)	0.101	(0.078)
Bond finance[2]	132	(61)	0.087	(0.078)

Notes

1 Solution period is eight quarters, 1986(1)–1987(4); stochastic shocks applied to first quarter only (hence these are one–step–ahead results); 250 replications using McCarthy technique, no antithetic variates. The standard error of the standard deviation is (approximately) $\sigma/\sqrt{2n}$.

2 Bond finance assumption is maintained over eight quarters and is applied using short term bill stocks to target £M_3.

3 Input disturbances are scaled down to one tenth of historical disturbances in order to prevent numerical failures and thus to avoid truncating the input distribution. Output units are £m and index points for GDP and the price level respectively but should then be re–scaled by a factor of 10.

4 LBS model is that used to forecast in Autumn 1985.

5 Figures in brackets are scaled down values calculated from the results of Fisher, Wallis and Whitley (1986) for an earlier non–forward–expectations version of the model.

was achieved. Output variability, while still larger under bond finance, is now only one standard error larger as opposed to more than seven in the previous vintage of model. Hence the incorporation of forward expectations in this model has substantially reduced the real effects and increased the nominal differences between money and bond finance.

The different reactions of the two models to the incorporation of forward expectations reflects the different ways in which this has been done. In the LBS model a fast moving real exchange rate coupled with sluggish wages and prices produces a very sensitive model. This vintage of the NIESR model has more pervasive expectations terms but the exchange rate movement is sluggish and thus the model is less sensitive. We generally find a greater variance of GDP under bond finance than under money finance; the reverse being true for the price level.

(iii) LPL model

In the LPL model the stochastic simulations under money finance involve changing an exogenous nominal money growth variable to hold government expenditure constant whereas bond finance requires nominal monetary growth to be held fixed whilst there is some movement in general government expenditure as the output level changes. Hence the results may not be strictly comparable with the quarterly models which assume that general government expenditure is fixed under both money and bond finance. The results for the LPL model are the complete opposite of those for the two quarterly models. Under bond finance, the output variance is lower and that of the price level higher than under money finance. The difference between the output standard deviations is substantially less than one standard error and so no strong conclusions may be drawn. However, the price level's standard deviations differ by about two standard errors.

This is a strong result when one considers the model's structure. The inflation equation is a function only of exogenous nominal monetary growth and endogenous real monetary growth. For the variance of the price level to decrease

TABLE 5.3: Stochastic simulation of the LPL model[4]

Standard deviations

Experiment[1]	GDP (£m)	Price level (index pts 80=100)
Balanced finance	6672	4.94
Bond finance[2]	6459	5.55

Notes

1 Solution period is fourteen years, 1988—2001; stochastic shocks applied to first quarter only (hence these are one—step—ahead results); 250 replications using McCarthy technique, no antithetic variates. The standard error of the standard deviation is (approximately) $\sigma/\sqrt{2n}$.

2 Bond finance assumes that government expenditure is held constant and no extra nominal money growth is allowed. Balanced finance assumes that government expenditure is held constant and the implied PSBR/GDP ratio changes to ensure that this results.

3 LPL model is annual, no comparison may be drawn between Table 6.3 and Tables 6.1, 6.2. For this model, there is no scaling down of the variance of the input distribution.

4 LPL model is that used to forecast in Autumn 1987.

when nominal monetary growth is being varied to hold government expenditure fixed implies strong covariance between real money demand and the "exogenous" variable.

These results demonstrate the substantial differences in model properties between the NIESR and LBS models on the one hand and the LPL model on the other. The application of stochastic simulation has allowed us to compare the model properties for higher order moments and not just the deterministic simulations. These reveal the lower variance of output under bond finance on the LPL model, even though real debt interest rapidly accumulates in line with the dynamically unstable case of Blinder and Solow (1973).

5.7 Summary

Stochastic simulation is a feasible computational option on conventional models (see the large literature surveys in Fisher and Salmon, 1986; reproduced and extended in Hall and Henry, 1987, pp248–249). The correct procedure for a forward expectations model is many times more expensive. To date it has only proved practical for a one–period–ahead forecast. However, for the smaller LPL model at least, it should be possible to extend to the dynamic case. Future improvements in computing speed could also make it practical for larger models.

The main interest in stochastic simulation is likely to arise for issues concerning the higher–order moments of the distribution, especially the variance. The example considered in Section 5.6 concerning alternative financing rules yields some useful results which appear to be consistent for the quarterly models. We find that bond financed changes in the PSBR due to stochastic shocks yield a higher variance for output and a lower variance for the price level than when such PSBR changes are financed by changes in the money stock. These results appear to be consistent with the theoretical literature.

We can also use estimates of higher–order moments to help assess the introduction of forward expectations into the quarterly models. The main differences compared with the results of Fisher, Wallis and Whitley (1985) on an earlier vintage of model, is seen to be a more significant reduction in the price level variance under bond finance. The LBS model has also been made more sensitive generally. These results for the variance are consistent with the findings of Whittaker *et al.* (1986) that the expectations formation mechanism does not affect the results of their model when investigating dynamic stability.

The opposite ranking of variances for the LPL model under alternative financing rules is a reminder that such exercises primarily reveal useful information about the model's (stochastic) structure and only tell us about the real economy if the model is a good approximation to it.

ALTERNATIVE MODEL FORMS AND SOLUTION MODES: HISTORICAL TRACKING

The aim of this chapter is threefold. First we draw out the links between different model forms (as laid out by Salmon and Wallis, 1982) and the alternative types of solution for macroeconomic models such as single–equation calculations, static and dynamic simulation. We then develop analogous model forms for rational expectations models and their corresponding solutions. In particular we propose single–equation and static simulation procedures for which particular attention must be paid to treating the expectations terms as estimates of agent's actual (aggregate) rational expectations. Finally, we present a study of the historical tracking performance of models of the United Kingdom economy which requires the use of single–equation, static and dynamic solution methods.

6.1 Conventional models

In a conventional dynamic econometric model there are a number of identical representations of the same system. Consider the structural form of the general dynamic linear model defined as in (4.1) above,

$$B(L)y_t + C(L)x_t = u_t. \tag{6.1}$$

Usually the model will be normalised so that B_0 has unit elements on the main diagonal. If $B_0=I$, the equations have no instantaneous coupling and each element of y_t can immediately be expressed as a function of predetermined variables only.

In the general case $B_0 \neq I$ and to express y_t as a function of predetermined variables we use the following reduced form (if B_0 is nonsingular):

$$y_t = -B_0^{-1}[B_1 y_{t-1} + \dots + B_p y_{t-p} + C(L)x_t] + B_0^{-1}u_t. \tag{6.2}$$

An alternative form of the model is given by expressing y_t as a function of exogenous variables and disturbance terms only; multiplying through by the inverse of B(L) rather than B_0 (under appropriate conditions on B(L), see Chapter 4).

$$y_t = -B(L)^{-1}C(L)x_t + B(L)^{-1}u_t \tag{6.3}$$

This is referred to by Salmon and Wallis (1982) as the final form. Now, a loose correspondence can be drawn between (6.1), (6.2) and (6.3), and the solution modes of an econometric model as follows.

Single–equation "solutions" to a simultaneous model depend on a particular normalization of the system. Each structural equation has a nominated dependent variable for which the single–equation solution is calculated. The calculation is performed holding all explanatory variables at observed historical values whilst setting the residual terms to zero. In a linear system estimated by OLS, the single equation solution calculated over the estimation sample corresponds to the fitted values of the regression (assuming the same normalization is used as in estimation). We can re–express the structural form (6.1) as

$$y_t = (I - B(L))y_t - C(L)x_t + u_t. \tag{6.4}$$

Each equation of system (6.4) can then be written as:

$$y_{it} = y_{it} - \sum_{j=1}^{n} B_{ij}(L)y_{jt} - \sum_{k=1}^{m} C_{ik}(L)\, x_{kt} + u_t, \qquad i = 1, \ldots, n. \tag{6.5}$$

where $B_{ij}(L)$, $C_{ik}(L)$ are scalar polynomials in the lag operator formed from elements (i,j) and (i,k) respectively of the matrix polynomials B(L) and C(L). If $B_{ii}(0)$ is unity, as is conventional, the first term on the right–hand side cancels out with $-B_{ii}(0)y_{it}$ in the second term. When u_t is set to zero, the solution to equation (6.5) is then the single equation solution which we denote y_t^{se}. The most useful products of this calculation however, are the single–equation residuals $\hat{u}_t = y_t - y_t^{se}$. We may think of y_t^{se} as being consistent with the (unsolved) structural form (6.1) and the single equation regression model.

The single equation solution to a simultaneous model is a useful tool in modelling work. For example, it provides a useful diagnostic check that an equation has been coded properly. Over an out–of–sample historical period, the single–equation residuals can be used to indicate when an equation has broken down. In stochastic simulation, single–equation residuals are used to provide an estimate of the estimation residuals and their variance–covariance matrix for the various techniques discussed in Section 2.3. The most frequent use of these residuals, however, is to provide residual adjustments in counter–factual solutions (Artis, 1984). In such analysis one requires a base run which coincides with the historical record. A single equation calculation is therefore made across all equations to estimate the entire vector $\hat{u}_t$. We demonstrate below that when these residual values are then used as adjustments to the solution, the model will reproduce the historical record. Finally, a single–equation solution can be used when a coefficient or equation (or sub–sector) of a model is altered (e.g. a policy rule is changed). A single–equation calculation, using as input a base solution to the original model, will determine the adjustments needed to leave the solution for the model unchanged when the new equation(s) is introduced. The new model properties can then be assessed around the original base: an important requirement for nonlinear systems which are state dependent.

A static solution to the model (6.1) is generated by setting all pre–determined variables, i.e. exogenous and lagged endogenous variables, to their actual values and solving only for current endogenous terms. The model can be re–expressed as:

$$B_0 y_t = -[B_1 y_{t-1} + \dots + B_p y_{t-p}] - C(L)x_t + u_t. \qquad (6.6)$$

With u_t set to zero, solving equation (6.6) for a particular value of t produces the static solution which we will denote y_t^{ss}. This solution can be thought of as corresponding to the reduced form (6.2). If the disturbance terms u_t are set to their observed values, rather than zero, then solving equation (6.6) will obviously reproduce the historical values for y_t as required in the counter–factual case.

The static solution mode is rather infrequently used in large–scale modelling. However, Fisher and Wallis (1988) argue that static solution is the most appropriate mode for historical tracking exercises which are designed to evaluate a model's performance in explaining the past. The arguments supporting this point of view are presented in Section 6.3.

A dynamic solution to the model (6.1) is generated over a period $t=1,\ldots,T$ by allowing the solution values from one period to feed through to the following periods. Hence only the exogenous variables are fixed at their observed values. The model can be re–expressed as:

$$B(L)y_t = -C(L)x_t + u_t, \qquad t=1,\ldots,T. \tag{6.7}$$

Solving equation (6.7) with the disturbance terms set to zero produces a full dynamic solution denoted y_t^d, $t=1,\ldots,T$. This is performed recursively after setting initial values for $y_0,\ldots,y_{1-p}$. It is equivalent to solving a repeated version of (6.6) for $t=1,\ldots,T$, where lagged values are set to solution values from $y_{t-1}^d,\ldots,y_{t-p}^d$. This solution is analogous to a finite–time solution to the final form although it may also be termed the dynamic reduced form. Unlike the single–equation or static modes, dynamic solution requires no historical data for the endogenous variables over the solution horizon (only the initial conditions and exogenous variables are needed). Hence, by forecasting the exogenous variables, dynamic solution is used to provide a genuine *ex ante* forecast. In this case the disturbance terms may be set to non–zero values or to known future disturbances. If a dynamic solution is conducted over an historical period and the residuals set to their single–equation values then, as noted previously, equation (6.7) will re-produce the historical values of y_t.

The correspondence of dynamic solution to the final form is not exact because the term $B(L)^{-1}$ in equation (6.3) implies an infinite distributed lag on x_t and u_t. The dynamic solution is thus perhaps more consistent with the final equations in which $B(L)^{-1}$ in (6.3) is written as $b(L)/|B(L)|$, where $b(L)$ is the adjoint matrix and $|B(L)|$ the determinantal polynomial, hence

$$|B(L)|y_t = -b(L)C(L)x_t + b(L)u_t. \qquad (6.8)$$

Dynamic solution is the standard operating mode for *ex ante* forecasting and policy analysis, counterfactual or otherwise.

There is one further solution of potential interest. The dynamic single-equation calculation is generated by taking individual structural equations such as equation (6.5) and solving them in a dynamic mode for the dependent variable. The model equations can be expressed as:

$$B_{ii}(L)y_{it} = -\sum_{j\neq i} B_{ij}(L)y_{jt} - C_i(L)x_t + u_{it}, \qquad (6.9)$$
$$t = 1, \ldots, T.$$

The solution to equation (6.9) for y_{it}, $t=1,\ldots,T$ is the dynamic single equation solution which we denote y_t^{ds}. This solution does not correspond with any of the system representations but may be thought of as the final form for each equation treated as an independent model (i.e. treating all other endogenous variables as exogenous).

The dynamic single-equation solution is rarely used in modelling work other than to describe the properties of individual equations. Of particular interest may be the long-run solution for y_t^{ds} when the equation is subjected to a shock.

From this analysis we have drawn a correspondence between conventional models (single equation; static; dynamic) and the models' alternative representations (structural form; reduced form; final form). For a linear system the algebraic forms given above can be used to deliver the different representations and hence the different solutions directly. When generalizing to nonlinear systems we cannot express the model in such convenient "closed" form expression. However, we can define the structural form, reduced form and final form as follows:

$$f(y_t, y_{t-1}, \ldots, y_{t-p}, x_t, x_{t-1}, \ldots, x_{t-q}; \Psi) = u_t, \tag{6.10}$$

$$y_t = g(y_{t-1}, \ldots, y_{t-p}, x_t, x_{t-1}, \ldots, x_{t-q}, u_t; \Psi), \tag{6.11}$$

$$y_t = h^*(x_t, x_{t-1}, \ldots, x_{-\infty}, u_t; \Psi), \tag{6.12a}$$

$$\text{or} \quad y_t = h(y_0, \ldots, y_{1-p}, x_t, x_{t-1}, x_{t-q}, u_t; \Psi), \tag{6.12b}$$

where Ψ is a vector of parameters; f, g, h, h^* are vectors of nonlinear functions; $y_0, \ldots, y_{1-p}$ are the p initial conditions required to generate an exact solution (6.12b) to the final form (6.12a). The model is usually constructed in the form of (6.10). The functional forms of equations (5.11, 5.12a,b) are not normally known and hence these solutions are generated numerically by the procedures described in Chapter 3. The correspondence to model solution modes then carries through to the numerical solutions generated by (6.10), (6.11) and (6.12b). Note that the stability and existence conditions can no longer be derived by linear matrix algebra and we usually assume that (6.11) and (6.12b) exist, and examine stability *ex post* by numerical methods.

6.2 Forward expectation models

Consider the following linear model with forward expectations:

$$B(L)y_t + C(L)x_t + A(F)y_{t+1|t-1} = u_t, \tag{6.13}$$

where the terms are defined as before and, in particular, the forward operator F acts as $Fy_{t+s|t-1} = y_{t+s+1|t-1}$. We then proceed to establish the corresponding model forms and solution modes. In order to establish these forms/modes we must remember that the forward expectations terms are generally unobserved variables representing the conditional expectations of economic agents. Different assumptions with respect to the content of the information set Ω_{t-1} will yield different solutions to the model.

The structural form of the model is given by equation (6.13). The corresponding single–equation solution is generated by taking an individual equation from (6.13) such as:

$$y_{it} = y_{it} - \sum_{j=1}^{n} B_{ij}(L)y_{jt} - \sum_{k=1}^{m} C_{ik}(L)\, x_{kt} - \sum_{j=1}^{n} A_{ij}(F)y_{jt+1|t-1} + u_t,$$
$$i = 1, \dots, n, \qquad (6.14)$$

Where, as before, the first term on the right–hand side of (6.14) (y_{it}) will usually cancel with $B_{ii}(0)y_{it}$ in the second term. Single equation calculations can only be performed *ex post* since we require knowledge of the exogenous variables and current–dated endogenous variables on the right–hand side of equation (6.14). The residuals and dependent variables are estimated by solving equation (6.14) for $\hat{y}_t = y_t^{se}$ and hence $u_t = \hat{y}_t - y_t^{se}$. However, in order to solve equation (6.14) we first need to determine values for the unobserved expectations $y_{jt+s|t-1}$. Given the use of a rational expectations assumption in estimation, and since most estimators proceed by using actual forward values subject to instrumental variables (e.g. McCallum, 1976a) one option is to replace the expectations with the actual forward values. This was our proposed method for stochastic simulation in Chapter 5. This perfect foresight assumption will also be appropriate when single equation residuals are required to fix a dynamic or static solution to a base run or the historical record. A simple one lead, one lag model gives:

$$
\begin{aligned}
y_t &= B_1 y_{t-1} + A y_{t+1|t} + C x_t + e_t, \\
y_{t+1|t-1} &= y_{t+1} + \epsilon_t \quad \text{and hence} \\
y_t &= B_1 y_{t-1} + A y_{t+1} + C x_t + e_t + A\epsilon_t.
\end{aligned}
\tag{6.15}
$$

Hence the single–equation residual contains the disturbance term of the behavioural equation and the error in the expectation.

In the single–equation context we do not wish to use the full model to deliver the expectations since the single–equation residuals for each equation should not reflect the (mis–)specification of other equations in the system. Perfect foresight is consistent with the rational expectations assumptions without invoking the full model structure. However, there are other possible assumptions we could make concerning the expectations terms. Dropping rational expectations, we could substitute in explicit time series rules (or adaptive expectations rules) as in Frenkel, Goldstein and Masson (1988, appendix).

The reduced form of the model can be derived for the equivalent representation to equation (6.6) by inverting B_0 in (6.13) to give:

$$
y_t = -B_0^{-1}[B_1 y_{t-1} + \ldots + B_p y_{t-p} + C(L)x_t + A(F)y_{t+1|t-1}] + B_0^{-1}u_t. \tag{6.16}
$$

The corresponding static solution mode is given by solving:

$$
B_0 y_t = -[B_1 y_{t-1} + \ldots + B_p y_{t-p} + C(L)x_t + A(F)y_{t+1|t-1}] + u_t, \tag{6.17}
$$

for $\dot{y}_t = y_t^{ste}$ with the disturbances u_t set to zero. The reduced form residuals can then be estimated as $v_t = B_0^{-1}\dot{u}_t = (y_t - y_t^{ste})$. Since the static solution mode requires knowledge of lagged endogenous variables it is valid only over an historical solution base. As before we need to define the unobserved expectations terms

$A(F)y_{t+1|t-1}$. In the full model context the rational expectation is that provided by the model's own forecast $\hat{y}_{t+1}$, t=1,...,T. This must be obtained by a dynamic consistent expectations solution to the model to which we now turn.

The final form of the model is given by re-writing equation (6.13):

$$[B(L) + A(F)]y_t = -C(L)x_t + u_t \tag{6.18a}$$
$$y_t = -[B(L) + A(F)]^{-1}C(L)x_t + [B(L) + A(F)]^{-1}u_t \tag{6.18b}$$

where $[B(L)+D(F)]^{-1}$ will give an infinite distributed lag and lead on x_t and u_t.

The dynamic solution implied by this model form now makes the consistent expectations assumption $y_{t+s|t-1}=\hat{y}_{t+s}$. It is obtained by solving equation (6.18a) for $\hat{y}_t=y_t^{de}$ subject to initial values for $y_0,\ldots,y_{1-p}$ and terminal values for $y_{T+1},\ldots,y_{T+k}$ as discussed in Chapters 3 and 4.

In dynamic consistent expectations simulation of a rational expectations model, the disturbance terms are set to their expectation of zero or to known values and the exogenous variables are similarly taken as given or projected either by time series models or judgmental forecasts. These assumptions are appropriate for *ex ante* forecasting or policy analysis in which policy changes are announced in advance. In Chapter 5 we showed how the dynamic solution can be extended to take account of anticipated and unanticipated shocks, temporary or permanent shocks and stochastic simulation.

The static simulation mode can then be derived in a similar fashion to the stochastic simulation procedure as a series of dynamic simulations, starting in successive periods to generate the expectations terms used as explanatory variables in each period of interest. Thus there are at least as many dynamic simulations required as there are time periods in the static simulation horizon. The first period of each dynamic simulation yields one observation of the static simulation. The differences from stochastic simulation are: firstly, that the series of shocks are the realised values of the exogenous variables becoming known; secondly that all initial conditions are taken from the historical record rather than from lagged solution values; third and finally, there is only one sequence of shocks.

As in Chapter 5, we can make alternative assumptions about the information set with respect to the exogenous variables. For example, we could assume that x_t is anticipated at the end of period t–1 whereas x_{t+s}, s>0 remains unanticipated. In this case we forecast values for x_{t+s}, s=1,...T–t and the dynamic solution which then produces $y_{t+1|t-1}$ will simultaneously yield the static simulation from its first period. Similarly, we can assume that x_{t+k} is anticipated at the end of period t–1, forecast values for x_{t+k+s}, s>0 as required, and then solve simultaneously for $y_{t+1|t-1}$ and y_t^{ste}. A more complicated alternative is to make the fully unanticipated assumption of Chapter 5. In this case, x_t only becomes anticipated at the end of period t but y_t still depends on the *ex post* realized value of x_t. We then need two solutions for each period of the static simulation. First we forecast x_s, s=t,...,T and solve for the expectations $y_{t+1|t-1}$ using a dynamic simulation. Holding expectations fixed at these values, we then set x_t equal to its historical value and re–solve for y_t^{ste}. The choice of which assumption is most appropriate will depend on what information assumptions were made in estimation.

Finally, we can use differing assumptions concerning the agent's expectations which are not based on the models' own forecasts. There are two obvious possibilities: perfect foresight or adaptive expectations. In a perfect foresight assumption we can replace agent's expectations with the actual value as in the single equation case. This might be a useful extreme and simplifying assumption — it is certainly computationally more convenient since it requires only a standard static simulation as for a backward–looking model. In an adaptive expectations mode, we could replace the expectations with extrapolative models as suggested by Frenkel *et al.* (1988). If agents are actually using rational expectations formed on the basis of the model, neither of these assumptions will yield a tracking fit as good as when using model generated expectations.

6.3 Historical tracking

In order to assess the adequacy and validity of large–scale economic models, it has long been the practice to compare the solutions of such models with the historical data record. This rest of this chapter presents such a "historical tracking" exercise, which compares the performance of six current models of the UK economy over the period 1978–85. This period began with a severe recession that at the time was forecast with limited success (Barker, 1985), so it is of interest to ask how well models revised in the light of that experience now perform overall. Also, given the interest in evaluating the economic consequences of the Thatcher regime, to quote Buiter and Miller's (1983) title, we consider whether the models' historical tracking record provides an adequate baseline for counterfactual policy analysis exercises. The models examined in this exercise are the quarterly models of the Bank of England (BE), Her Majesty's Treasury (HMT), the National Institute of Economic and Social Research (NIESR), and the London Business School (LBS); and the annual models of the Liverpool University Research Group in Macroeconomics (LPL) and the City University Business School (CUBS). The versions of each model are those deposited with the ESRC Macroeconomic Modelling Bureau in the autumn of 1987. Their properties are reported in Fisher *et al.* (1988) and references therein.

Our general approach is to compute for each period the "static" solution for the endogenous variables of the model, with all lagged variables and current exogenous variables taking their known values and with no other adjustments or interventions to the model. These solutions are then compared with the actual outcomes. This form of solution represents a one–step–ahead *ex post* forecast from the model; it corresponds to the "hands–off *ex post* forecast" calculated as a means of decomposing the error of published *ex ante* forecasts in the Bureau's forecast assessments (Wallis *et al.*, 1986, Ch.4; 1987, Ch.4). Attention is not restricted to the use of the model in a practical forecasting setting however, since the period over which a historical tracking exercise is conducted often includes part of the sample period over which the model was estimated. Indeed, historical tracking performance over the sample period is often used as a measure of system performance during the model specification stage.

This approach should be contrasted with that of dynamic simulation. As mentioned above, this is the standard mode of operating an econometric model in forecasting and policy analysis exercises, but it is inappropriate in the present context, for reasons discussed below. Of the recent examples of historical exercises using UK models, Britton and Whittaker (1982, HMT model), Brooks and Henry (1983, NIESR), Holly and Longbottom (1982, LBS), Dunn *et al.* (1984, BE) and Beenstock *et al.* (1986, CUBS) all concentrate on the dynamic simulation approach, and only the latter two additionally quote static simulation results.

As noted above, historical tracking also has a specific part to play in assessing the adequacy of a model for use in economic policy evaluation and counterfactual analysis, on which a substantial literature exists. The recent U.K. literature has followed largely from experiments by Artis and Green (1982) on the HMT model, although the methods can be traced back to Blinder and Goldfeld (1976) and earlier to Godley and Hopkin (1965). Other recent examples have used a variety of models and include Artis *et al.* (1984, HMT and NIESR models), Saville and Gardiner (1986, NIESR), Matthews and Minford (1987, LPL) and Mackie *et al.* (1989, BE). A necessary (but not sufficient) condition for the validity of such counterfactual exercises is that the model is an adequate approximation to the underlying economic process. The judgment of economic adequacy has a certain subjective element, but we might objectively require a satisfactory statistical approximation. If a model cannot explain the actual systematic variation in the data then it cannot be expected to discriminate between the sources of that variation. However, none of the above cited references contain more than a cursory prior evaluation although, as noted, tracking exercises on these models are available elsewhere.

In the rest of this chapter we first address the methodological issues in historical tracking, leading to our choice of static simulation. For rational expectation models we use the methods described in Section 6.2. In Section 6.5 we then present the results of a systematic historical tracking exercise on each of the models over the period 1978–85, assessing the historical record model by model. In Section 6.6 we undertake cross–model comparisons, using Theil inequality coefficients and forecast encompassing tests. Finally, in Section 6.7 we evaluate the use of the models for counterfactual exercises and suggest some adjustments.

6.4 Static and dynamic simulation residuals

Consider a simplified structural form of a standard dynamic linear model given by equation (2.2)

$$B_0 y_t + B_1 y_{t-1} + C\, x_t \;=\; u_t, \tag{6.19}$$

As noted in Section 6.1, the static solution of the model is based on the reduced form representation of (6.19), and for given initial conditions y_0 and known values of the exogenous and lagged endogenous variables is obtained as

$$\hat{y}_t = -(B_0^{-1} B_1 y_{t-1} + B_0^{-1} C\, x_t), \qquad t=1,\ldots,T. \tag{6.20}$$

The resulting prediction errors $v_t = y_t - \hat{y}_t$ are given by the reduced form disturbances $v_t = B_0^{-1} u_t$.

In a dynamic simulation, the solution values satisfy the recursion

$$B_0 \bar{y}_t + B_1 \bar{y}_{t-1} + C x_t \;=\; 0, \qquad t=1,\ldots,T,\ \bar{y}_0 = y_0 \tag{6.21}$$

hence the dynamic reduced form solution can be expressed as

$$\bar{y}_t \;=\; A^t y_0 - \sum_{i=0}^{t-1} A^i B_0^{-1} C\, x_{t-i}, \qquad t=1,\ldots,T, \tag{6.22}$$

where $A=-B_0^{-1}B_1$. The dynamic simulation errors $w_t = y_t - \bar{y}_t$ likewise satisfy the recursion

$$B_0 w_t + B_1 w_{t-1} \;=\; u_t, \qquad t=1,\ldots,T, \quad w_0 = 0 \tag{6.23}$$

and hence are given as

$$w_t = \sum_{i=0}^{t-1} A^i B_0^{-1} u_{t-i}, \qquad t=1,\ldots,T. \tag{6.24}$$

As is evident from this analysis, both the static and dynamic reduced form errors are linear combinations of the structural disturbances. There is therefore no extra information present in simulation errors: it is simply presented in a different way. Under the "classical" assumptions that the u's are independently and identically distributed (i.i.d.) with mean zero and constant covariance matrix Σ, it follows that v_t and w_t also have mean zero. However, whereas the v's are also i.i.d. with constant covariance matrix $(B_0^{-1})'\Sigma B_0^{-1}$, the w's are heteroskedastic and serially correlated and hence difficult to interpret. In particular, sample second moments of the w's should not be calculated, as their theoretical counterparts, the auto– and cross–covariances, are not constant but vary over the period.

The preceding discussion has not considered coefficient estimation error, and this serves to complicate matters further. For a correctly specified and consistently estimated model, however, the above results hold in the limit. Turning to finite–sample considerations, Pagan (1989) shows that if the structural residuals $\hat{u}_t$ sum to zero over the estimation period, which is true for most econometric estimators, then the reduced form residuals $\hat{v}_t$ also sum to zero over the estimation period. That is, for the estimation period t=1,...,T, we have

$$\sum_{t=1}^{T} \hat{v}_t = B_0^{-1} \sum_{t=1}^{T} \hat{u}_t = 0. \tag{6.25}$$

Turning to the dynamic simulation residuals $\hat{w}_t$, equation (6.24) gives

$$B_0 \sum_{t=1}^{T} \hat{w}_t + B_1 \sum_{t=1}^{T} \hat{w}_{t-1} = \sum_{t=1}^{T} \hat{u}_t \tag{6.26}$$

so that, again summing over the estimation period, we have

$$(B_0 + B_1) \sum_{t=1}^{T} \hat{w}_t - B_1 \hat{w}_T = 0. \tag{6.27}$$

Thus the mean of the dynamic simulation residuals is approximately zero, depending on the value of $T^{-1}\hat{w}_T$, and Pagan describes conditions under which $T^{-1}\hat{w}_T \xrightarrow{p} 0$. The general conclusion is that the property of unbiasedness is delivered automatically, if the estimation and simulation periods coincide. Hence unbiasedness does not represent a useful model validation criterion, being independent of the misspecification or otherwise of the model.

Hendry and Richard (1982) criticise the use of dynamic simulation as a model selection criterion, since the accuracy of a dynamic simulation track depends on the extent to which the model attributes data variance to factors that it treats as being outside the model, irrespective of the justification for doing so. It is not enough simply to agree on a set of variables to be taken as exogenous for the purpose of cross–model comparison, as the winning model may still have little to commend it: the validity of the simulation should be assessed by testing these exogeneity claims. Nevertheless in our experiments we find that the quarterly models have a very similar classification of exogenous variables and similar assumptions about the underlying stance of policy; where potential differences exist we attempt to standardize. For the annual models, however, there are not only important differences in exogeneity and policy stance but also the difficulties of temporal aggregation. As a result the simulations on the LPL and CUBS models are not directly comparable with those on the quarterly models.

In the tracking exercises reported below we accordingly choose to use only one–step–ahead *ex post* forecast errors, that is reduced form residuals, as a measure of system performance. This approach avoids the problems of interpreting dynamic simulation residuals discussed above. The link between the reduced form and structural disturbances must be borne in mind, together with the zero sample mean result, although other problems may limit its applicability. First, the models are typically estimated by single–equation techniques, and in practice the estimation periods of these individual equations may not coincide. Secondly, the empirical models of the UK economy all contain a number of imposed relationships and technical equations, whose residuals may not have zero mean over any period and may not satisfy homoskedasticity or serial independence assumptions. Thus the reduced form residuals may exhibit non–zero means particularly when, as below, the performance over a sub–sample of the historical period available is under consideration. Nevertheless a non–zero mean for the reduced form residuals indicates that the model is incompatible with the data, and hence can be interpreted as evidence of model failure. Such systematic effects in the reduced form residuals can sometimes be traced to single–equation properties, which can then be examined by standard econometric techniques. Errors arising from inappropriate estimation methods are more difficult to attribute, however. Information concerning covariances that have not been picked up in estimation may also emerge in the reduced form residuals, but this again is difficult to attribute to specific parts of the structure.

Finally, it must be remembered that the practical models are nonlinear in variables, and that in general a nonlinear model does not possess an explicit reduced form. The reduced form errors defined by the deterministic solution of a nonlinear system in general do not have zero mean, since they are nonlinear transformations of the structural disturbances (see, for example, Fisher and Salmon, 1986). Hence it may be possible to obtain errors that are systematically non–zero simply due to nonlinearity, although such effects should be of second order of importance relative to misspecification. Nevertheless the interpretation and analysis of summary statistics based on the reduced form residuals, such as mean absolute errors, auto–correlations, variances etc., may be undermined by the nonlinearity of the model, and in the first instance such quantities are used in a purely descriptive manner.

6.5 The historical tracking record of the models

In this section, results are presented on the static simulation errors (reduced form residuals) for a selection of variables across the six models. We focus on four key variables — unemployment (thousands), the annual rate of GDP growth (output measure, percentage points per annum), the annual rate of inflation (consumption price, percentage points per annum) and the nominal exchange rate (trade weighted basket; index points, 1975=100). The historical values of these variables, on a quarterly basis, are shown in Figure 6.1. These exhibit the well–known general movements over this period, including in respect of output growth the recession of 1979–81 and the miners' strike of 1984. In addition two particular short–term episodes should be noted. The jump in price inflation in 1979(3), with an off–setting fall in 1980(3), reflects the introduction of a higher, unified rate of VAT in the June 1979 budget. The sudden fall in output in 1979(1), reversed in 1979(2), is mainly caused by a road haulage strike which delayed exports at the beginning of the year.

The reported errors are calculated as $\hat{y}_t - y_t$, which is conventional in the modelling literature, rather than as $y_t - \hat{y}_t$, which is conventional in the econometric literature; a positive value therefore represents an over–prediction. For each case, the mean error, mean absolute error (MAE), standard deviation (s), root mean square error (RMSE), and first–order autocorrelation coefficient (ρ) are reported in Table 6.1; a descriptive account is given below, model by model.

In using historical tracking results as a model diagnostic, it is helpful to identify the structural equations which may be the source of particular difficulties in the full model solution. To assist in this we consider regressions of the reduced form residuals on single–equation errors, analogous to the relation $v_t = B_0^{-1} u_t$ discussed above. In general everything depends on everything else, but we see that in a number of cases relatively few single–equation residuals explain a substantial proportion of the variance in the reduced form residuals. If the latter exhibit aberrant behaviour then this variance decomposition helps to locate equations of the

Figure 6.1 The historical data record 1978–85.

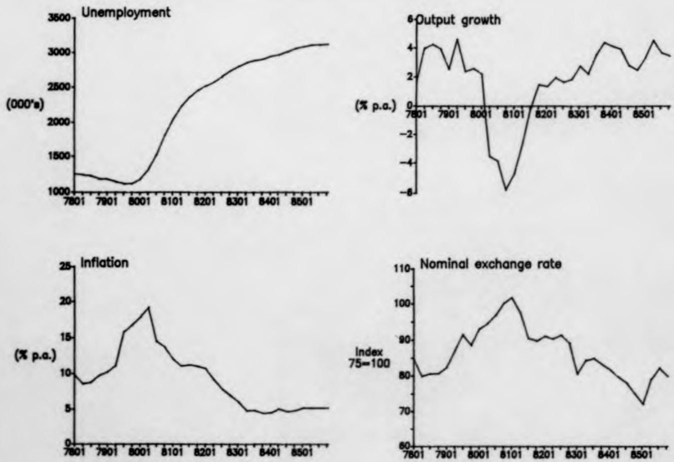


TABLE 6.1: Static simulation residuals, summary statistics

	$\bar{x}$	MAE	s	RMSE	ρ
Unemployment (000's)					
BE	–69	76	57	90	0.79
HMT	157	169	116	195	0.76
NIESR	40	73	77	86	0.47
LBS	–115	80	109	110	0.45
LPL	–48	269	283	288	0.34
CUBS	–190	372	392	436	0.24
Annual Price Inflation (% points)					
BE	–0.13	0.55	0.71	0.73	0.31
HMT	7.20	7.20	1.30	7.34	0.54
NIESR	0.30	0.68	0.77	0.82	0.09
LBS	–0.85	1.13	1.06	1.40	0.76
LPL	–6.50	6.50	2.92	7.12	0.48
CUBS	–0.08	1.38	1.77	1.77	0.40
Nominal exchange rate (index pts 80=100)					
BE	1.5	5.0	6.2	6.3	0.16
HMT	–3.0	4.1	4.2	5.1	0.22
NIESR	–10.8	11.1	5.8	12.7	0.25
LBS	43.6	53.8	53.7	69.2	0.63
LPL	6.3	6.9	5.6	8.4	0.48
CUBS	0.9	2.3	2.7	2.7	0.30
Annual Output growth (% points)					
BE	1.15	1.43	1.51	1.90	0.41
HMT	–0.19	1.64	2.14	2.14	0.18
NIESR	0.003	1.27	1.73	1.73	0.11
LBS	–0.66	1.22	1.30	1.46	0.15
LPL	–2.00	2.80	2.73	3.66	0.70
CUBS	–0.01	1.20	1.40	1.40	–0.63

Notes

$\bar{x}$ – mean error
MAE – mean absolute error
s – standard deviation
RMSE – root mean squared error
ρ – first order auto–correlation coefficient

model that require attention. In the nonlinear case, any transformation applied to the "dependent" variable (the endogenous variable on which the equation is normalised) is reversed before single–equation residuals are calculated. These are then given in the same units as the variable in question, but may not correspond to estimation residuals. In this calculation, right–hand–side endogenous variables assume their known values.

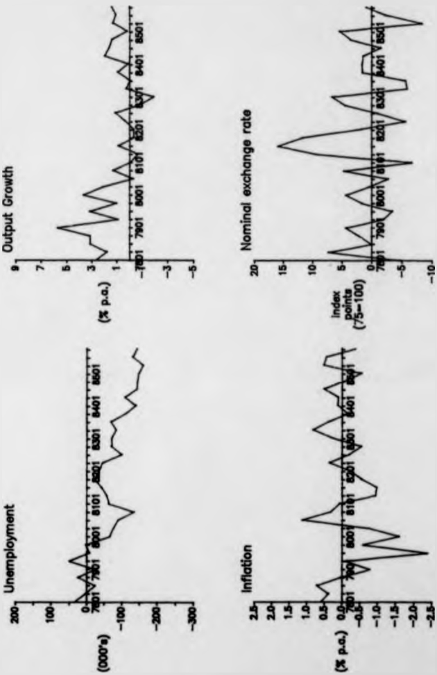
(i) Bank of England (BE)

The historical tracking errors for the BE model are shown in Figure 6.2. The unemployment residuals show two separate periods of consistent under–prediction 1980(1)–1981(3) and 1982(2)–1985(4). The early portion is mainly due to large, systematically positive errors in the total hours equation, which feed positively into manufacturing employment. The drift away over the latter period can be attributed partly to this hours equation and partly to the manufacturing employment equation itself. In an econometric analysis of labour market models, this latter equation, estimated up to 1981(4), exhibited predictive failure and autocorrelation (Wallis *et al.*, 1986, Tables 6.9, 6.13).

On the inflation rate, the single striking feature is the apparent failure to pick up in full the effects over the period 1979(3)–1980(2) of the increase in the VAT rate, particularly in its effects on the price of durables. Much of the remaining error is transmitted from the exchange rate.

The exchange rate residuals are large, with a standard deviation of 6.2 index points and a mean error of 1.5. The main problem is a failure to pick up the depreciation of the nominal rate 1981(1)–(3). The BE exchange rate equation in this model version is not estimated and is entirely backward looking, depending on a number of fourth order lags of explanatory variables. The result is that it heavily lags exchange rate movements. Furthermore, the equation is not homogeneous in prices due to over–differencing of the explanatory variables. The reduced form residuals and single–equation residuals have a near one–to–one relationship, with the latter explaining some 86% of the reduced form error variance.

Figure 6.2 Static simulation residuals: Bank of England model.



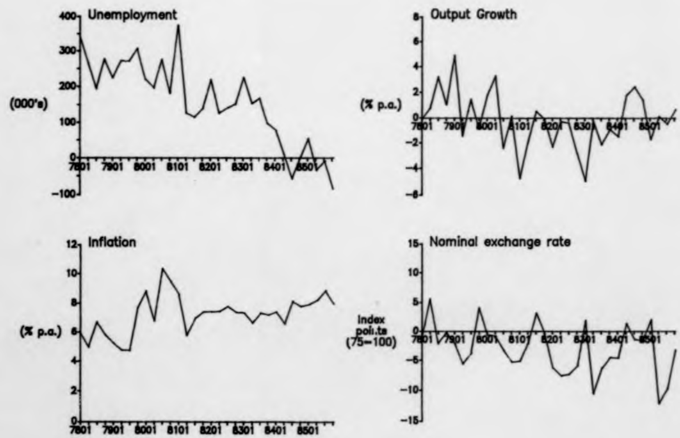
GDP growth is over–predicted in the period 1978(1)–1980(2) and to a lesser extent from 1984(2)–1985(4). There are significant errors on all the components over this period, but exports, imports and stockbuilding are relatively important. In general the residuals on imports from 1980(3)–1985(4) offset errors on consumption, exports and stockbuilding. In particular the reduced–form residual variance can be traced back to systematic single–equation errors on four equations, namely exports of non–manufactured goods, and three stockbuilding equations. Of these latter equations, those for manufacturers' finished goods and work in progress and distributors' stocks failed tests of predictive performance in previous econometric analysis (Wallis *et al.*, 1987, Ch. 5).

(ii) Her Majesty's Treasury (HMT)

When solved in a "hands–off" mode the HMT model produces strikingly large errors in unemployment, as shown in Figure 6.3. The causes can be traced to three technical relationships for employment (by local authorities, central government and nationalized industries) and a price sector equation for the tax content of "other" retail prices. At first sight the unemployment residuals appear to improve over 1984(1)–1985(4), but this is caused by errors elsewhere offsetting those noted above. Principal among these is an over–estimate of unemployed school–leavers (up to 200,000!) which in turn can be attributed to the properties of the three equations determining the number of school leavers. Two other substantial sources of error in this sector, which also account for errors in other sectors, are the treatment of average earnings in the private sector and labour taxes. Average earnings in the private sector (manufacturing and non–manufacturing) are assumed to be proportional to manufacturing earnings only, despite the fact that the non–manufacturing private sector is the larger component: grossing–up earnings in this fashion does not appear to be reliable.

The inflation residuals largely reflect the tax content equation for "other" retail prices, and there are significant jumps due to the higher VAT rate. A number of price equations have systematic single–equation errors, but these equations do

Figure 6.3 Static simulation residuals: Treasury model.



not appear to have been estimated, and in the reduced form their errors mostly cancel out. Other important influences come through from earnings and taxes (see above) and the exchange rate (see below).

The exchange rate tracking errors do not appear to be random, but yield an apparent cycle of some six to seven periods. The exchange rate predictions tend to lag exchange rate movements as well as exhibiting a consistent tendency to under–predict. The exchange rate equations comprise three behavioural relationships that determine the rate itself, the expected rate next quarter and the equilibrium rate. However the latter two variables are not observed, and none of the three equations (nor their reduced form) is freely estimated. In the HMT dataset, the value of the expected exchange rate next period is equal to the actual value in the current period, implicitly representing a no–change or random walk assumption, but this is not consistent with the structure of the expected exchange rate equation. Similarly the equilibrium rate "data" do not appear to satisfy the corresponding equation. Given that all three "structural" equations are determined from a single estimated equation, it is impossible to apportion the error between them.

Output growth is over–predicted in 1978, 1980(1) and 1984(2)–1984(4) and substantially under–predicted in 1981(1), a period of output decline, and 1983(1). The over–prediction in 1978 occurs in a number of GDP components, but particularly exports and investment. Over the whole period the growth errors mask large off–setting errors in the growth of individual GDP components, with consistent over–prediction of stockbuilding and imports and under–prediction of consumption. Reduced form errors in all these sectors can be traced back to single equation residuals in those sectors without any effects dominating.

(iii) National Institute of Economic and Social Research (NIESR)

The NIESR model is a consistent expectations model with 40 forward–looking terms in 10 variables and is sensitive to the assumptions concerning expectations formation. Figure 6.4 shows the results obtained when expectations are based on an information set dated t–1, (i.e. not including the actual values of x_t) using the method described in Section 6.2. The exogenous variable forecasts used in forming expectations are those supplied with the model and used in an earlier exercise by Hall (1987).

The unemployment residuals in Figure 6.4 show one period of consistent under–prediction, namely the two years 1983(4)–1985(4). This is consistent with positive single–equation errors on three of the four employment equations (manufacturing, other private industries, and public services) which were estimated up to 1983(4) only.

The residuals on inflation are generally quite small, although the reduction in inflation in the latter half of the period is not well captured. Two of the largest errors occur in 1979(3) and 1980(3), reflecting the imposition of the higher rate of VAT. The reduced form error variation for inflation can be largely ascribed to the single–equation performance of the price and exchange rate equations.

The exchange rate residuals are the most dramatic and the axes on Figure 6.4 have been stretched to accommodate them. The exchange rate reacts sharply to changes in the balance of trade term, which has exports measured in foreign currency units, and this reaction is exacerbated by forward–looking expectations, since the model also persistently overestimates the deterioration in this measure of trade performance. The result is a consistent under–prediction of the exchange rate, the mean error being –10.8. The variance of the nominal exchange rate residuals around this mean can be largely explained by the single–equation errors on the real exchange rate, which do not exhibit any systematic pattern.

The main features of growth errors are two positive outliers in 1978(4) and 1979(2) and an under–prediction for three consecutive quarters in 1982, but the mean is very close to zero. These results obscure offsetting errors in GDP components. Most notable are two investment equations, for distribution and business services and private sector dwellings, both of which tend to under–predict over the sample. Other consistent errors come through single–equation errors on factor cost adjustment, stocks (particularly distributive trades) and imports. The stocks and investment equations for the distributive sub–sector both failed

Figure 6.4 NIESR model: static simulation residuals.

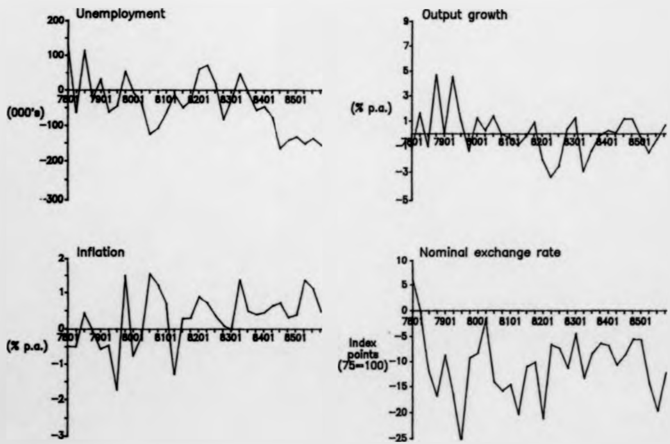
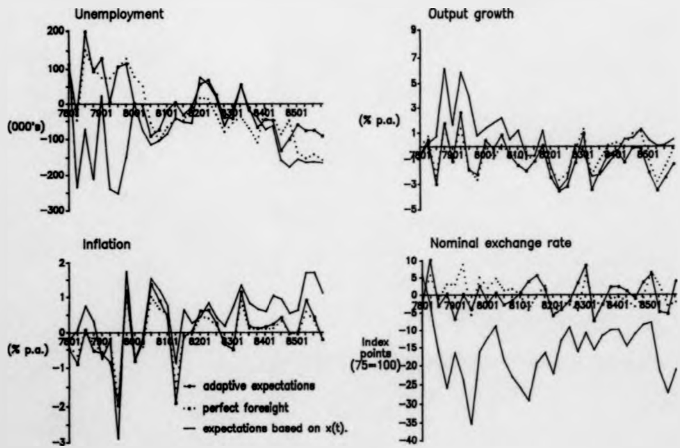


Figure 6.5 Static simulation residuals: NIESR model: variant solutions.



forecasting tests in earlier econometric analysis (see Wallis *et al.*, 1987, Table 6.7, Figure 6.1).

Figure 6.5 shows some variant solutions in which expectations are formed using different information assumptions concerning the exogenous variables. The adaptive and perfect foresight assumptions give results which are very similar to the *ex post* solution shown in Figure 6.4 except for the exchange rate forecast. Now the effect of the balance of trade errors is limited to the current period, and the exchange rate residuals more closely resemble zero mean i.i.d. errors. The relatively poor performance of the model–consistent treatment of exchange rate expectations is suggestive of misspecification of the exchange rate equation. Within the model–consistent expectations framework, using an information set dated at time t makes all the forecasts much worse except for the inflation rate, implying that for this model this assumption is less valid than the t–1 dating used above.

(iv) London Business School (LBS)

In attempting to solve the LBS model over an historical period, considerable problems arise due to extreme movements in the exchange rate. This model has three forward expectations terms, one of which is for the exchange rate (defined as the domestic price of foreign assets, that is, as the reciprocal of the definition used elsewhere in this paper), and in order to generate complete solutions for the expectations variables it proved necessary to make a residual adjustment to the model. The source of the extreme behaviour of the exchange rate is the large financial sector of the model, whose calibration has been criticised (Courakis, 1988) and which, in *ex ante* forecasting, is subject to substantial residual adjustment. The largest single–equation errors occur on overseas bank loans and it was found that adjusting for these errors resulted in a reduction in the exchange rate residuals sufficient to allow the model to solve in all periods. In periods which solved without these adjustments, it was found that their addition had no significant impact on the other key variables – unemployment, output growth and inflation – nor on the variance of the exchange rate residuals, but only their mean.

Figure 6.6 LBS model: static simulation residuals.

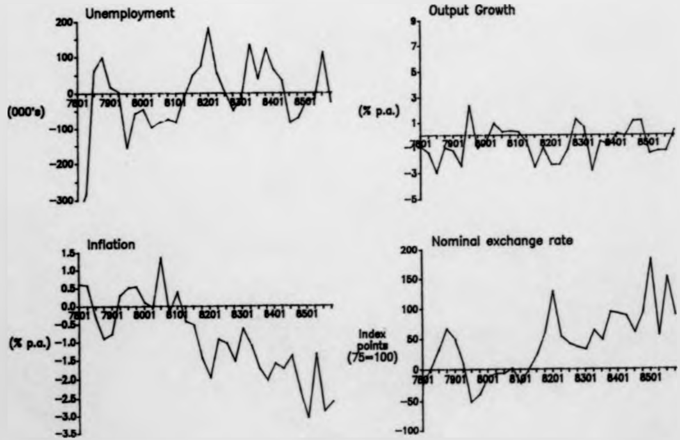
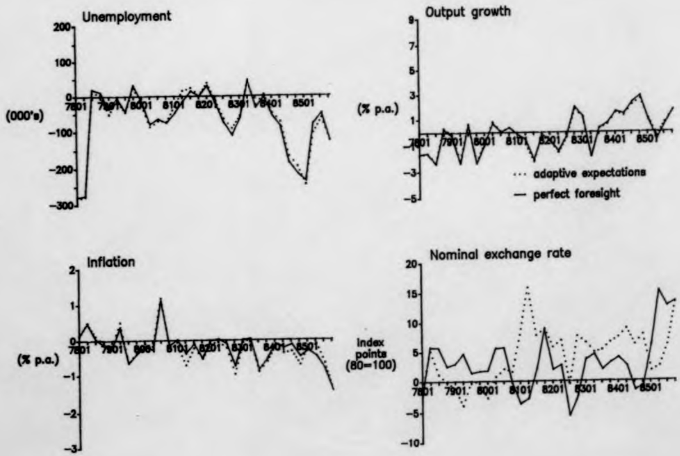


Figure 6.7 Static simulation residuals: LBS model: variant solutions.



The unemployment residuals shown in Figure 6.6 do not exhibit any trend behaviour although there are several sub–periods during which residuals are consistently either positive or negative for at least 4 successive quarters – 1979(3)–1981(1), 1981(3)–1982(2), 1983(2)–1984(2), 1984(3)–1985(2). Also there are two large outliers in 1978(1) and 1978(2), which are due in turn to outlying residuals on the female working population and male working population variables, respectively. Overall, the residuals can be traced back partly to single–equation residuals in the labour sector, but also to price effects feeding through from the rate of change of the exchange rate.

The inflation residuals are inversely proportional to the exchange rate residuals. The residuals on the level of the exchange rate are reflected in every period in its first difference and thereby in the inflation rate for import prices, oil prices etc. The exchange rate residuals are therefore the key problem with this model, as reflected in a mean error of 44 index points and a standard deviation of 54. These residuals are ultimately due to a great many systematic and large errors in the financial sector which are then accentuated by the expectations process. Under perfect foresight or adaptive expectations the maximum residual is 15 rather than the 180 observed in Figure 6.6. These two variant solutions are presented in Figure 6.7, where it is seen that the errors on growth are very similar, those on unemployment slightly less variable and the inflation residuals substantially less variable than in the previous case. In general, the output growth residuals once again mask much larger offsetting errors in all the GDP components.

(v) Liverpool University Research Group in Macroeconomics (LPL)

Static simulation residuals for the Liverpool model are shown in Figure 6.8, using an information set dated t–1 as discussed in Section 6.2. One–step–ahead forecasts become one–year–ahead forecasts in the annual models, which produce larger residuals than the quarterly models, and the scales of the graphs are correspondingly altered.

The unemployment residuals in 1979 and 1980 represent approximately 20 and

Figure 6.8 Static simulation residuals: LPL model.

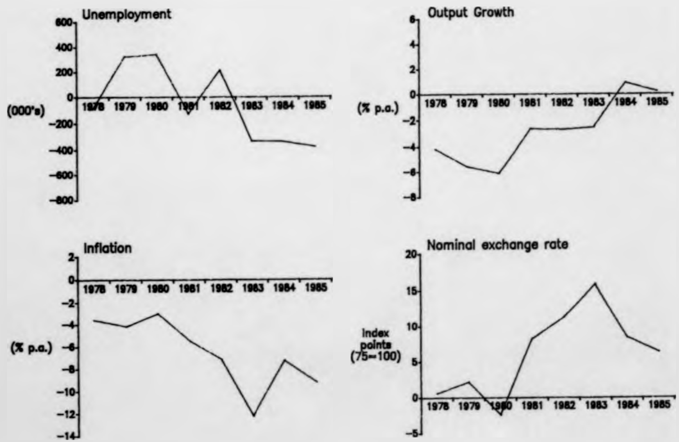
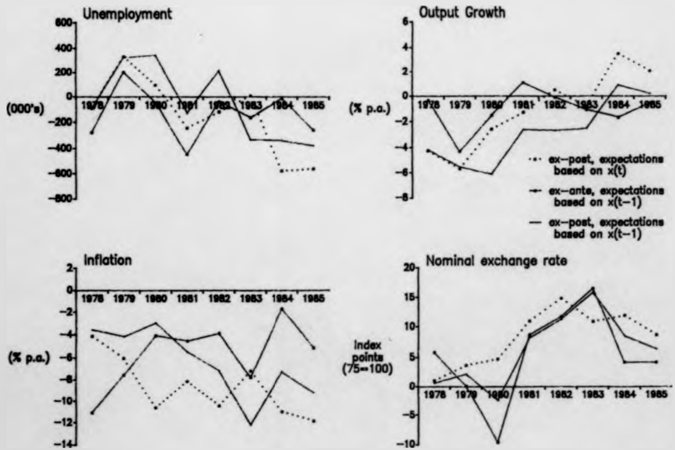


Figure 6.9 Static simulation residuals: LPL model: variant solutions.



25 per cent of the actual values, respectively, and the inflation residuals are negative throughout (mean −6.5). The nominal exchange rate is substantially overpredicted from 1981 to 1985 giving an overall mean error of 6.3, and growth is substantially underpredicted.

The errors on unemployment, inflation and the exchange rate can be traced back to the same fundamental causes. A large number of the equations of this model have systematic and large single–equation residuals. Most notable of these are government expenditure, the stock of goods, real money demand and real wages, these four errors accounting for most of the reduced form error variance for all four key variables. The poor single–equation performance of the model is linked to the way in which the model's coefficients are derived. The original version of the model was based on equations which were freely estimated from data up to 1980 (reported in Minford *et al.*, 1984). However all the important parameters in the current version are imposed (see Wallis *et al.*, 1986, Ch.5; 1987, Ch.5 for tests of the parameterization of the unemployment, wage and goods equations). In *ex ante* forecasting exercises, the remainder of the coefficients – constants, time trends – are calibrated using one or two periods of recent data. It is quite clear that no attempt has been made to fit this version of the model to the data over the period 1978–1985.

A further reason for the reduced form mean error on inflation is due to one of the forecasting rules for the exogenous variables (all of which were supplied with the model). The exogenous "temporary" component of nominal money growth is forecast as zero although it clearly does not have a zero mean over this period, and this contributes to the consistent underprediction of inflation.

Variant solutions have been calculated in which expectations are formed using an information set including exogenous variables dated t and under a perfect foresight assumption and the results are shown in Figure 6.9. Both of these assumptions produce worse results than those reported in Figure 6.8.

(vi) City University Business School (CUBS)

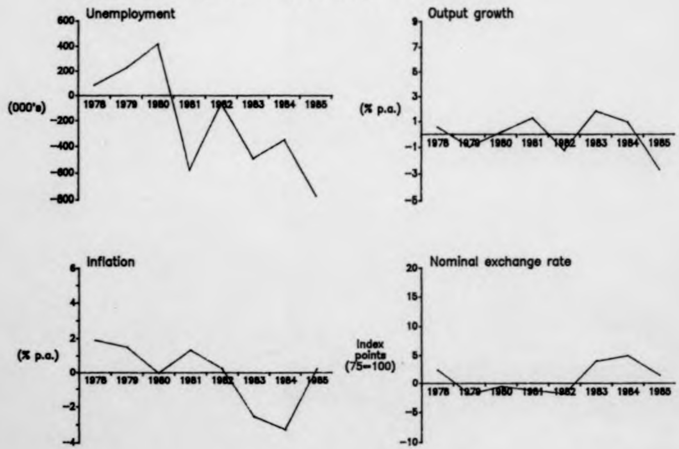
The reduced form residuals for the CUBS model are shown in Figure 6.10. In general the results reveal no systematic evidence of model failure although the variances of the residuals for unemployment and inflation are large (standard deviations 392 and 1.77 respectively). The unemployment residuals can be traced back to outlier single–equation errors on both labour supply (1984–85) and labour demand (1979, 1981 and 1983) which, however, are not obviously systematic. Similarly the main source of the inflation residual seems to be its own single–equation performance which explains 97% of the variance of the static simulation residuals. The inflation single–equation errors also explain 62% of the exchange rate residuals. The over–prediction of growth in 1983–84 can be attributed to a combination of the inflation and labour market errors. It appears that some of these errors are associated with post–sample predictive failure of the corresponding equations.

(vii) Concluding comments

Of the four variables considered across six models, in very few cases do the residuals have the appearance of the zero–mean independent random errors of the econometrics textbooks. Few common explanations for the performance of the models emerge from the preceding discussion, however, and we offer two general observations.

First, with respect to the 1979–81 recession, the historical tracking performance represents an improvement over the poor *ex ante* forecasting performance of those modelling groups who were publishing forecasts at the time (described, for example, by Barker, 1985), with the notable exception of the LPL growth figures. The model modifications that were introduced in the period following these forecasting failures (see, for example, Wallis, 1989, Section 4) appear to dominate the benefit of using actual realized values of exogenous variables, and have indeed yielded better results. On the other hand, the miners' strike of 1984–85 is treated as an extraneous event, and the output residuals for this period

Figure 6.10 Static simulation residuals: CUBS model.



consistently across the models indicate that not only was this an unanticipated event but that no model modifications are called for in its aftermath.

Secondly, of the four variables considered, it is the exchange rate that in general appears to be the most difficult to track. Whereas the determination of the exchange rate has received considerable attention in the theoretical and applied literature, this does not yet appear to have borne empirical fruit, neither in single–equation nor in full model analysis. The treatment of the exchange rate is somewhat different across the models, as noted above, so there is no common explanation except that modelling the exchange rate is hard. Similarly there are few common features in the exchange rate residuals, and their correlations across models are small, the greatest similarity being between the two official models, BE and HMT, which have the most similar exchange rate systems, neither of which is directly estimated.

6.6 Cross–model comparisons

In this section the relative performance of the models is assessed more formally, using a standardized summary statistic (Theil's inequality coefficient) and encompassing tests. Comparisons of macroeconometric models face a number of well–known difficulties: models may differ in size, in temporal and sectoral aggregation, in exogeneity assumptions and in the underlying treatment of policy variables. The different size of models often reflects the degree of sectoral disaggregation, although in the case of the largest of the present models (HMT), this is due to its detailed treatment of the public sector. In focusing on a small number of key macroeconomic variables, we implicitly neglect the information advantage that a larger model may have by virtue of its accounting for a larger number of variables. With respect to temporal aggregation, that is, in this case, the difference between annual and quarterly models, we use a standardized summary measure in an attempt to ensure comparability. With respect to exogeneity, in general there is a consistent treatment of external and policy variables across the quarterly models, but some differences emerge in the annual models. For example, the world price of oil, which is treated as an exogenous variable in four models and a causally prior endogenous variable in a fifth (LBS), does not appear in the LPL model. However this model treats government expenditure as an endogenous variable, determined so as to ensure a constant PSBR/GDP ratio, hence residuals on this variable make a contribution to overall model performance from which other models do not suffer. Both LPL and CUBS treat the interest rate as endogenous and the growth of the nominal money stock as exogenous – a contrary assumption to that of the quarterly models.

(i) Theil inequality coefficients

The Theil inequality coefficient is defined as the ratio of the RMSE of the model forecast to the RMSE of a no–change forecast, that is, the root mean square of the first difference of the relevant variable. For a given variable, a value greater than one indicates that the one–step–ahead *ex post* forecast under consideration here performs less well (in the root mean square sense) than a simple random walk model. The coefficient is unit–free, and standardizes for the difficulty of forecasting different variables, and so can be compared across variables. Finally, by calculating the first difference over a quarter or a year as appropriate, quarterly and annual models can be compared.

The results for the four key variables and six models are shown in the first six rows of Table 6.2. It is seen that only 10 of the 24 coefficients are less than one, indicating that in more than half of the cases considered, the model fits no better than a random walk model; a statement which is true, and more dramatically so, for all explanations of the exchange rate except that of CUBS. Of the four variables considered, the average inequality coefficient is smallest in the case of unemployment.

The ranking of the models according to their performance on each variable is of course the same as that based on the RMSE's given in Table 6.1. If the inequality coefficients are averaged across variables to obtain an overall ranking, then CUBS and BE are ranked first and second, respectively. This calculation

TABLE 6.2: Theil inequality coefficients

Model	Unemployment	Inflation	Nominal Exchange Rate	Output Growth
BE	0.921	0.493	1.826	1.234
HMT	1.996	4.960	1.478	1.390
NIESR	0.880	0.554	3.681	1.123
LBS	1.126	0.950	20.06	0.95
LPL	0.819	2.011	1.654	1.476
CUBS	1.240	0.500	0.531	0.565
HMT(a)	1.242	0.751	1.348	1.435
NIESR(a.e.)	0.798	0.520	1.269	1.218
LBS(a.e.)	1.00	0.363	1.820	0.882
LPL(a)	1.047	0.808	1.087	1.139

Notes

(a) adjusted model – see text
(a.e.) adaptive expectations variant

weights the four variables equally (with zero weight, implicitly, on all other variables), and the fact that the rankings differ across variables indicates that the overall ranking is sensitive to the chosen weights. Thus if unemployment is considered an important variable and its relative weight is increased, then the ranking of the CUBS model declines. Likewise increasing the weight attached to the exchange rate worsens the standing of the BE model. There are no obvious correlations of ranking, variable by variable or overall, with model size.

(ii) Encompassing tests

Forecast encompassing tests between rival models (Chong and Hendry, 1986) rest on the idea of combining forecasts (Bates and Granger, 1969). Consider a combination of one forecast, $\hat{y}_t^1$, based on model M_1, with a competitor, $\hat{y}_t^2$, based on model M_2; if the combined forecast has an error variance that is not significantly smaller than that of $\hat{y}_t^1$, then $\hat{y}_t^2$ appears to offer no additional information, and M_1 encompasses M_2. A comparison may be based on the regression equation

$$y_t = \beta \hat{y}_t^1 + (1-\beta)\, \hat{y}_t^2 + u_t \tag{6.28}$$

and a test of the null hypothesis $\beta = 1$. Chong and Hendry (1986) propose an encompassing test of M_1 by testing the null hypothesis $\alpha_2 = 0$ in the regression

$$y_t - \hat{y}_t^1 = \alpha_2 \hat{y}_t^2 + u_t \tag{6.29}$$

that is, by seeing whether the second forecast helps to explain the error in the first forecast. Their development rests on two assumptions that are not satisfied in our present exercise, however. First, the forecast period is assumed to be outside the estimation sample period and second, the forecasts are assumed to have satisfied within–model tests before encompassing tests are entertained. A common example of such a test is provided by the realization–forecast regression

$$y_t = \alpha + \beta\hat{y}_t + u_t, \tag{6.30}$$

where the relevant null hypothesis is $\alpha = 0$, $\beta = 1$: if this is rejected, then forecast errors could be reduced by exploiting their correlation with the forecast. Given that this occurs in the majority of our cases, we undertake cross–model comparisons by considering the generalized realization–forecast regression

$$y_t = \alpha + \beta_1\hat{y}_t^1 + \beta_2\hat{y}_t^2 + u_t. \tag{6.31}$$

The relation to the Chong–Hendry regression (6.29) is noted by rewriting (6.31) as

$$y_t - \hat{y}_t^1 = \alpha + (\beta_1 - 1)\,\hat{y}_t^1 + \beta_2\hat{y}_t^2 + u_t. \tag{6.32}$$

The coefficient β_2 thus indicates the extent to which M_2 explains M_1's error, given that the performance of M_1 can by itself be improved. Similarly, an encompassing test of M_2 is based on the coefficient β_1 in (6.31). Each model can be treated as the null hypothesis in turn, the pair of tests being based on the single regression (6.31).

The results of a comparison of the two annual models, CUBS and LPL, are shown in Table 6.3, which gives the t–ratios of the regression coefficients in (6.31) for each of four variables. It is seen that for unemployment, LPL encompasses CUBS, in that the CUBS null hypothesis is rejected but the LPL null hypothesis is not rejected, whereas for the exchange rate, CUBS encompasses LPL. For the remaining two variables, neither model can reject the other, treated as a null hypothesis.

Corresponding pairwise comparisons of the quarterly models are shown in Table 6.4. In each case the model treated as the null hypothesis is denoted by the row and that against which it is tested denoted by the column. Also included in the final column is a joint test against all three competing models: denoting the null model as M_1, the table presents the F–statistic (with 3,27 degrees of freedom) for the null hypothesis $\beta_2 = \beta_3 = \beta_4 = 0$, in an obvious generalisation of the regression

TABLE 6.3 Forecast encompassing tests, annual models

	M_1 : CUBS M_2 : LPL	M_1 : LPL M_2 : CUBS
Unemployment	2.06	0.05
Inflation	1.61	1.35
Exchange rate	0.28	4.26
Output growth	1.27	1.21

Note

Entries are (absolute) t–ratios (5 d.f.) in a regression of M_1's error on M_2's forecast.

TABLE 6.4 Forecast encompassing tests, quarterly models

	BE	HMT	NIESR	LBS	All
Unemployment					
BE		0.30	0.03	1.32	0.58
HMT	11.28		4.99	4.92	42.44
NIESR	8.25	2.15		2.66	23.05
LBS	11.87	5.08	5.54		43.94
Inflation					
BE		1.92	1.80	3.51	13.21
HMT	3.86		4.25	5.18	20.82
NIESR	3.46	3.95		7.01	19.24
LBS	2.65	2.70	4.95		10.36
Exchange Rate					
BE		4.11	2.44	2.51	5.37
HMT	0.39		0.49	0.02	0.085
NIESR	4.51	6.02		1.53	11.27
LBS	5.40	7.03	3.00		15.57
Output Growth					
BE		1.06	3.18	4.36	6.98
HMT	4.47		4.79	5.66	16.96
NIESR	4.88	3.77		3.79	12.59
LBS	5.05	3.70	2.72		9.13

Note

Each entry is the (absolute) t–ratio associated with the regression of the residual error from the model denoted by the row, on the forecast from the model denoted by the column. In the column labelled "all", the entries are F–statistics with (3,27) degrees of freedom in a regression of the model error on the forecasts from all the other models. Regressions include a constant and the model's own forecast.

(6.31). It is seen that, for unemployment, the BE model is not rejected, nor is the HMT model for the exchange rate. In all other cases there is no encompassing model, each having something to contribute in explaining the other models' errors.

These comparisons indicate that the models each have deficiencies, and that these differ across the models. Each model could, in some general sense, be improved by taking account of the properties of the other models.

6.7 Using the models for counterfactual simulations

One of the motivations for historical tracking exercises is to assess the reliability of the models in providing a base for counterfactual analysis. In such exercises the models are usually adjusted by adding back their single-equation errors, so that the base trajectory is the historical record itself. Various "what was the effect of ..." or "what would have happened if ..." questions are then addressed by changing domestic policy variables or external variables and comparing the resulting solution with the base. The effect of alternative policy rules may be investigated in a similar fashion. Such exercises produce reliable results, however, only if the underlying model provides an adequate representation of the relevant economic behaviour. As noted above, the assessment of "adequacy" has a subjective element, relating to the theoretical foundations of the model, its overall properties, and the amount of detail it provides on the subject of interest. However, unless the model also provides a good statistical account of the behaviour of variables of interest, it must be deemed inadequate for counterfactual exercises. If the model itself represents a poor explanation of the historical record, omitting various systematic effects that require considerable residual adjustment before the solution is on track, then the estimates it provides under alternative scenarios are correspondingly unreliable. A perturbed solution is subject to the same specification errors as the base solution, and as these are not in general simple additive constants, their effect does not cancel out when the two solutions are compared: two wrongs don't make a right! In some cases, however, it may be possible to make relatively straightforward modifications to a model in order to improve matters. We consider the extent to which the need for such improvements is indicated by the results presented above, and the extent to which they are possible.

One of the most straightforward alterations to a model is simply to drop a variable, that is, to treat a recalcitrant endogenous variable as exogenous, switching out the appropriate equation of the model. In the present case, the obvious candidate for such treatment is the exchange rate, and in *ex ante* forecasting the exchange rate equation is often overridden or, equivalently, subject to heavy residual adjustment. In their counterfactual experiments with the BE model, for example, Mackie *et al.* (1989) treat the exchange rate as exogenous. In general this produces biased results, however, because feedbacks through the deleted equation are automatically suppressed, and because the responses of other endogenous variables to the circumstances that produce the given values of the exogenized variable have to be ignored, as those circumstances are not described. The question "what if the exchange rate were x per cent lower" cannot be answered sensibly unless a plausible scenario in which this occurs is described, and its other ramifications assessed. Of course the problems are less acute if the exogenized variable is relatively constant, but this is typically not the case, a variable being exogenized precisely because its unusual movements cannot be explained.

In the HMT model, seven technical relations dealing with disaggregated public sector employment, school leavers' unemployment and the tax content of certain retail prices cause problems. Although dealing with rather detailed matters, these equations have errors that permeate the full model solution, and they are obvious candidates for switching out. In this case the variables thereby made exogenous would seldom be regarded as of central importance, and corresponding variables on the other models are treated as exogenous. For example, the conventional unemployment measure used in this Chapter excludes school leavers, and so unless unemployed school leavers are a focus of attention, or represent an important transmission mechanism in the labour market, treating this variable as exogenous loses little information. Dropping these seven equations improves the tracking performance overall. In particular the Theil inequality coefficient for inflation is no longer an outlier, but is reduced below 1 in common with the other quarterly models, as shown in the "HMT adjusted" row of Table 6.2.

The LPL model is used in a historical tracking exercise by Matthews and Minford (1987), whose claim that the model "fits past data well" does not appear to be well supported by the results presented above. In fact Matthews and Minford use an amended version of the LPL model. The amendments comprise various adjustments to the money demand equation and intercept adjustments to the stock of goods equation: these remove the systematic error in the growth and inflation trajectories. In the case of the money demand equation, certain tax and benefit variables which proxy a black economy effect are deleted, in order to keep this effect unchanged in their experiments, and compensating intercept adjustments are made. Such amendments centre the growth and inflation residuals plotted in Figure 6.8 on zero, while retaining a similar time profile. The Theil inequality coefficients for these two variables (and for the exchange rate) are reduced, that for inflation being reduced below one, as in other models (see Table 6.2), although the coefficient for unemployment is increased.

It is noted in discussion of the LPL model above that several of the calibrated equations of the model have little empirical support, and in other models examples are cited in which equations have obvious statistical deficiencies. There is then an obvious course of action to obtain a better base solution for counterfactual exercises, namely to respecify such equations in order to improve matters. In several cases preliminary investigation indicates that this can be done.

6.8 Summary and conclusions

This Chapter begins by examining the link between model forms and solution modes. The standard analysis of model forms is extended to the forward–expectations case and solution modes are developed corresponding to the standard single–equation, static and dynamic solutions of a conventional model. The presence of unobservable forward expectations terms allows us to propose a variety of solutions in each case depending on differing assumptions concerning the information set used to form expectations.

In the later sections, a historical tracking exercise is presented which requires the use of all three solution modes and allows us to assess the impact of the information assumptions. We begin this exercise by considering the methodology of historical tracking. We argue that static simulation rather than the more commonly employed dynamic simulation is the appropriate mode of analysis. We apply this approach to six current models of the UK economy over the period 1978–85 including three forward expectations models. The results are disappointing, with Theil inequality coefficients exceeding one in more than half of the cases tabulated (six models, four key variables). There is no clear distinction between consistent–expectations models and more traditional backward–looking models in this respect: the major difficulties lie elsewhere.

Static simulation over a historical period reveals the goodness of fit of the model as a complete system. In some cases it is then possible to trace errors back to particular equations of the model, perhaps inappropriately specified or estimated technical or behavioural relations, and so to identify areas of the model that require attention. Some examples specific to individual models emerge in our analysis, but the most general example concerns those parts of the models that deal with the exchange rate. The difficulty of empirical modelling of the exchange rate is well known, and the poor results in the present paper conform to the prevailing, disappointing consensus.

Irrespective of this particular difficulty, one might ask why it is that models that are in such regular use in forecasting and policy analysis exercises are so poorly validated historically. It would appear that, whereas individual equations are often thoroughly tested against the historical record, once a satisfactory specification at that level is obtained and the equation inserted into the complete model, attention is concentrated on the use of the model in "forward–looking" forecasting and policy analysis. These may be the more pressing problems facing the model proprietor, and "backward–looking" historical validation, system–wide, is correspondingly neglected. When the model does not perform well historically, however, and no good explanation of its failure is available, the reliability of analysis based on the model is called into question.

One specific use of the models which demands greater prior attention to their historical tracking performance is in counterfactual analysis. Here a good account of the historical behaviour of relevant variables is a prerequisite for the study of their possible behaviour under alternative scenarios — questions of what might have been. The device of deleting a troublesome variable may create as many problems as it solves. While in some cases simple alterations to the models can be made to improve matters, the overall inadequacy of the historical solution as a baseline for counterfactual analysis serves to underline the general deficiency of the models.

OPTIMAL CONTROL AND INFLATION–UNEMPLOYMENT TRADE–OFFS

In this chapter we develop methods for the optimal control of large–scale nonlinear models containing forward expectations. We then show how optimal control methods can be used and interpreted in the context of model analysis rather than policy analysis. Finally, we present an empirical examination of the inflation–unemployment trade–offs in three large–scale models which serves to evaluate the cost of the optimal control algorithms and to demonstrate their use in model analysis.

7.1 Optimal control algorithms for forward expectations models

The general optimal control problem was discussed in Section 2.4 and we briefly restate the problem here. In policy optimization we generally have conflicting objectives. These objectives are explicitly stated in a scalar–valued loss function. Optimal policies are those which minimize that loss function subject to the econometric model. It is conventional to employ a quadratic objective function which we re–state from equation (2.15):

$$\min_{X} J^0 = \tfrac{1}{2}(Y-\tilde{Y})'W_y(Y-\tilde{Y}) + \tfrac{1}{2}(X-\tilde{X})'W_x(X-\tilde{X}), \tag{7.1}$$

where capital letters denote stacked vectors of variables e.g. $Y'=(y_1',\ldots,y_T')$; the weighting matrices W_x, W_y are positive semi–definite; tilde denotes a target or desired value (e.g. $\tilde{Y}$). Equation (7.1) is minimized with respect to elements of X subject to the econometric model which may be written in stacked form as in equation (2.14) e.g.:

$$B_0 y_t = B_1 y_{t-1} + Cx_t + u_t \tag{7.2a}$$

becomes

$$DY = GX + BY_0 + U, \tag{7.2b}$$

where

$D = \begin{bmatrix} B_0 & & 0 \\ B_1 & \ddots & \\ 0 & B_1 & B_0 \end{bmatrix}$, $G = \begin{bmatrix} C & & 0 \\ & \ddots & \\ 0 & & C \end{bmatrix}$, $B = \begin{bmatrix} B_1 & & 0 \\ & 0 & \\ & & \ddots \\ 0 & & 0 \end{bmatrix}$ and $Y_0' = (y_0', \ldots 0)$ is an initial condition.

In the forward expectation case it is convenient to separate out the expectations terms as separate variables as in Section (3.6) since we will no longer be imposing consistent expectations in every solution. Hence the model becomes:

$$DY = GX + BY_0 + HY^e \tag{7.3}$$

where $Y^{e\prime}$ is $(y_2^{e\prime}, \ldots, y_{T+1}^{e\prime})$ and (e.g.) $H = \begin{bmatrix} A & & 0 \\ & \ddots & \\ 0 & & A \end{bmatrix}$.

For linear systems with forward expectations, different optimal solutions can be obtained by the use of different first order conditions as shown in Section 2.8. These conditions correspond to different scenarios and, for example, vary between the time–consistent and time–inconsistent optimization. For linear models such solutions have been examined, *inter alia*, by Miller and Salmon (1985), and Cohen and Michel (1988) in continuous time; Hughes Hallett (1987) and Ghosh, Gilbert and Hughes Hallett (1987) in discrete time. In this section we are concerned to develop numerical procedures for determining analogous solutions in nonlinear models. In particular we wish to consider and build on the methods suggested by Hall (1984, 1986), Rustem (1989), Holly and Zarrop (1979, 1983) (used by Bray, 1988). In particular these will be compared to the method used in Wallis *et al.* (1987, Ch.3).

The various methods can be characterised by interpreting them as dynamic games played between two players – the policy maker calculating optimal trajectories, and economic agents setting expectations. The policy maker will always have an explicit objective function. Other agents will behave rationally by formulating expectations to be consistent with the model's predictions. This will be regarded as an optimisation on the part of economic agents. This optimisation may be implicit in a consistent expectations solution procedure or can be made explicit in a loss function as in the penalty function solution for consistent expectations (Section 3.6). It is not necessary for economic agents to act collectively if we assume that they are an homogeneous set acting independently but identically so that in aggregate their actions are like those of one player.

Three solutions will be considered and existing procedures will be critically assessed. Where necessary alternative procedures are proposed. The three solutions can be viewed as:

(a) time–consistent and expectations consistent,

(b) expectations–inconsistent/cooperative,

(c) time–inconsistent and expectations–consistent.

In practice type (b) methods have been used to attempt a type (c) solution. The methods we propose for type (a) and type (c) solutions can be interpreted as nonlinear, discrete time analogues of the solutions proposed by Cohen and Michel (1988).

(a) Time–consistent, expectations–consistent

A time–consistent optimal policy is defined as the best policy from amongst those which the policy–maker has no incentive to change simply due to the passage of time i.e. if the policy–maker re–optimises starting from a later time period, the optimal solution remains unchanged. In section 2 we showed that the optimal policy for our demonstration model was not time–inconsistent if it was not forward looking i.e. changes in the optimal policy in period 2 did not affect the outcome for period 1. This result leads to a general method for obtaining the time–consistent solution as the outcome of a non-cooperative Nash game. A non-cooperative Nash game is played between two players who each take the others reaction function as given. A Nash solution is one for which neither player can unilaterally improve their position. The solution must therefore have the property that it minimizes the objective function for the policy-maker taking agents' expectations as given, and minimizes economic agents' objective functions given the optimal policy settings. We then define the optimization process for economic agents as that of finding consistent expectations given policy settings. By separating out the forward expectations from the policy optimization calculation, we automatically remove any time-inconsistency by artificially setting to zero the derivative which causes inconsistency. Expectations consistency is then enforced by searching across time-consistent strategies until expectations-consistency is achieved.

The solution may be achieved numerically by a Cournot-style process which iterates between the policy maker's optimization problem and the agent's consistent expectations problem as follows. This is analogous to the "guess and iterate" procedure of Cohen and Michel (1988).

Algorithm (a) Nash/non-cooperative

(i) Obtain arbitrary values for the expectations y^e_{t+1}, $t=1,\ldots,T$ (e.g. by a consistent-expectations solution to the base).

(ii) Solve the policy optimization problem by minimizing J^O with respect to the policy instruments subject to fixed expectations, using a conventional nonlinear optimization package, to obtain y^*_t, $t=1,\ldots,T$ and a terminal value y^*_{T+1}.

(iii) Check for consistency of expectations i.e. $|(y^*_{t+1}-y^e_{t+1})/y^e_{t+1}| < \tau$, $\forall$ $t=1,\ldots,T$ for some small τ. If consistency has been achieved then stop. Else go to step (iv).

(iv) Solve the model for consistent expectations given the optimal instrument values calculated in step (ii). Return to step (ii).

If these iterations converge the solution will possess the desired properties of optimality for policy makers, time–consistency of the optimal policy and expectations consistency. The numerical cost may be substantial, equal to $pC_1 + pC_2$ for p iterations of steps (ii)–(iv); where C_1 is the average cost of a consistent expectations solution for given exogenous variable values and C_2 is the average cost of an optimization under fixed expectations.

This algorithm can be made more efficient in a number of ways. Since the process is a two–part iteration such as that used to solve for expectations, we could employ incomplete iterations on either step (ii) or step (iv) as in Chapter 2. One incomplete iteration would be generated by a partial optimisation at step (ii), e.g. only one iteration of the nonlinear optimisation algorithm will usually give a good approximation to the exact optimum. Once expectations have converged, the number of optimization iterations would be extended. An alternative incomplete iteration could be made by taking one (or a few) steps of a penalty function solution for consistent expectations at step (iv). Both step (ii) and step (iv) could be made faster by restricting convergence tests to the variables of interest. Incomplete iterations will generally increase the necessary number of outer loop iterations (p) between steps (ii) and (iv). The most efficient algorithm would be that which finds the best trade–off between C_1, C_2 on the one hand and p on the other.

The only prior attempt to generate this type of optimal time–consistent solution appears to that by Hall (1984, 1986). Hall's algorithm is a special case of our general procedure in which step (iv) is replaced by the following:

(iv′) Set $y^e_{t+1} = \hat{y}_{t+1}$, t=1,...,T where $\hat{y}_{t+1}$ is obtained from step (ii). Return to step (ii).

This algorithm is equivalent to an incomplete loop of the consistent–expectations solution at step (iv) which uses only one iteration of Hall's (1985) first order method. However, this is unlikely to be appropriate for more general models than that investigated by Hall. In his case the consistent–expectations problem is very specific — the expectations terms are confined to one consumption equation.

This apparent short cut could actually make the procedure very expensive and possibly non–convergent in more general models. By using the restricted step (iv′) the cost will be approximately p_2C_2 for p_2 iterations of steps (ii)–(iv′), but p_2 will be at least the number of iterations required to solve the model for consistent expectations using the Hall (1985) algorithm, in which the model stacked over time, is treated as one large equation system. For the NIESR model, p_2 will therefore be in excess of one hundred. For the LBS and LPL models the Hall (1985) algorithm has not yet been found to work successfully at all due to numerical instabilities (see Fisher, Holly and Hughes Hallett, 1985, pp.22–23). Furthermore, the changes induced in the policy variable at step (ii) will affect the solution path. In general this increased variation in the policy variables will increase p_2 above its value for the base solution. At worst this could cause the iterations to become non–convergent. Since both C_1 and C_2 are likely to be large we have an interest in minimizing p and so the restriction implied by (iv′) is unlikely to be beneficial.

A second method for making this algorithm efficient is to obtain good start values. Since the time–consistent solution is of particular interest alongside the time–inconsistent solution then the latter could be used as a start value (or *vice versa* if the time–consistent solution was found to be cheaper to construct).

Alternatively we could use extrapolation parameters to speed up convergence between steps (ii) and (iv). The most obvious possibility here is an extrapolation/damping factor on the optimal policy trajectory i.e. $x^*(s)=\gamma x^*(s-1)+(1-\gamma)\bar{x}(s)$ at iteration (s) where $\bar{}$ denotes the result of step (ii) and * denotes the input into step (iv). Extrapolation parameters are likely to be especially important if incomplete iterations are being employed.

Finally, the time–consistent solution could be generated by linearised versions of the models in which case the computation is much easier. This approach has been developed by Gaines, al–Nowaihi and Levine (1987). In applications where the time–consistency is itself the most important issue, linearisation may be the most efficient method but it is always possible that removing the nonlinearities will alter the conclusions of the study.

The algorithm for time–consistent solutions can be extended to cover games played between economic agents and several policy makers. Such games are particularly appropriate for multi–country models where there are policy makers for each country. A particular extension has recently been given by Bray (1989) following Rustem (1989). In this case step (ii) is replaced by a loop in which the policy optimization problem is solved for each player in turn. In this extension it may be more convenient to retain the penalty function method for solving out the expectations since the same software can be applied. As noted above incomplete optimizations (one step) may reduce the overall cost.

It is important to note that this algorithm is intended to find the optimal time–consistent policy for a government using a discretionary optimization approach i.e. one which is willing, if able, to set policies so as to take advantage of agents' expectations formed in the past. In the new–classical model of Kydland and Prescott, the time–consistent optimal solution only exists if a government cares about inflation as well as unemployment. If a government only cared about reducing unemployment, it is not possible for a time–consistent solution to be achieved: there is always an incentive for such a government to deceive, agents always know that such an incentive exists and the result is an inflation rate of infinity (i.e. the solution given at (2.31) no longer exists). In a new–classical model any persistent attempts by a government to reduce unemployment below the natural rate leads to accelerating inflation. Although this result is based on extreme assumptions which are unlikely to exactly replicated in an empirical model, the feasible solution space of a nonlinear model is generally bounded and relatively small (see the following section). Hence, for certain objective functions there may be no time–consistent, expectations–consistent solution that is a feasible outcome given a government intent on using discretionary policy optimization. The above algorithm may therefore be non–convergent.

(b) Cooperative, expectations–inconsistent

The solution to algorithm (a) is an optimal non–cooperative outcome. The cooperative analogue could arise as the outcome of a cooperative Nash game in which policy makers agreed to stick to their announced policy and agents behaved as if their expectations were inconsistent – such an agreement might still be perfectly rational if agents were compensated in some way. An example might be an agreed incomes policy in which union wage bargainers reduced their claims (as if their expectations of inflation were lower than they really were) in return for agreement on government economic and social policy (e.g. a "social contract"). Such an outcome could be consistent with the optimal cheating solution of the Kydland–Prescott model at equation (2.32).

The solution to a cooperative game is obtained by a joint minimization of the objective function e.g.:

$$\min J^e = J^0 + \tfrac{1}{2}(Y^+ - Y^e)' W_e (Y^+ - Y^e) + \tfrac{1}{2}(Y^e)' W_v (Y^e) \tag{7.4}$$

where Y^e is given as for equation (7.3) but $Y^{+\prime}$ is now given by $(y_2^+, \ldots, y_{T+1}^+)$. The weighting matrices W_e, W_v are positive semi-definite and the magnitude of W_e, W_v relative to W_x, W_y must be agreed by the two cooperating partners. The Y^e variables are now treated as control instruments along with X in J^0 (equation 7.1). If J^0 is empty then this is a generalization of the objective function used in the penalty function solution method examined in Chapter 3.

Algorithm (b) Nash cooperative

(i) Obtain arbitrary values for the expectations y_{t+1}^e, $t=1,\ldots,T$ (e.g. by a consistent expectations solution to the base).

(ii) Solve the joint optimal control problem by minimizing J^e with respect to the policy instruments and the expectation terms.

This solution has certain computational attractions in that the optimal values and expectations are calculated simultaneously and it can be implemented in a standard optimisation software package without any special adjustment. Since the outcome is for a cooperative game, the notion of time–consistency is not relevant. The outcome can only be implemented if the two players contract to stick to their agreed policies/expectations.

This type of algorithm was first suggested by Holly and Zarrop (1979, 1983) not for the cooperative case of the expectations–inconsistent solution but for an expectations–consistent outcome. In order to achieve this solution the weights W_e must be set to be large relative to the W_y, W_x and W_v elsewhere in J^e, so that at the optimum expectations are consistent. In this variant this procedure is then using an unconstrained penalty function approach to give policy optimization subject to the constraint of consistent expectations. In the control literature this is well known to be an ill–conditioned problem (see, *inter alia*, Bertsekas, 1975; Luenberger, 1973, Ch.12) although it does not appear to have been realized in connection with this economic application. The ill–conditioning arises near the optimum when the constraint part of the objective function achieves a near–zero outcome whereas the other part remains non–zero. The problems emerge in a practical sense in that it is difficult to choose W_e. If these weights are too large the solution will remain at the nearest consistent expectations solution; too small and expectations remain inconsistent. In order to overcome this difficulty, the weights W_e should be gradually increased during the iterations. Unfortunately there is no general algorithm in the literature for doing this automatically for this context. However we could amend algorithm (b) as follows. Start with W_e relatively small in step (i), and add a further step (iii).

(iii) Test for consistency of expectations $|(\dot{y}_{t+1} - y^e_{t+1})/(y^e_{t+1})| < \tau$, $\forall$ t=1,...,T. If expectations consistent then stop. If not, then increase the weights W_e and go to step (ii).

Bray (1988) uses algorithm (b) with step (iii) applied manually and found considerable difficulty in locating the optimum. This experience is entirely consistent with our analysis and our own failed empirical attempts to employ this algorithm to optimise subject to consistent expectations.

A further problem with the Holly and Zarrop application is the interpretation of the results. It is claimed (Holly and Zarrop, 1983, pp28–31) that a consistent–expectations solution generated by algorithm (b) is also time–consistent. However, since the optimization is undertaken with expectations changing simultaneously, the forward–looking part of the model is affecting the optimal policy calculation. Hence, the solution must be potentially time–inconsistent in the sense defined in this chapter. This criticism of Holly and Zarrop has also been made recently by Wohltmann and Kromer who argue that the notions of time–consistency and time–inconsistency used by Holly and Zarrop are directly contrary to those of Kydland and Prescott. This criticism is confirmed by our analysis. Since the Holly and Zarrop solution is numerically problematic, we propose to find the time–inconsistent solution from a more appropriate formulation of the problem.

(c) Time–inconsistent, expectations–consistent

The optimal time–inconsistent solution yields the minimum to the policy–makers objective function over all expectations–consistent solutions. The problem takes the form of a Stackelberg leader–follower game since the policy maker takes into account the fact that expectations are formed consistently when he optimizes. Economic agents, however, simply accept the announced policy and form expectations as if that policy is to be pursued. As with all dynamic Stackelberg solutions, this is potentially time–inconsistent as discussed by Kydland and Prescott (1977). With no cheating this would deliver the no–loss solution to the Kydland–Prescott model given by equation (2.31).

We generate the solution by setting up an objective function for the policy maker which is minimised taking into account the consistent expectations formation process and not just a vector of expectations as in algorithm (a). This is a standard optimisation with the additional feature that all derivative evaluations and intermediate iterations use a consistent expectations solution procedure.

Algorithm (c) Stackelberg leader–follower

(i) Solve the optimal control problem to minimize J^O using a consistent expectations solution for all solutions to the model including derivative calculations.

Rather than minimize subject to a consistent–expectations constraint as in the Holly and Zarrop version of algorithm (b), we have implicitly substituted the constraint into the model and this removes the numerical ill–conditioning from the problem. This approach can be viewed as a natural extension to part of the optimal control literature for discrete time models in which the system is stacked over time. For linear models the extension to rational expectations is considered by Hughes Hallett (1987) and Ghosh, Gilbert and Hughes Hallett (1987, pp.253–261). For nonlinear models this general approach was briefly described (but not used) by Fair (1984, p.385) and is used implicitly (but not described) by Hall (1984, 1986).

Computationally algorithm (b) appeals because an expectations model could be slotted into any optimal control algorithm. However when we first used this method (see Wallis *et al*, 1987, Ch.3), it was found that certain features of the expectations model require changes to the routines which have been developed for conventional models. Convergence of the expectations usually requires that the instruments are on a stable trajectory near the terminal date. If this criterion is not met, there may not be a stable solution path for the solution algorithm to locate (see Chapter 4). Hence we specify the control problem over a sub–period of the time–horizon with appropriate smooth projection of the instruments up to the terminal date. These long–run considerations then become an explicit part of the optimization problem.

The procedure also complicates the control algorithm itself. The derivative–based procedures usually construct a Jacobian matrix to control the descent direction. Elements of this matrix take the form $\partial y_{it}/\partial x_{js}$ for target variables $i=1,\dots,n_y$; instrument variables $j=1,\dots,n_x$; and time periods $s,t=1,\dots,T$. This matrix is approximated using finite differences calculated from dynamic solutions to the model. In principle the matrix (which is $(n_y.T)\times(n_x.T)$) needs $n_x.T$ model solutions to construct. Because this is very costly, two restrictions are usually observed:

$$\frac{\partial y_{it}}{\partial x_{js}} = 0, \qquad t<s \tag{7.5a}$$

$$\frac{\partial y_{it}}{\partial x_{js}} = \frac{\partial y_{it+k}}{\partial x_{js+k}}, \; k=1,\dots,T. \tag{7.5b}$$

These two assumptions will hold exactly in a linear model without forward expectations. The first states that current dated variables are not affected by future policy changes, the second states that the effects of policy are not state dependent.

Given these assumptions, only n_x model solutions are required to evaluate the numerical derivatives and the derivative matrix is lower triangular. In a model with forward expectations, the first of these restrictions (7.5a) cannot hold because variables are linked backwards in time as well as forwards. At least $2n_x$ solutions are therefore required, one to construct the lower triangle and one to construct the upper triangle for each instrument. This latter derivative evaluation is performed using an impulse shock to the terminal period. In our experience this is sufficient for well behaved models and instruments, but not generally.

The second assumption (7.5b) is unlikely to be very accurate as an approximation because, apart from the nonlinearity of the model, there is increased dependence on initial and terminal values and we find differences in anticipated and unanticipated effects. This assumption is therefore particularly suspect in the regions of t=0 and t=T. One might therefore need to compute $n_x.T$ solutions just to evaluate the derivatives. For example in the NIESR model the response to a shock can be very sensitive to its introduction date and the derivatives need to be calculated for every period and $n_x.T$ model solutions are required.

These two changes to a standard optimisation procedure can be viewed as beneficial rather than undesirable. Even in a conventional model there is always a terminal period problem in which the algorithm can take advantage of the finite time horizon to attain implausible results near the end period. This can be dealt with by applying a suitable penalty near the terminal date to act as a transversality condition (Dixit, 1976, p114). In a forward expectations model, the problem is more explicit and attention is focused on retaining stability of the solution. Similarly, the response of current variables to future policy changes focuses attention on the question of anticipated and unanticipated effects.

Summary

In this section we have proposed three routines for optimizing policy on nonlinear forward expectations models. The three routines can be shown to yield different solutions. A general method for producing a non–cooperative time–consistent solution is proposed and several possibilities for improving its efficiency are presented. The algorithm of Hall (1984, 1986) is found to be a special case of this general method with a efficiency gain incorporated which is specific to the problem considered. An analogous solution is presented for a cooperative game. This algorithm incorporates that of Holly and Zarrop (1983) for the time–consistent solution as a special case. However, it is found that the Holly and Zarrop version is both numerically problematic and is actually time–inconsistent rather than time–consistent as had been claimed. Finally we present a method for finding the optimal time– inconsistent solution and note some amendments to existing routines which enable us to use the algorithm in a conventional optimisation package.

7.2 Calculating trade–offs

In this section we propose the use of control techniques for model analysis. In particular we show how control and targeting techniques are used to select solutions of interest from the feasible solution space of a model. Within this context we discuss the derivation of trade–offs between conflicting objectives. The definition of the trade–off must be carefully made in terms of the range of possible outcomes for different policies. These trade–offs may then, in principle, be treated as a menu for policy choice (although one would be very cautious in such interpretation due to possible side–effects on other variables). We therefore start by examining the geometry of a model's solution space and how particular outcomes are achieved by different experimental designs. We go on to discuss why certain outcomes are not attainable. Finally we demonstrate how trade–offs can be defined in particular examples.

We begin by examining the range of possible solutions for a macroeconomic model in a particular type of targeting problem. In this problem we choose an instrument (set) to target an endogenous variable (set). We may be able to achieve the target(s) exactly in every period using equation inversion techniques. This solution is compared to a base in which the instrument values are unchanged. These two solutions will then define a range of possible outcomes for the instruments and target values. We then apply optimal control techniques to the same exercise. It will be shown that optimal control can be used as a device for selecting a particular solution, or set of solutions, from within the range defined by the two simulations. In certain circumstances, addressed below, equation inversion techniques fail to produce a solution. In these cases optimal control can be used to obtain a solution for the endogenous variable(s) as close to the target values as possible.

We begin with a simple two variable, static relationship:

$$y_t = \alpha + \beta x_t + u_t, \quad t = 1,\dots,T, \tag{7.6}$$

where y_t is an endogenous variable, x_t is exogenous and u_t is a normally distributed random disturbance term with mean 0 and variance σ_u^2. We assume that the coefficients α, β are known. The solution to (7.6) for some base trajectory of $x_t = \bar{x}_t$ and setting u_t to its expectation of zero is simply:

$$\bar{y}_t = \alpha + \beta \bar{x}_t, \quad t = 1,\dots,T. \tag{7.7}$$

Let us assume that, *ceteris paribus*, $\bar{x}_t$ is the desired level for x_t and that there is also a desired level for $y_t = \dot{y}_t \neq \bar{y}_t$. For $\beta \neq 0$ we can invert equation (7.7) and thus choose $y_t = \dot{y}_t$ to yield $\dot{x}_t$:

$$\tilde{x}_t = \tilde{y}_t/\beta - \alpha/\beta, \quad t = 1,\ldots,T. \tag{7.8}$$

Now $\bar{x}_t$ and $\tilde{y}_t$ represent the desired values for x_t, y_t and we can set up the optimization problem to minimize the joint loss arising from deviations about these desired levels. The optimum choice (x_t^O, y_t^O) will lie between the two outcomes obtained above if we assume that the loss function is constant over time, and quasi–concave. Then we can write

$$\begin{aligned} y_t^O &= k\,\bar{y}_t + (1-k)\,\tilde{y}_t \\ x_t^O &= k\,\bar{x}_t + (1-k)\,\tilde{x}_t \\ 0 &< k < 1 \end{aligned} \tag{7.9}$$

The outcomes $(\bar{x}_t, \bar{y}_t)$ and $(\tilde{x}_t, \tilde{y}_t)$ represent limiting solutions as the cost of deviations from $\bar{x}_t$ relative to those from $\tilde{y}_t$ go to zero and infinity respectively. For a general class of objective functions we can therefore obtain limiting results by two simulations to obtain $\bar{y}_t, \bar{x}_t$ and $\tilde{x}_t, \tilde{y}_t$.

For example, let us assume a simple quadratic objective function to be minimized:

$$\min_x J = \frac{1}{2} \sum_{t=1}^{T} [\, w_y\,(y_t - \tilde{y}_t)^2 + w_x(x_t - \bar{x}_t)^2 \,], \tag{7.9}$$

where w_x, $w_y > 0$, $w_y + w_x = 1.0$. Equation (7.9) is minimized subject to the equation of the econometric model (7.5). The choice of w_y/w_x then characterizes the control problem and yields a unique solution:

$$y_t^o = \alpha + \beta x_t^o \tag{7.10}$$

$$x_t^o = \frac{w_x\bar{x}_t + \beta w_y\bar{y}_t - \alpha\beta w_y}{(w_x + \beta^2 w_y)} = \frac{w_x\bar{x}_t/\beta^2 + w_y\bar{x}_t}{(w_x/\beta^2 + w_y)}$$

$$\Rightarrow k = w_x/(w_x + \beta^2 w_y), \qquad (1-k) = w_y/(w_y + w_n/\beta^2)$$

In Figure 7.1a, we show these solutions graphically for the one target – one instrument case. In the linear model there are, in general, no bounds to the function and the constraint line given by AA is therefore open ended. The preference function can be shown as a set of indifference curves around the bliss point (e.g. lines BB, CC, DD) and the optimal solution as a tangency between an indifference curve and the constraint line i.e. point O_1. As the relative weight between target and instrument changes, the indifference curves are pulled along the axes e.g. to become line EE and a new optimum O_2. The optimal solutions then trace out the model solution space as defined by the constraint line.

In practice models are usually nonlinear such as

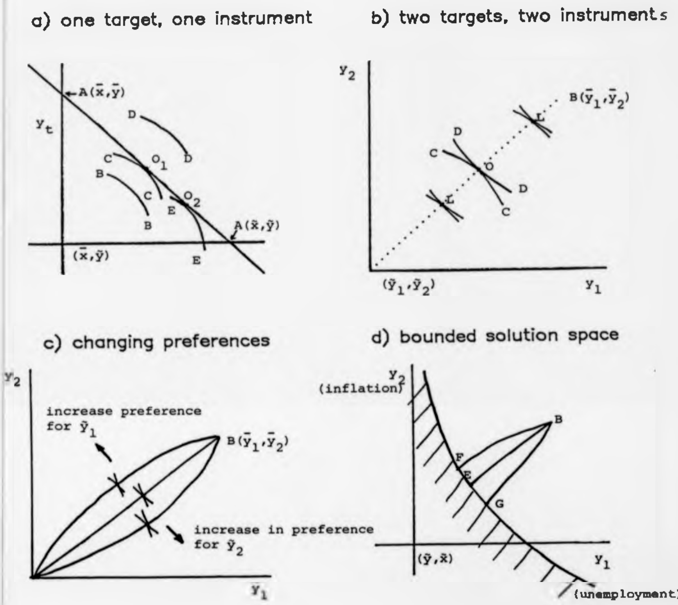
$$f(y_t, x_t\,;\,\theta) = u_t, \quad ,t = 1,\ldots,T. \tag{7.11a}$$

where $f(\cdot)$ is a general nonlinear function and θ is a (vector of) parameter(s). We assume the existence and uniqueness of the implicit functions

$$y_t = g(x_t, u_t\,;\,\theta), \;\; x_t = h(y_t, u_t\,;\,\theta). \tag{7.11b}$$

The presence of nonlinearities simply alters the shape of the constraint line (AA in Figure 7.1a). The boundary points mark limits for x as long as the relationship between x_t and y_t is concave. Static controllability (see below) is preserved as long as $\partial y_t/\partial x_t \neq 0$ i.e. that h exists along the length of the constraint line. However it is possible that nonlinearity will now bound the constraint line at one or both ends

Figure 7.1 Trade-off calculations



(e.g. bounds such as y, x>0) or even to rule out some intermediate solutions.

These results hold for all time periods in this static model and for each equation in a non–simultaneous (uncoupled) model. Dynamic and simultaneous models can be dealt with jointly in the following manner. We stack up the system over time and turn to matrix notation for a simultaneous system as for equation (7.3):

$$DY = GX + BY_0 + U. \tag{7.12}$$

We assume for simplicity that X has as many elements as Y and that all elements of Y and X that are not targets or instruments have been substituted out of the model.

Now we have the base solution, setting U to its expectation of zero:

$$\bar{Y} = D^{-1}\,[G\bar{X} + BY_0] \tag{7.13}$$

and the other limiting solution is given by

$$\begin{aligned} &\tilde{Y} = D^{-1}\,[G\tilde{X} + BY_0] \\ \Rightarrow\quad &\tilde{X} = G^{-1}\,[D\tilde{Y} - BY_0] \end{aligned} \tag{7.14}$$

The optimal control problem is defined by minimization of the objective function (7.1) subject to the model (7.12). The solution is given by :

$$X^o = [\Pi' W_y \Pi + W_x]^{-1}\,[W_x\bar{X} + \Pi' W_y \tilde{Y} - \Pi' W_y D^{-1} BY_0] \tag{7.15a}$$

where $\Pi = D^{-1}G$. Since $D\tilde{Y} - BY_0 = G\tilde{X}$ we have

$$X^o = [\Pi' W_y \Pi + W_x]^{-1}\,[W_x\bar{X} + \Pi' W_y \Pi\tilde{X}] \tag{7.15b}$$

Hence X^o is a linear combination of the vectors $\bar{X}$ and $\bar{X}$ with matrices of weights which sum to the identity matrix.

The simultaneous case is demonstrated graphically in Figure 7.1b for a single–period, two–target, two–instrument problem. We transform the exogenous variable values into endogenous variable space by solving the model. The origin is used to indicate the desired combination of the two targets. The base solution is assumed to give the desired values for the instruments and this solution $(\bar{y}_1, \bar{y}_2)$ is marked as point B. The objective functions can be traced out separately for the targets (e.g. line CC) and the instruments (e.g. line DD) around their separate bliss points. Given a quadratic objective function, the preference functions for the targets will be ellipses. For the instruments, they are ellipses mapped into the target space via the transformation of the econometric model. The two bliss points coincide by assumption with the targeting and the base solutions. Optimal solutions are given by the tangency of two indifference curves around the two bliss points e.g. point O. Different relative weights between targets and instruments will trace out a locus of optimal solutions (e.g. line LL).

The effects of changing the target weights relative to each other (and similarly for the instrument weights) are shown in Figure 7.1c. The immediate effect is to change the shape of the indifference curves thus pulling the optimal locus to one side in favour of the variable who's weight has been increased. Hence we can induce two types of movement: along the optimal locus by changing instrument weights relative to target weights; shifts in the optimal locus by changing the relative preference within the targets (or within the instruments).

The above analysis holds for systems and objective functions where the solutions are feasible (e.g. linear models) and the models are controllable (see below) over that space. We next consider the case where the desired values are not feasible either through non–controllability or nonlinearity.

A solution embodying a set of targets to be achieved by a set of instruments may be made infeasible either because the model is non–controllable (with respect to those targets and instruments) or because it is nonlinear. We observe both these effects on a nonlinear model when it fails to solve (e.g. due to invalid arithmetic calculations such as taking the log of a non–positive number) or the optimisation routine fails to find a descent direction. Linking those observed effects formally to theoretical causes is no easy task but we consider potential reasons why a particular limiting solution may be infeasible.

(i) Controllability

A particular set of n target values y_t can be attained using x_t if $B_0^{-1}C_1$ has rank n . This is the familiar Tinbergen counting rule that there must be as many independent, effective instruments as there are independent targets. If met, this gives us static controllability (see, e.g. Turnovsky, 1977, p.309). The optimal control literature also considers dynamic point controllability – attaining a value y_T with instruments x_t, $t=1,\ldots,T$ and dynamic path controllability – attaining a path $y_p,\ldots,y_T$, $p>1$, with instruments x_t, $t=1,\ldots,T$ and conditions for these are given, *inter alia*, in Turnovsky, (1977, p.333) and Buiter and Gersovitz, (1981, footnote 10, p.42) respectively for the discrete time model. The dynamic solution (7.14) requires a dynamic path for y starting in period 1. In this special case the sufficient and necessary conditions for dynamic path controllability collapse to those of static controllability (see Buiter and Gersovitz's footnote 10). Indeed the solutions are obtained by a recursive sequence which treats each period as a static problem. Hence solution (7.14) requires as many independent, effective instruments as there are independent targets.

In cases where we do not have static controllability, optimal control can still be used to obtain a solution as near as possible to (7.14). Such a solution may or may not possess dynamic path controllability after some point p. In all the cases we consider in this Chapter static controllability will be assumed for a *local* linearisation of the nonlinear model.

(ii) Nonlinearity

Nonlinearity can destroy static controllability in a number of ways. Nonlinear transformations or logical "if" conditions may place bounds on the solution of both endogenous and exogenous variables — potentially reducing the rank of the reduced form coefficient matrix at these boundaries e.g. tax rates, nominal interest rates, unemployment, prices and output may all be bounded at zero.

Secondly, although a model may be statically controllable in a local neighbourhood, this property may be lost gradually as one moves away from this solution, i.e. the multipliers may be state dependent. This is particularly noticeable near the end of a dynamic solution where the state is likely to be further away from the base state. Thus the multipliers and hence the static controllability property may be changing over the time period. The feasible solution space of the model may well vanish, e.g. as unemployment tends to zero. This is often observed in practice when the instrument trajectories are erratic or unstable and the dynamic solution fails.

As long as a base solution exists, optimal control can usually find some solution in a local neighbourhood of that base at the expense of not achieving the targets. It is often difficult to ascertain the precise cause when a nonlinear model fails to solve and sensitivity analysis using optimal control often provides an alternative "second best" solution.

The effect of non–controllability or nonlinearity on our graphical representation is therefore to place bounds on the feasible solution space. Plausible bounds for the case where the two targets are unemployment and inflation are shown in Figure 7.1d . One can see that the optimal loci may be truncated and hence there is a boundary whose location may be of interest.

The aim of this analysis is to show that for feasible objectives, the solution space is defined by two simple simulations. The base solution $(\bar{X}, \bar{Y})$ is a straightforward model solution as long as $\bar{X}$ is a feasible input. The other limit

$(\bar{X}, \bar{Y})$ may be obtainable by simple algorithms for equation inversion. In the linear, model the solution can be obtained directly. For the nonlinear model we use iterative methods known as Type 2 fixes (Rampton, 1984). The solution space then defines the range of possible optimal outcomes for different relative weights between W_y and W_x. In the general simultaneous or static case, the fact that we have weighting matrices means that the solution space is is no longer a straight line but a hyper–surface. Hence each element of X or Y is no longer bounded by the extreme solution and this is implicit in Figures 7.1c,d.

Any one point within this set may be the optimal solution for many sets of weights and every point must be an optimal solution for at least one set of weights (or some other functional form for the objective function). *The choice of weights is therefore simply a question of choosing the solution of interest from the feasible set.*

This section so far has attempted to synthesize a view of the optimization problem as a means of choosing any one solution from a large subset and proposes that the limits of the solution space may themselves be of interest. These limits can often be ascertained by relatively simple simulation techniques. The shape of the feasible set — particularly in the long run for dynamic models — may be of more direct interest to the economist than any one specific solution.

Trade–offs are traced out by linking up particular solutions within, or at the edge of, the feasible solution space. The solutions are generated following the suggestion of Chow and Megdal (1978) by altering the weights within the objective function. If we find a point on the edge of the space (e.g. point E in Figure 7.1d) then we can trace out the boundary by adjusting the relative weights between the targets and adjusting the weights of instruments to targets to ensure that the model solves and no further gains are possible (i.e. by finding points F, G in Figure 7.1d).

It should be noted that when a subset of the exogenous variables are used as instruments, the boundary will depend on which instruments are chosen. The boundary will also depend on the state around which the exercises take place. There is therefore no *unique* boundary to the feasible solution space of a nonlinear model: it is defined only for a particular objective function.

If we begin with a solution within the boundary edge then some normalization is needed to map out a trade–off. Let us consider the specific example of the unemployment–inflation trade–off explored by Chow and Megdal (1978), Henry *et al.* (1982), Mellis (1984) and Wallis *et al.* (1987). We choose two demand–side instruments such as government spending and the income tax rate to target the zero inflation — zero unemployment bliss point. Under fixed nominal interest rates, changes in fiscal policy automatically induce a simultaneous monetary shock. In most macroeconomic models there is either an implicit or explicit Phillips Curve giving a negative trade–off between inflation and (log) unemployment which ensures that the bliss point cannot be obtained. A point on the edge of feasible space is determined by gradually reducing the instrument weights to a minimum. Eventually, very large changes in the instruments will not achieve any further reduction in the objective function. At this point the weights between inflation and unemployment can be altered in order to trace out the nearest achievable points to the bliss point.

Alternatively a trade–off can be traced out by taking an optimal solution within the boundary space and altering the target weights according to some external criterion. One way of deciding on the changes in weight is to maintain a constant ratio of target to instrument costs. Using the analysis given above this has the effect of ensuring that points are traced out by shifts in the locus of feasible points rather than by movements along (caused by changing instrument costs relative to target costs). Alternative suggestions include changing the weights such that the intervention costs remain the same e.g. an inflation–unemployment trade–off evaluated at a constant PSBR.

Finally, trade–offs can be drawn out without the explicit use of optimal analysis. We can take the instruments one–by–one and generate repeated solutions varying the amount by which they are shocked. We can then derive trade–offs between the targets by linking the various combinations available. These solutions (and hence the trade–off) will all be optimal for some unspecified objective function.

7.3 Inflation–unemployment trade–offs: empirical results

One issue in optimal control analysis is to pose reasonable policy problems for which it might be appropriate to use a large–scale nonlinear model rather than a small linear model. The commonest example in the literature seems to be the inflation–unemployment trade–off mentioned briefly in Section 7.3. One possible explanation for this is the distinct nonlinearity caused by some formulations of the Phillips curve relating inflation inversely to unemployment and the implicit restriction of unemployment to be strictly positive. Hence this particular issue may be more interesting to examine on a nonlinear model. In what follows we will treat the inflation–unemployment trade–off as an unknown model property to be elucidated. The use of such a trade–off for policy analysis requires much wider considerations such as the examination of all other variables of interest (e.g. GDP growth or the current account deficit).

As an optimisation algorithm we adopt the time–inconsistent approach (type (c) algorithms) which should allow us to achieve the nearest possible points to the bliss point. Cooperative methods (type (b) algorithms) would achieve a smaller value for the objective function. However, the most important forward expectation in the empirical models is that of the exchange rate. It is unlikely that foreign exchange markets would be willing to undertake an explicit contract with a government over economic policy.

For comparison, we also discuss the results of some optimizations using time–consistent procedures (type (a) algorithms). Our aim in these exercises is then threefold. We wish to evaluate the inflation–unemployment trade–offs in the models; the costs of optimisation of a forward–expectations model; and we wish to examine the different possible solutions from the alternative algorithms.

In a previous exercise (Wallis *et al*, 1987) we were concerned only with the trade–offs and employed optimal control on only one forward expectations model. In this section we summarise that exercise with some additional computational information and compare it with results on two further models, one of which gives conflicting evidence.

TABLE 7.1: Specification of the objective function

(a) Base control exercise

	Desired value	Weight
Targets		
Inflation (% p.a.)	0	1.5
Unemployment (thousands)	500	2.7×10^{-5}
Instruments		
Central government expenditure on procurement (1980 £m)	Base trajectory (4000 rising to 4600)	2.5×10^{-6}
Income tax rate	Base trajectory (0.27 declining to 0.25)	2500
Damping terms		
Change in central government expenditure on procurement	0	2.5×10^{-6}
Change in income tax rate	0	2500

(b) Weight combinations used for trade–off calculations (instrument weights constant)

Relative weight[1]	Inflation	Unemployment
×19	2.85	2.7×10^{-6}
×3	2.25	1.35×10^{-5}
Neutral	1.5	2.7×10^{-5}
÷3	0.75	4.05×10^{-5}
÷19	0.15	5.13×10^{-5}

Note

[1]Relative weight w_1/w_2 in neutral case is multiplied by factor indicated. These values leave the relative costs of instruments to targets unaltered.

(i) LBS model

We use the Autumn 1986 version of the LBS model which is consistent with that used by Wallis *et al* (1987). The inflation–unemployment trade–off is generated using two fiscal policy instruments: government spending and the income tax rate. The model has fixed nominal interest rates and bond sales — hence any PSBR changes are implicitly financed by changes in the money stock. Details of the objective functions employed to generate the trade–off are shown in Table 7.1. Each objective function is minimized over the first 20 quarter period of a 32 quarter solution. The relative preference between inflation ($\dot{P}$) and unemployment (U) is then altered by changing the diagonal weights in the W_y matrix in such a way as to keep the relative preferences of targets to instruments constant (see Table 7.1). A set of 5 results is obtained for a variety of relative preferences. Alternative trade–offs are generated using step shocks to the instruments of varying magnitude.

The results for $\dot{P}$, U in the last four quarters of the twenty, are averaged and plotted in $\dot{P}$, U space. The result is shown in Figure 7.2 as the AA curve. The BB and CC curves are the results of step change simulations using government expenditure and the income tax rate respectively. The slope of the optimized trade–off is clearly seen to be a combination of the slopes derived from simulation analysis but lies nearer the overall target of zero inflation and zero unemployment. The slope and position of this trade–off curve is sensitive to different combinations of relative preferences for between $\dot{P}$ and U and hence is not unique. This particular trade–off is not designed to trace the edge of the solution space but is quite close to it. In general we would expect to be able to predict roughly where that edge will be by combining the two simulation trade–offs.

Even though Figure 7.2 illustrates the fifth year of the simulation, the trade–off generated is quite horizontal and it does not tend to the vertical over time. The results suggest that there is no estimate of the natural rate of unemployment available from the LBS model. Furthermore the trade–off does not

Figure 7.2 LBS model: optimized inflation–unemployment trade–off.

AA, optimized trade–off; BB government expenditure simulation; CC, income tax rate simulation.

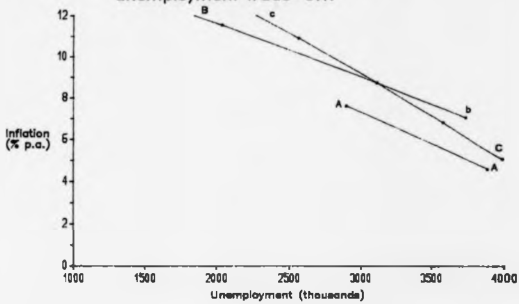
appear to be very nonlinear. The nonlinearities only show in restricting the length of the trade–off (the model won't solve for extreme instrument changes and this places limits on the BB and CC curves).

The costs of the optimization depend heavily on the particular set–up of the objective function and the closeness of the base run to the optimal solution. In this example we have two instruments so there are initial costs of four derivative evaluations (upper and lower triangle for each instrument) and one preliminary solution. Using simulation analysis it was possible to predict rough optimal values for the instruments and so a reasonable preliminary solution was constructed which cost a little more than a standard consistent expectations solution as given in Table 3.1 (whichever scheme is used). The derivative evaluations each cost much less than a standard solution because they are just impulse shocks. Using incomplete inner iterations, each of the four derivatives were obtained for about the same cost as a fixed expectations solution ($\simeq$1000 iterations, per derivative). Each optimization for a point on the trade–off then required between 3 and 5 iterations of the nonlinear optimization routine, each iteration costing the same as a multiplier calculation. The total cost of each optimization exercise (one point on the trade–off) including all calculations, lies between 6 and 8 times those of a single multiplier run.

(ii) NIESR model

The same experiment is repeated for the NIESR model retaining the same instruments and objective function weights. The NIESR model has much less scope for generating a trade–off because suitable instruments all have very similar effects. The optimal trade–off shown in Figure 7.3, for income tax rate and government expenditure, is therefore some distance from the origin. It is also quite short because some of the weight settings caused the model to fail to solve hence this trade–off must be very close to the boundary of the solution space. In other respects the results are very similar to the LBS model but the slope is somewhat flatter. The trade–off slope is generated as a combination of those generated by the

Figure 7.3 NIESR model: optimized inflation–unemployment trade-off.



AA, optimized trade–off; bB government expenditure simulation; cC, income tax rate simulation.

two instruments treated separately but the optimal trade–off lies nearer to the origin.

The NIESR model, with its near unit root (see Chapter 4) and forty expectation terms, is much more difficult to optimise than the LBS model. The model derivatives are much more sensitive to the date that the shocks are introduced. For example, the effect of an income tax rate shock in any one quarter may change sign if introduced in an adjacent quarter instead. For this reason it was found necessary to calculate derivatives in every period. Thus the initial costs are 46 derivative evaluations (23 for each instrument). Each derivative evaluation is used to construct both the upper and lower triangles of the derivative matrix. In this model, each derivative evaluation is relatively more expensive than for the LBS model, about 2/5 the costs of a multiplier calculation using incomplete iterations. This reflects the sensitivity of the model to a shock in any period. Finally, the costs of the optimization iterations were 4 to 8 steps, each costing the same as a multiplier run. Thus the total optimization cost lies between 23 and 27 multiplier calculations, considerably more than for the LBS model.

(iii) LPL model

Using the same instruments, targets and objective function weights, The LPL model gives a trade–off which completely contradicts the results of the NIESR and LBS models for important reasons. The result of optimisation is shown in Figure 7.4 where the trade–off actually has a positive slope bending back at zero inflation yielding a minimum unemployment level of 1.5 million. For this model zero inflation (and hence a natural rate of unemployment) can therefore be an achievable target.

The policy mix to produce this result is simultaneous tax cuts and government expenditure cuts (under balanced financing of the PSBR this implies a decline in real monetary growth). Both income tax and government expenditure simulations alone indicate strong inflation effects and weak unemployment effects. However the income tax rate has potentially strong supply–side effects in this model. In

Figure 7.4 LPL model: optimized inflation–unemployment trade–off.

AA, optimized trade–off; bB government expenditure simulation; cC, income tax rate simulation.

simulation, these beneficial supply–side effects of a tax cut are offset by inflation effects arising from the presence of the tax rate in the money demand equation. This term is meant to represent the effects of the black economy. When the tax rate cuts are made in association with government expenditure reductions the full supply–side effects come through because the inflation rate can be controlled.

In fact the Phillips curve in this model is actually near vertical as indicated by the simulation results and the apparent positive slope of the trade–off reveals shifts in its position as public sector expenditure becomes a smaller share of total expenditure i.e. the supply side effect shifts the Phillips curve.

For the LPL model, the derivatives each have a similar cost to those of a multiplier run as do each of the optimization iterations. The optimization took between 4 and 11 steps, reflecting the distance travelled towards the origin. Each optimization exercise therefore cost between 7 and 14 times a basic multiplier run.

(iv) Time–consistency

Applying algorithm (a) to the time–inconsistent solutions of the NIESR and LBS models (starting from the mid–point of the trade–off) yields convergence in one iteration: there is no incentive to re–optimize in these models for this policy problem. In the Kydland–Prescott example, the incentive to re–optimize depended on the short–run trade–off being non–vertical whereas the long–run trade–off was vertical. This difference rested on the ability of the policy–maker to reduce unemployment (increase output) only through surprise policy changes. In Chapter 5 we found that unanticipated and anticipated policy changes in the NIESR and LBS models had almost identical output (and hence unemployment) effects. Hence there is no ability to reduce unemployment by deceit. Furthermore, the trade–offs in these two models do not appear to become vertical in the long–run. To the extent that they change at all, they become flatter. Therefore, there is no incentive from any source for reneging on optimal policy announcements. This is confirmed numerically by the result that re–optimization with fixed expectations reveals no further gain.

This result suggests that the time–inconsistent algorithm is useful to apply prior to the time–consistent algorithm. It also emphasizes that the results of studies of time–inconsistency are only valid within the context of the properties of the model being used. Simple studies of those properties will reveal, *a priori*, whether there is likely to be a problem of time–inconsistency.

The LPL model is based on the new–classical paradigm. Applying algorithm (a) starting from the best solution ($\dot{P}$=0.0, U=1500) of the time–inconsistent algorithm causes the model to fail to solve. The LPL model has a near–vertical trade–off in the short–run and a vertical trade–off in the long–run. Hence there would appear to be little incentive to re–optimize. The positive trade–off observed in Figure 7.4 exists because there are two independent and effective instruments for both inflation and unemployment. Thus there is no need to indulge in cheating behaviour in order to minimize unemployment and maintain zero inflation. However, there is some non–zero reduction in unemployment for a surprise expansion of government expenditure. The starting point taken embodies a much greater preference for reducing unemployment than inflation (weights from the fourth line of Table 7.1b). The preference for unemployment reductions is such that any possibility of reducing unemployment is taken, whatever the inflation cost. Hence algorithm (a) produces huge increases in government expenditure (and therefore inflation) in the first period for which the model cannot solve.

If we commence algorithm (a) from the "worst" of the solutions produced by algorithm (c) then we find no incentive to re–optimise for the reasons given above. This result shows that the issue of time–inconsistency is not only model dependent but objective function dependent. An objective function sufficiently weighted towards reducing unemployment is more likely to produce time–inconsistency than one weighted to reducing inflation in this model. These results suggests that time–inconsistency (and the generation of high inflation rates) might be a function of unreasonable preferences given the model structure rather than the process of discretionary optimization itself. This conclusion clearly requires further investigation to establish generality.

7.4 Summary and conclusions

In this Chapter we have sought to achieve three objectives. Firstly we have developed a general set of algorithms for the optimal control of nonlinear forward expectations models. These algorithms cover the optimal time–consistent, time–inconsistent and cooperative solutions. Previous attempts to design such algorithms (e.g. Hall, (1984, 1986), Holly and Zarrop, (1983)) are shown to be special cases of our more general approach. Particular attention is paid to computational considerations and improving the efficiency of the algorithms.

Secondly we have examined how optimal control can be used as a tool in model analysis to choose particular solutions from a model's solution feasible space. In particular, we show how the calculation of trade–offs between competing policy targets can be better understood by such an interpretation.

Thirdly we have presented an empirical exercise which obtains the optimal inflation–unemployment trade–offs from three large–scale, nonlinear, forward–expectations models. These calculations allow us some insight into model properties. For example, we find that the NIESR and LBS models have trade–offs which are relatively flat and do not tend to the vertical over time. Thus no estimate of a natural rate of unemployment emerges. Following this result we find no evidence of potential time–inconsistency on these models. For the LPL model we find that time–inconsistency only exists for objective functions heavily weighted towards reducing unemployment. As the new–classical model predicts, this leads to accelerating inflation.

Finally, our calculations also allow us to assess the costs of optimizing such models. We find that the cost varies across models and that a single optimization could cost between 6 and 27 times the cost of a standard multiplier run i.e. expensive but feasible. The time–inconsistent optimisation algorithm worked very well in producing these solutions without numerical difficulties but benefits from the use of good start values and the use of a model with derivatives which are not heavily state–dependent.

Chapter 8

CONCLUSIONS

8.1 Summary

In this thesis we have developed a comprehensive set of techniques for solving, simulating and analysing large scale, nonlinear, econometric models with rational expectations of future–dated variables. We have demonstrated these techniques on three empirical models of the United Kingdom economy.

For solving the basic consistent expectations, two point boundary value problem, we recommend the use of first–order iterative techniques applied to the system stacked over time. These techniques can be substantially improved by taking advantage of the system's block structure to create families of solution algorithms based on two–part algorithms. In comparison with first–order methods, we find the penalty function approach (or the Newton's method equivalent to our approach; Holly and Zarrop, 1983) to be feasible but inefficient for the models considered here. Finally, we find that shooting techniques (Lipton *et al.*, 1982) are an inappropriate formulation of the problem and are unlikely to be feasible on many nonlinear model structures.

It is clear from our results that no one first–order strategy will be dominant for all models. Different structures may require different variations and our methods allow for this possibility.

Each finite–horizon, consistent–expectation solution requires a set of terminal values. If the model is constructed with a saddlepoint property it will have a unique stable long–run solution given the future trajectory of the exogenous variables. The choice of the terminal value is designed to approximate that solution over a finite period. As long as the model clearly possesses the required stability properties, we propose the use of endogenously generated values which assume constant or zero growth rates. Conditions based on the long–run equilibrium should only be used when the model has unstable roots close to the unit circle.

Sensitivity testing should always be carried out on any model to ensure that the solution of interest is insensitive to the terminal date. Any terminal condition which meets this requirement is valid. In simulations, one must take care that the terminal condition does not artificially yield a solution to an otherwise globally unstable model, particularly when policy or control rules are implemented. Terminal conditions constructed from long–run equilibrium analysis should be avoided in simulation wherever possible as they may restrict the range of simulations available.

In a conventional model we perform a variety of simulation experiments: static, dynamic, single equation, and these can be linked with alternative model forms i.e. structural form, reduced form, final form. We present alternative model forms for rational expectation models and derive the analogous solution modes. We can then perform all the usual simulations on a rational expectation model as on a conventional model. In particular, we can undertake static simulations for historical tracking. Our results from such an exercise on six U.K. models over the period 1978–85, suggest that none has a particularly good tracking performance and that tracking the exchange rate is especially difficult. The three models with rational expectations in our exercise perform no better than the three without.

We next consider solutions to the model in the face of shocks, deterministic and stochastic. In all such simulations one must make explicit assumptions about the content of the information set used to generate the expectations. In particular we demonstrate the difference between anticipated and unanticipated, temporary and permanent shocks. This analysis leads to a proposed method for stochastic simulation of rational expectation models. We undertake a stochastic simulation exercise which evaluates the variance of output under alternative financing rules for the PSBR. On the quarterly NIESR and LBS models we find that the variance of output is higher under bond finance than under money finance, whereas the variance of the price level is lower. The annual LPL model gives contradictory results.

Finally, we consider the optimal control of expectations models and its use in model analysis. We propose three general approaches for optimal control algorithms which are shown to deliver different outcomes. These three approaches correspond to different formulations of the policy problem and cover the time–consistent, time–inconsistent and cooperative solutions.

We show how optimal control can be used as a solution selection device from the feasible solution space of a large–scale model and how trade–offs between competing targets can be derived. We then apply our time–inconsistent optimization algorithm on the three models to derive inflation–unemployment trade–offs using government expenditure and the income tax rate as instruments. We find that the quarterly models have trade–offs which are relatively flat. We also find that there is no incentive to re–optimize when the time–consistent algorithm is applied and that this is explained by the properties of the models. The LPL model generates a horizontal optimised trade–off due to the supply–side effects of the income tax rate which shifts the (near–vertical) Phillips curve in the model. On this model we find that extreme preference function weights can generate an incentive to re–optimise when the time–consistent algorithm is applied.

8.2 Directions for future research

This thesis offers a comprehensive analysis of numerical methods for solving, simulating and analysing large–scale, nonlinear models with rational expectations terms. Having established the methods, there are obvious extensions in their application. Firstly, the stochastic simulation approach could be extended to other problems in which higher order moments are of interest. The extension to dynamic simulations will probably require the use of more powerful computing facilities to generate sensible numbers of replications.

The main direction for future research is the application of these methods for policy analysis. In particular the optimal control methods could be applied to a wider range of problems. The models used in this exercise do not reveal properties which lead to substantial problems of time–inconsistency except for extreme preference functions. This relationship between the properties of empirical models, preference functions and time–inconsistency requires further investigation.

Finally, rational expectations models could be used to investigate issues such as uncertainty, sustainability and credibility. These issues have been beyond the scope of this thesis since they require models of how agents learn about, and form expectations of, policy variables. In the first instance, such topics may be best considered with small analytical models and then quantified with the large–scale models.

BIBLIOGRAPHY

ANCOT J.P. (ed.) (1984). Analysing the Structure of Econometric Models. The Hague: Martinus Nijhoff.

ANDERSON, P.A. (1977). 'Rational' forecasts from 'nonrational' models. Discussion paper no.61, Federal Reserve Bank of Minneapolis.

ANDERSON, P.A. (1979). Rational expectations forecasts from nonrational models. Journal of Monetary Economics, 5, 67–80.

AOKI, M. (1976). Optimal Control and System Theory in Dynamic Economic Analysis. Amsterdam: North–Holland.

AOKI, M. and CANZONERI, M. (1979). Reduced forms of rational expectations models. Quarterly Journal of Economics, 93, 59–71.

ARTIS, M.J. and GREEN, C.J. (1982). Using the Treasury model to measure the impact of fiscal policy 1974–1979. In M.J. Artis *et al.* (eds.), Demand Management, Supply Constraints and Inflation. Manchester: Manchester University Press.

ARTIS, M.J., BLADEN–HOVELL, R., KARAKITSOS, E. and DWOLATZKY, B. (1984). The effects of economic policy: 1979–82. National Institute Economic Review, 108, 54–67.

AUSTIN, G.A. and BUITER, W.H. (1982). 'Saddlepoint': A programme for solving continuous time linear rational expectations models. University of Bristol, Discussion Paper no.82/132.

BACKUS, D. and DRIFFILL, J. (1985a). Inflation and reputation. American Economic Review, 75, 530–538.

BACKUS, D. and DRIFFILL, J. (1985b). Rational expectations and policy credibility following a change in regime. Review of Economic Studies. 52, 211–222.

BARKER, T.S. (1985) Forecasting the economic recession in the U.K. 1979–1982: a comparison of model–based ex ante forecasts. Journal of Forecasting. 4, 133–151.

BARRO, R.J. (1976). Rational expectations and the role of monetary policy. Journal of Monetary Economics. 2, 1–33.

BARRO, R.J. (1977). Unanticipated money growth and unemployment in the United States. American Economic Review. 67, 101–115.

BARRO, R.J. (1986). Reputation in a model of monetary policy. Journal of Monetary Economics, 17, 101–122.

BARRO, R.J. and GORDON, D. (1983a). A positive theory of inflation in a natural–rate model. Journal of Political Economy, 91, 589–610.

BARRO, R.J. and GORDON, D. (1983b). Rules, discretion and reputation in a model of monetary policy. Journal of Monetary Economics, 12, 101–121.

BASAR, T. and OLSDER, G.J. (1982). Dynamic Noncooperative Game Theory. New York: Academic Press.

BATES, J.M. and GRANGER, C.W.J. (1969). The combination of forecasts. Operational Research Quarterly, 20, 451–468.

BEENSTOCK, M., WARBURTON, P., LEWINGTON, P. and DALZIEL, A. (1986). A macroeconomic model of aggregate supply and demand for the U.K.. Economic Modelling, 3, 242–268.

BEGG, D.K.H. (1982). The Rational Expectations Revolution in Macroeconomics. Oxford: Philip Allan.

BELLMAN, R., (1957). Dynamic programming. Princeton: Princeton University Press.

BERGSTROM, A.R. (1967). The Construction and Use of Economic Models. London: The English Universities Press.

BERTSEKAS, D.P. (1975). Combined primal–dual and penalty function methods for constrained minimisation. S.I.A.M. Journal of Control. 13, 521–543.

BLACK, F. (1974). Uniqueness of the price level in monetary growth models. Journal of Economic Theory. 7, 53–65.

BLACKBURN, K. (1987). Macroeconomic policy evaluation and optimal control theory: a critical review of some recent developments. Journal of Economic Surveys. 1, 111–138.

BLAKE, A.P. (1986). The Treasury model under rational expectations. Queen Mary College, London, Discussion Paper no.152.

BLAKE, A.P., CHRISTODOULAKIS, N.C. and WEALE, M. (1989). The use of a linear representation in the consistent expectations solution of a non–linear macroeconomic model. Paper presented at the sixth annual conference of the ESRC Macroeconomic Modelling Bureau, University of Warwick, July.

BLANCHARD, O.J. and KAHN, C.M. (1980). The solution of linear difference models under rational expectations. Econometrica. 48, 1305–1311.

BLINDER, A.S. and GOLDFELD, S.M. (1976). New measures of fiscal and monetary policy, 1958–73. American Economic Review. 66, 780–796.

BLINDER, A.S. and SOLOW, R.M. (1973). Does fiscal policy matter? Journal of Public Economics, 2, 319–337.

BRANDSMA, A.S. and HUGHES HALLETT, A.J. (1984). Non–causalities and time inconsistency in dynamic games. Economics Letters. 14, 123–130.

BRAY, J. (1988). Policies for exchange rate stabilisation on the U.K. Treasury model. Paper presented at the fifth annual conference of the ESRC Macroeconomic Modelling Bureau, University of Warwick, July.

BRAY, J. (1989). International economic coordination in the G7 as a dynamic Nash game. Paper presented at the IFAC symposium on Dynamic Modelling and Control of National Economies, Edinburgh, June 27–29.

BRITTON, A. and WHITTAKER, R. (1982). The tracking performance of the Treasury model. Government Economic Service Working Paper no.58.

BROOKS, S. and HENRY, S.G.B. (1983). A tracking exercise. In A. Britton (ed.) Employment, Output and Inflation. pp133–142. London: Heinemann.

BUDD, A., DICKS, G., HOLLY, S., KEATING, G. and ROBINSON, B. (1984). The London Business School econometric model of the U.K.. Economic Modelling. 1, 355–420.

BUITER, W.H. (1984). Saddlepoint problems in continuous time rational expectations models: a general method and some macroeconomic examples. Econometrica. 52, 665–680.

BUITER, W.H. and GERSOVITZ, M. (1981). Issues in controllability theory and the theory of economic policy. Journal of Public Economics. 15, 33–43.

BUITER, W.H. and MILLER, M.H. (1983). Changing the rules: economic consequences of the Thatcher regime. Brookings Papers on Economic Activity. 2, 305–365.

BURMEISTER, E. (1980). On some conceptual issues in rational expectations modelling. Journal of Money Credit and Banking, 12, 800–816.

CALVO, G.A. (1978). On the time inconsistency of optimal policy in a monetary model. Econometrica. 46, 1411–1428.

CALVO, G.A. (1979). On models of money and perfect foresight. International Economic Review, 20, 83–103.

CALZOLARI, G. (1979). Antithetic variates to estimate the simulation bias in nonlinear models. Economic Letters. 4, 323–338.

CALZOLARI, G. and STERBENZ, F. (1989). Alternative specifications of the error process in the stochastic simulation of econometric models. Journal of Applied Econometrics, forthcoming.

CAGAN, P. (1956). The monetary dynamics of hyper–inflation. In M. Friedman (ed.), Studies in the quantity theory of money. Chicago: University of Chicago Press.

CANZONERI, M.B. (1985). Monetary policy, games and the role of private information. American Economic Review. 75, 1056–1070.

CHONG, Y.Y. and HENDRY, D.F. (1986). Econometric evaluation of linear macro–economic models. Review of Economic Studies. 53, 671–690.

CHOW, G.C. (1975). Analysis and Control of Dynamic Economic Systems. New York: John Wiley.

CHOW, G.C. (1980). Estimation of rational expectations models. Journal of Economic Dynamics and Control, 2, 241–255.

CHOW, G.C. (1981). Econometric Analysis by Control Methods. New York: John Wiley.

CHOW, G.C. and CORSI, P. (eds.) (1982). Evaluating the Reliability of Macroeconomic Models. Chichester: John Wiley.

CHOW, G.C. and MEGDAL, S.B. (1978). An econometric definition of the inflation–unemployment tradeoff. American Economic Review. 68, 446–453. Reproduced in G.C. Chow (ed.) (1981). Econometric Analysis by Control Methods. New York: John Wiley.

CHRIST, C.F. (1968). A simple macroeconomic model with a government budget restraint. Journal of Political Economy. 76, 53–67.

COHEN, D. and MICHEL, P. (1988). How should control theory be used to calculate a time–consistent government policy? Review of Economic Studies. LV, 263–274.

COOPER, A. (1987). Modelling uncertainty in simulation analysis. Paper presented at the fourth annual conference of the ESRC Macroeconomic Modelling Bureau, University of Warwick, July.

COURAKIS, A.S. (1988). Modelling portfolio selection. Economic Journal. 98, 619–642.

CURRIE, D.A. (1976). Optimal stabilisation policies and the government budget constraint. Economica. 42, 159–167.

CURRIE, D.A. (1978). Macroeconomic policy and government financing: a survey of recent developments. In M.J. Artis and A.R. Nobay (eds.) Studies in Contemporary Economic Analysis, 1, 65–99, London: Croom Helm.

CURRIE, D.A. (1985). Macroeconomic policy design and control theory – a failed partnership? Economic Journal. 95, 285–306.

DELEAU, M., LE VAN, C. and MALGRANGE, P. (1989). The long run of macroeconometric models. ESRC Macroeconomic Modelling Bureau, Discussion Paper no.21, University of Warwick.

DIXIT, A. (1976). Optimisation in Economic Theory. Oxford: Oxford University Press.

DON, F.J.H. and GALLO, G.M. (1987). Solving large sparse systems of equations in econometric models. Journal of Forecasting. 6, 167–180.

DRIEHUS, W., FASE, M.M.G. and DEN HARTOG, H. (eds.) (1988). Challenges for Macroeconomic Modelling. Amsterdam: North–Holland.

DRIFFILL, J. (1987). Macroeconomic policy games with incomplete information: some extensions. CEPR discussion paper no.159, London.

DUNN, G.P., JENKINSON, N.H., MICHAEL, I.M. and MIDGELY, G. (1984). Some properties of the Bank model. Technical Series Discussion Paper no.9, Bank of England.

FAIR, R.C. (1977). An analysis of a macro–econometric model with rational expectations in the bond and stock markets. Paper presented at the meeting of the Econometric Society, Ottawa, Canada, June 22.

FAIR, R.C. (1979). An analysis of a macro–econometric model with rational expectations in the bond and stock markets. American Economic Review. 69, 539–552.

FAIR, R.C. (1984). Specification, Estimation and Analysis of Macroeconometric Models. Harvard University Press.

FAIR, R.C. and TAYLOR, J.B. (1983). Solution and maximum likelihood estimation of dynamic nonlinear rational expectations models. Econometrica. 51, 1169–1186.

FISCHER, S. (1980). Rational Expectations and Economic Policy. Chicago: University of Chicago Press.

FISHER, P.G. (1989). SLIM: User's manual for version 6.2. Mimeo, University of Warwick.

FISHER, P.G., HOLLY, S. and HUGHES HALLETT, A.J. (1985). Efficient solution techniques for dynamic nonlinear rational expectations models. ESRC Macroeconomic Modelling Bureau, Discussion Paper no.4, University of Warwick.

FISHER, P.G., HOLLY, S. and HUGHES HALLETT, A.J. (1986). Efficient solution techniques for dynamic nonlinear rational expectations models. Journal of Economic Dynamics and Control, 10, 139–145.

FISHER, P.G. and HUGHES HALLETT, A.J. (1987). The convergence characteristics of iterative techniques for solving econometric models. Oxford Bulletin of Economics and Statistics, 49, 231–244.

FISHER, P.G. and HUGHES HALLETT, A.J. (1988a). An efficient solution strategy for solving dynamic nonlinear rational expectations models. Journal of Economic Dynamics and Control. 12, 635–657.

FISHER, P.G. and HUGHES HALLETT, A.J. (1988b). Iterative techniques for solving simultaneous equation systems: a view from the economics literature. Journal of Computational and Applied Mathematics, 24, 241–255.

FISHER, P.G. and SALMON, M.H. (1986). On evaluating the importance of nonlinearity in large macroeconometric models. International Economic Review. 27, 625–646.

FISHER, P.G., TANNA, S.K., TURNER, D.S., WALLIS, K.F. and WHITLEY, J.D. (1988). Comparative properties of models of the U.K. economy. National Institute Economic Review, 125, 69–87.

FISHER, P.G., TANNA, S.K., TURNER, D.S., WALLIS, K.F. and WHITLEY, J.D. (1989). Comparative properties of models of the U.K. economy. National Institute Economic Review, 129, 69–88.

FISHER, P.G. and WALLIS, K.F. (1988). The historical tracking performance of U.K. macroeconomic models, 1978–1985. ESRC Macroeconomic Modelling Bureau, Discussion Paper no.17, University of Warwick. Forthcoming in Economic Modelling, April, 1990.

FISHER, P.G., WALLIS, K.F. and WHITLEY, J.D. (1985). Financing rules and output variability: evidence from UK macroeconomic models. ESRC Macroeconomic Modelling Bureau, Discussion Paper no.7, University of Warwick.

FRENKEL, J., GOLDSTEIN, M., and MASSON, P. (1988). Simulating the effects of some simple coordinated versus uncoordinated policy rules. Discussion paper presented at the Brookings Institution meeting on Macroeconomic Policies in an Interdependent World, Washington, D.C., 12/13 December.

FRIEDMAN, B.J. (1971). Econometric simulation difficulties: an illustration. Review of Economics and Statistics, 53, 381–384.

FRIEDMAN, M. (1968). The role of monetary policy. American Economic Review. 53, 381–384.

GABAY, D., NEPOMIASTCHY, P., RACHDI, M, and RAVELLI, A. (1980). Numerical methods for simulation and optimal control of large-scale macroeconomic models. In A. Bensoussan, P. Kleindorfer and C.S. Tapiero (eds.) Applied Stochastic Control in Econometrics and Management Science. pp115–158. Amsterdam: North-Holland.

GAINES, J. AL-NOWAIKI, A. and LEVINE, P.L. (1987). An optimal control package for rational expectations models. Centre for Economic Forecasting, Discussion paper no.18–87, London Business School.

GANDOLFO, G. (1981). Qualitative Analysis and Econometric Estimation of Continuous Time Dynamic Models. Amsterdam: North-Holland.

GHOSH, S., GILBERT, C.I. and HUGHES HALLETT, A.J. (1987). Stabilizing Speculative Commodity Markets. Oxford: Clarendon Press.

GODLEY, W.A.H. and HOPKIN, W.A.B. (1965). An analysis of tax changes. National Institute Economic Review. 32, 33–42.

GOURIEROUX, C. LAFFONT, J.J. and MONTFORT, A. (1982). Rational expectations in dynamic linear models: analysis of the solutions. Econometrica. 50, 409–425.

GRANGER, C.W.J. (1981). Some properties of time series data and their use in econometric model specification. Journal of Econometrics. 16, 251–276.

GRUNBERG, E. and MODIGLIANI, F. (1954). The predictability of social events. Journal of Political Economy. 62, 465–478.

HAGEMAN, L.A. and YOUNG, D.M. (1981). Applied Iterative Methods. New York: Academic Press.

HALL, S.G. (1984). An investigation of time inconsistency and optimal policy formulation in the presence of rational expectations using the National Institute's Model 7. NIESR Discussion Paper no.71, NIESR.

HALL, S.G. (1985). On the solution of large economic models with consistent expectations. Bulletin of Economic Research. 37, 157–161.

HALL, S.G. (1986). An investigation of time inconsistency and optimal policy formulation in the presence of rational expectations. Journal of Economic Dynamics and Control. 10, 323–326.

HALL, S.G. (1987). Analysing economic behaviour 1975–1985 with a model incorporating consistent expectations. National Institute Economic Review, 120, 75–80.

HALL, S.G. and HENRY, S.G.B. (1985a). Rational expectations in an econometric model, NIESR Model 8. National Institute Economic Review. 114, 58–68.

HALL, S.G. and HENRY, S.G.B. (1985b). Some dynamic experiments on a rational macroeconometric model. Paper presented to the Fifth World Congress of the Econometric Society, Boston, August.

HALL, S.G. and HENRY, S.G.B. (1986). A dynamic econometric model of the UK with rational expectations. Journal of Economic Dynamics and Control, 10, 219–233.

HALL, S.G. and HENRY, S.G.B. (1987). Macroeconomic Modelling. Amsterdam: North–Holland.

HANSEN, L.P. (1982). Large sample properties of generalised methods of moments estimators. Econometrica. 50, 1029–1054.

HANSEN, L.P. and SARGENT, T.J. (1980). Formulating and estimating dynamic linear rational expectations models. Journal of Economic Dynamics and Control. 2, 7–46.

HENDRY, D.F. and RICHARD, J–F. (1982). On the formulation of empirical models in dynamic econometrics. Journal of Econometrics. 20, 3–33.

HENRY, S.G.B., KARAKITSOS, E. and SAVAGE, D. (1982). On the derivation of the "efficient" Phillips curve. The Manchester School. 50, 151–177.

HOLDEN, K., PEEL, D.A. and THOMPSON, J.L. (1982). Modelling the U.K. Economy. Oxford: Martin Robertson.

HOLDEN, K., PEEL, D.A. and THOMPSON, J.L. (1985). Expectations: Theory and Evidence. London: Macmillan.

HOLLY, S. and BEENSTOCK, M. (1980). The implication of rational expectations for the forecasting and simulation of econometric models. E.F.U. Discussion Paper no.72, London Business School.

HOLLY, S. and CORKER, R. (1984). Optimal feedback and feedforward stabilisation of exchange rates, money, prices and output under rational expectations. In A.J. Hughes Hallett (ed.) Applied Decision Analysis and Economic Behaviour. Lancaster: Martinus Nijhoff.

HOLLY, S. and LONGBOTTOM, J.A. (1982). The historical tracking performance of the LBS model. Centre for Economic Forecasting, Discussion Paper no.117, London Business School.

HOLLY, S., RUSTEM, B. and ZARROP, M.B. (eds.) (1979). Optimal Control for Econometric Models. London: Macmillan Press.

HOLLY, S. and HUGHES HALLETT, A.J. (1989). Control, Expectations and Uncertainty. Cambridge: Cambridge University Press.

HOLLY, S. and ZARROP, M.B. (1979). Calculating optimal economic policies when expectations are rational. PROPE Discussion Paper no.30, Imperial College, London.

HOLLY, S. and ZARROP, M.B. (1983). On optimality and time consistency when expectations are rational. European Economic Review, 20, 23–40.

HOWREY, E.P. (1972). Dynamic properties of a condensed version of the Wharton model. In B.G. Hickman (ed.), Econometric Models of Cyclical Behavior. pp601–671. New York: Columbia University Press.

HOWREY, E.P. and KELEJIAN, H. (1971). Simulation versus analytical solutions:the case of econometric models. In T.H. Naylor (ed.), Computer Simulation Experiments With Models of Economic Systems. pp299–319. New York: John Wiley.

HUGHES HALLETT, A.J. (1981). Some extensions and comparisons in the theory of Gauss–Seidel iterative techniques for solving large equation systems. In E.G. Charatsis (ed.), Proceedings of the 1979 Econometric Society Meeting: Essays in Honour of Stefan Valavanis. Amsterdam: North–Holland.

HUGHES HALLETT, A.J. (1985). Techniques which accelerate the convergence of first order iterations automatically. Linear Algebra and Applications, 68, 115–130.

HUGHES HALLETT, A.J. (1986). The convergence of accelerated over–relaxation iterations. Mathematics of Computation, 47, 219–223.

HUGHES HALLETT, A.J. (1987). Forecasting and policy evaluation in economies with rational expectations: the discrete time case. Bulletin of Economic Research, 39, 40–70.

HUGHES HALLETT, A.J. and FISHER, P.G. (1987). Should econometricians use Newton's method for model solution ? ESRC Macroeconomic Modelling Bureau, Discussion Paper no.10, University of Warwick.

HUGHES HALLETT, A.J. and REES, H. (1983). Quantitative Economic Policies and Interactive Planning. Cambridge: Cambridge University Press.

ISARD, P. (1988). Exchange rate modelling: an assessment of alternative approaches. In R.C. Bryant *et al.* (eds.), Empirical Macroeconomics for Interdependent Economies. pp183–201. Washington D.C.: Brookings Institution.

KELLER, H.B. (1968). Numerical Methods For Two–point Boundary Value Problems. Waltham, Mass.: Blaisdell.

KEYNES, J.M. (1936). The General Theory of Employment, Interest and Money. London: Macmillan Press.

KLEIN, L.R. (1986). Economic policy formation: theory and implementation (applied economics in the public sector). In Z. Griliches and M.D. Intrilligator (eds.), Handbook of Econometrics, pp2057–2093. Amsterdam: North–Holland.

KUH, E., NEESE, J. and HOLLINGER, P. (1985). Structural Sensitivity in Econometric Models. New York: John Wiley.

KYDLAND, F.E. and PRESCOTT, E.C. (1977). Rules rather than discretion: the inconsistency of optimal plans. Journal of Political Economy, 85, 473–491.

LAHTI, A. and VIREN, M. (1989). The Finnish rational expectations QMED model: estimation, dynamic properties and policy results. Bank of Finland, discussion paper 23/89, presented at the European meeting of the Econometric Society, Munich, August.

LEVINE, P. and HOLLY, S. (1987). The time inconsistency issue in macroeconomics: a survey. Centre for Economic Forecasting, discussion paper no.CEF–05–87, London Business School.

LIPTON, D., POTERBA, J., SACHS, J. and SUMMERS, L. (1982). Multiple shooting in rational expectations models. Econometrica. 50, 1329–1333.

LONDON BUSINESS SCHOOL (1985). Economic Outlook. vol.10, no.1. London: Gower.

LUCAS, R.E. Jr. (1972a). Econometric testing of the natural rate hypothesis. In O. Eckstein (ed.), Econometrics of Price Determination Conference. Washinton D.C.: Board of Governors, Federal Reserve System.

LUCAS, R.E. Jr. (1972b). Expectations and the neutrality of money. Journal of Economic Theory, 4, 103–124.

LUCAS, R.E. Jr. (1973). Some international evidence on output–inflation trade–offs. American Economic Review, 65, 326–334.

LUCAS, R.E. Jr. (1975). An equilibrium model of the business cycle. Journal of Political Economy. 83, 1113–1144.

LUCAS, R.E. Jr. (1976). Econometric policy evaluation: a critique. In K. Brunner and A.H. Meltzer (eds.), The Phillips Curve and Labour Markets. Supplement to the Journal of Monetary Economics.

LUCAS, R.E. Jr. and RAPPING, L. (1969). Real wages, employment and inflation. Journal of Political Economy. 77, 721–754.

LUCAS, R.E. Jr. and SARGENT, T.J. (eds.) (1980). Rational Expectations and Econometric Practice. Minnesota: University of Minnesota Press.

LUENBURGER, D.G. (1973). Introduction to Linear and Nonlinear Programming. Reading, Mass.: Addison–Wesley.

MACKIE, D., MILES, D. and TAYLOR, C. (1989). The impact of monetary policy on inflation: modelling the U.K. experience 1978–86. In A. Britton (ed.), Policymaking With Macroeconomic Models, pp151–191. Aldershot: Gower

MARWAHA, S.S. (1983). Practical considerations when solving nonlinear rational expectations models. Mimeo, London School of Economics.

MARIANO, R.S. and BROWN, B.W. (1984). Residual based procedures for prediction and estimation in nonlinear simultaneous systems. Econometrica. 52, 321–343.

MARIANO, R.S. (1985). Finite–sample properties of stochastic predictors in nonlinear systems: some initial results. Discussion paper no.266, University of Warwick.

MATTHEWS, K.G.P., MARWAHA, S.S. and PIERSE, R. (1981). RATEXP Mark 2 – a program to solve for rational expectations. Department of Economics Working Paper no.8102, University of Liverpool.

MATTHEWS, K.G.P and MINFORD, A.P.L. (1987). Mrs Thatcher's economic policies 1979–1987. Economic Policy. 5, 57–101.

McCALLUM, B.T. (1976a). Rational expectations and the estimation of econometric models: an alternative procedure. International Economic Review. 17, 484–490.

McCALLUM, B.T. (1976b). Rational expectations and the natural rate hypothesis: some consistent estimates. Econometrica. 44, 43–52.

McCALLUM, B.T. (1983). On non–uniqueness in rational expectations models: an attempt at perspective. Journal of Monetary Economics, 11, 139–168.

McCARTHY, M.D. (1972). Some notes on the generation of pseudo–structural errors for use in stochastic simulation studies. Appendix to Evans, Klein and Saito in B.G. Hickman (ed.), Econometric Models of Cyclical Behavior. pp185–191. New York: Columbia University Press.

MELLISS, C.L. (1984). Some experiments with optimal control on the Treasury model. Government Economic Service Working Paper no.83.

MELLISS, C.L., MEEN, G., PAIN, N. and WHITTAKER, R. (1989). The new Treasury model project. Government Economic Service Working Paper no.106.

MILLER, M.H. and SALMON, M.H. (1985). Dynamic games and the time inconsistency of optimal policy in open economies. Economic Journal. 85, 124–137.

MINFORD, A.P.L., MARWAHA, S.S., MATTHEWS, K.G.P. and SPRAGUE, A. (1984). The Liverpool macroeconomic model of the United Kingdom. Economic Modelling, 1, 24–62.

MINFORD, A.P.L., MATTHEWS, K.G.P. and MARWAHA, S.S. (1979). Terminal conditions as a means of ensuring unique solutions for rational expectations models with forward expectations. Economic Letters. 4, 117–120.

MINFORD, A.P.L., MATTHEWS, K.G.P. and MARWAHA, S.S. (1980). Terminal conditions, uniqueness and the solution of rational expectations models. Mimeo, Department of Economics, University of Liverpool.

MINFORD, A.P.L. and PEEL, D.A. (1983). Rational Expectations and the New Macroeconomics. Oxford: Martin Robertson.

MISHKIN, F.S. (1981). Are market forecasts rational? American Economic Review. 71, 295–306.

MURPHY, C.W., BRIGHT, I.A., BROOKER, R.J. GEEVES, W.D. and TAPLIN, B.K. (1986). A macroeconometric model of the Australian economy for medium term policy analysis. Commonwealth of australia, Economic Planning Advisory Council, Technical Paper no.2.

MURPHY, C.W. (1989). The macroeconomics of a macroeconomic model. Mimeo, Australian National University.

MUTH, J.F. (1961). Rational expectations and the theory of price movements. Econometrica. 29, 315–335.

NAGAR, A.L. (1969). Stochastic simulation of the Brookings econometric model. In J.S. Duesenberry, G. Fromm, L.R. Klein and E. Kuh (eds.) The Brookings model: some further results. pp425–456. Amsterdam: North–Holland.

NATIONAL INSTITUTE OF ECONOMIC AND SOCIAL RESEARCH (1985a). National Institute Economic Review, 114. London: NIESR.

NATIONAL INSTITUTE OF ECONOMIC AND SOCIAL RESEARCH (1985b). Nimodel: a user's guide. NIESR discussion paper, February.

NELSON, C.R. (1975). Rational expectations and the estimation of econometric models. International Economic Review, 16, 555–561.

NIJMAN, T. and PALM, F. (1989). Generalized least squares estimation of linear models containing future expectations. Center for Economic Research, Tilburg University, Discussion Paper no.8902. Presented at the European meeting of the Econometric Society, Munich, August.

ORTEGA, J.M. and RHEINBOLDT, W.C. (1970). Iterative Solution of Nonlinear Equations in Several Variables. New York: Academic Press.

PAGAN, A. (1986). Two stage and related estimators and their applications. Review of Economic Studies. 53, 517–538.

PAGAN, A. (1989). On the role of simulation in the statistical evaluation of econometric models. Journal of Econometrics. 40, 125–139.

PETERSEN, C.E. (1987). Computer simulation of large scale models. The International Journal of Supercomputer Applications. 1.

PRESTON, A.J. and PAGAN, A.R. (1982). The Theory of Economic Policy. Cambridge: Cambridge University Press.

RAMPTON, D. (1984). AMODEL: A program for solving econometric models. Mimeo, Her Majesty's Treasury.

RAU, N. (1985). Simplifying the theory of the government budget restraint. Oxford Economic Papers. 37, 210–229.

ROBERTS, S.M. and SHIPMAN, J.S. (1972). Two–point Boundary Value Problems: Shooting Methods. New York: American Elsevier.

RUSTEM, B. (1989). Optimal, time consistent robust feedback rules under parameter, forecast and behavioural uncertainty. Paper presented at the IFAC symposium on the Dynamic Modelling and Control of National Economies, Edinburgh, June 27–29.

RUSTEM, B. and ZARROP, M.B. (1979). A Newton–type method for the optimization and control of non–linear econometric models. Journal of Economic Dynamics and Control. 1, 283–300.

RUSTEM, B. and ZARROP, M.B. (1981). A quasi–Newton algorithm for the control of large nonlinear econometric models. Large Scale Systems. 2, 105–111.

SALMON, M.H. and WALLIS, K.F. (1982). Model validation and forecast comparisons: theoretical and practical considerations. In G.C. Chow and P. Corsi, (eds.), *op. cit.*, pp.219–249.

SARGAN, J.D. (1983). A reformulation and comparison of alternative methods of estimating models containing rational expectations. Paper presented at the European meeting of the Econometric Society, Pisa, September.

SARGAN, J.D. (1984). Alternative models for rational expectations in some simple irregular cases. DEMEIC Discussion Paper no.A.47, London School of Economics.

SARGENT, T.J. (1973). Rational expectations, the real rate of interest and the natural rate of unemployment. Brookings Papers on Economic Activity, 2, 429–472. Correction of errors, pp799–800.

SARGENT, T.J. (1976). A classical macroeconomic model for the United States. Journal of Political Economy, 84, 207–237.

SARGENT, T.J. (1979). Macroeconomic Theory. New York: Academic Press.

SARGENT, T.J. and WALLACE, N. (1973). Rational expectations and the dynamics of hyperinflation. International Economic Review. 14, 328–350.

SARGENT, T.J. and WALLACE, N. (1975). Rational expectations, the optimal monetary instrument and the optimal money supply rule. Journal of Political Economy. 83, 241–254.

SARGENT, T.J. and WALLACE, N. (1976). Rational expectations and the theory of economic policy. Journal of Monetary Economics. 2, 169–183.

SAVILLE, I.D. and GARDINER, K.L. (1986). Stagflation in the U.K. since 1970: a model–based explanation. National Institute Economic Review, 117, 52–69.

SCHINK, G.R. (1971). Small sample estimates of the variance–covariance matrix of forecast error for large econometric models: the stochastic simulation technique. Phd dissertation, University of Pennsylvania.

SHILLER, R.J. (1978). Rational expectations and the dynamic structure of macroeconomic models. Journal of Monetary Economics, 4, 1–44.

SIMS, C.A. (1980). Macroeconomics and reality. Econometrica, 48, 1–48.

SPENCER, P.D. (1985). Bounded shooting: a method for solving large nonlinear econometric models under the assumption of consistent expectations. Oxford Bulletin of Economics and Statistics, 47, 79–82.

STROTZ, R.H. (1955). Myopia and inconsistency in dynamic utility maximization. Review of Economic Studies. 23, 165–180.

TAYLOR, J.B. (1977). Conditions for unique solutions in stochastic macroeconomic models with rational expectations. Econometrica. 45, 1377–1387.

TAYLOR, J.B. (1979). Estimation and control of a macroeconomic model with rational expectations. Econometrica. 47, 1267–1286.

THEIL, H. (1964). Optimal Decision Rules for Government and Industry. Amsterdam: North–Holland.

TINBERGEN, J. (1936). Kan hier te lande, al dan niet na overheidsingrijpen, een verbbetering van de binnenlandse conjuctuur intreden, ook zonder verbetering van onze exportpositie? Pae–adviezen voor de Vereeniging voor de Staathuishoudkunde en de Statistiek, 's–Gravenhage, pp62–108. Republished in English (1959) as "An economic policy for 1936". In L.H. Klaasen, L.M. Koyck and H.J. Witteveen (eds.), Jan Tinbergen Selected Papers. pp37–84. Amsterdam: North–Holland.

TURNER, D.S., WALLIS, K.F. and WHITLEY, J.D. (1987). Evaluating special employment measures with macroeconometric models. Oxford Review of Economic Policy, 3, xxv–xxxvi.

TURNER, D.S., WALLIS, K.F. and WHITLEY, J.D. (1989). Using macroeconometric models to evaluate policy proposals. In A. Britton (ed.), Policymaking With Macroeconomic Models, pp103–150. Aldershot: Gower

TURNOVSKY, S.J. (1977). Macroeconomic Analysis and Stabilization Policy. Cambridge: Cambridge University Press.

VARGA, R.A. (1962). Matrix Iterative Analysis. Englewood Cliffs, NJ: Prentice–Hall.

VICKERS, J. (1985). Signalling in a model of monetary policy with incomplete information. Oxford Economic Papers. 18, 199–217.

WALLIS, K.F. (1980). Econometric implications of the rational expectations hypothesis. Econometrica. 48, 49–72.

WALLIS, K.F. (1988). Some recent developments in macroeconometric modelling in the United Kingdom. Australian Economic Papers. 27, 7–25.

WALLIS, K.F. (1989). Macroeconomic forecasting: a survey. Economic Journal. 99, 28–61.

WALLIS, K.F. (ed.), ANDREWS, M.J., BELL, D.N.F., FISHER, P.G. and WHITLEY J.D. (1984). Models of the UK Economy: A Review by the ESRC Macroeconomic Modelling Bureau. Oxford: Oxford University Press

WALLIS, K.F. (ed.), ANDREWS, M.J., BELL, D.N.F., FISHER, P.G. and WHITLEY J.D. (1985). Models of the UK Economy: A Second Review by the ESRC Macroeconomic Modelling Bureau. Oxford: Oxford University Press.

WALLIS, K.F. (ed.), ANDREWS, M.J., FISHER, P.G., LONGBOTTOM, J.A. and WHITLEY J.D. (1986). Models of the UK Economy: A Third Review by the ESRC Macroeconomic Modelling Bureau. Oxford: Oxford University Press.

WALLIS, K.F. (ed.), FISHER, P.G., LONGBOTTOM, J.A., TURNER, D.S. and WHITLEY J.D. (1987). Models of the UK Economy: A Fourth Review by the ESRC Macroeconomic Modelling Bureau. Oxford: Oxford University Press.

WESTAWAY, P. (1989a). Does time inconsistency really matter? Paper presented at the IFAC symposium on Dynamic Modelling and Control of National Economies, Edinburgh, June 27–29.

WESTAWAY, P. (1989b). Partial credibility: A solution technique for econometric models. Paper presented at the IFAC symposium on Dynamic Modelling and Control of National Economies, Edinburgh, June 27–29.

WESTAWAY, P. and WHITTAKER, R. (1986). Consistent expectations in the Treasury model. Government Economic Service Working Paper no.87.

WHITTAKER, R., WREN–LEWIS, S., BLACKBURN, K. and CURRIE, D.A. (1986). Alternative financial policy rules in an open economy under rational and adaptive expectations. Economic Journal. 96, 680–695.

WICKENS, M.R. (1982). The efficient estimation of econometric models with rational expectations. Review of Economic Studies. 49, 55–67.

WILCOXEN, P.J. (1989). A fast algorithm for solving rational expectations models. Mimeo, Impact Research Centre, University of Melbourne.

WOHLTMANN, H–W. and KRÖMER, W. (1989). On the notion of time–consistency. European Economic Review, 33, 1283–1288.

WREN–LEWIS, S. (1989). Comment on Chapter 6, Mackie *et al.*, in A. Britton (ed.), Policymaking With Macroeconomic Models. pp103–150. Aldershot: Gower

YOUNG, D.M. (1971). Iterative Solution of Large Linear Systems. New York: Academic Press.

Official Publications

Cmnd. 7148 (1978). Report of the Committee on Policy Optimisation. London: HMSO.

APPENDIX: MODEL VINTAGES

The models used in this thesis are those supplied by the ESRC Macroeconomic Modelling Bureau on its User service over the period Autumn 1984 to Autumn 1989. Each chapter used the following vintage of model.

Chapter 3

LPL and LBS models: Autumn 1985.

Chapter 4

LPL model: Autumn 1984 and Autumn 1985.

NIESR and LBS models: Autumn 1985.

Chapter 5

LPL, NIESR and LBS models: Autumn 1985.

Chapter 6

BE, HMT, CUBS, LPL, NIESR and LBS models: Autumn 1987.

Chapter 7

LPL, NIESR and LBS models: Autumn 1986.

D91830