Policy Evaluation and Design in the Light of Rational Expectations

Andrew John Snell

Submitted for the Degree of Doctor of Philosophy

University of Warwick
Department of Economics

November 1984

To Clare, Ellen and Melissa

Contents

Preface

The bulk of this thesis was written at Warwick University during the years 1979 to 1982 and I gratefully acknowledge the financial support of the SSRC throughout this period.

I am much indebted to my research supervisor Professor K.F. Wallis without whose experienced guidance and encouragement this thesis would not have materialised. Of the several others who were generous with their time in providing constructive criticism and advice, Mark Salmon, Norman Ireland, Ian Tonks and Richard Pierse deserve a special mention.

I would also like to thank Shirley Seal and Bobbie Coe for their efforts in typing a script that for the most part was in mathematics and Greek, languages alien to them both.

Finally I wish to express gratitude to my family and friends, and in particular to my mother and father, Patricia Hartley and Scarlett Palmer, for their encouragement and support throughout the writing of this thesis.

Declaration

This thesis contains my own work and no part of it was a result of any research done in collaboration with other people.

This thesis examines certain key problems that the existence of forward rational expectations poses for policy analysis. The separate stages of estimating, testing and solving an econometric model are dealt with in turn.

The main body of original work is in chapters four and six. In chapter four the problem of the existence of a continuum of solutions to a rational expectations model is addressed. We show that the existing practice of imposing terminal conditions is arbitrary and a procedure is advanced which in principle at least, can be used to estimate the solution jointly with the parameters. In chapter six analytical closed forms for the first order conditions of the likelihood function of the endogenous variables of a general rational expectations model are derived. We believe this is a major contribution to the literature because it opens the door to computationally efficient and cheap likelihood estimation, something not previously available. The first order conditions for a class of models with no predetermined variables has been programmed in Fortran IV and this has been used in chapter seven to estimate a model of financial asset demands. A likelihood ratio test of restrictions implied by rational expectations is comfortably passed so establishing empirical support for the hypothesis.

Other original work in the thesis is contained in chapter five. Here we scrutinise the validity of a simulation technique advanced by Fair and Anderson which it is claimed solves a standard non rational model to yield an approximate rational expectations solution. The results of the chapter suggest the method is better in certain circumstances than in others and these circumstances pertain to the make up of the model in question.

Finally, chapters two and three cast a critical eye over the policy analysis literature to which a minor contribution is made.

1. Introduction

1.1 The subject matter of policy analysis

It is often argued that for economic research to be worthwhile it must involve economic policy at some stage. It was not until 1939 however that a formal framework for the study of economic policy was developed. Since this time it has evolved as a subject within macroeconomics in its own right. The subject, called policy analysis has two components: policy evaluation and policy design.

The first of these concerns itself with an assessment of the effects of various policy 'instruments' such as tax rates, interest rates and government expenditure on various 'targets' such as nominal GDP and the rate of price inflation.

The procedure starts with the development and estimation of an econometric model as large in size and as complex in detail as the objective of the study demands. More often tha not nowadays such models are large, explaining most or all of the key variables in an economy. The coefficients of this model are then scrutinised either analytically or numerically so that the effects of various policies can be quantified. These end products are called policy multipliers. When the instrument(s) in question is (are) discrete in nature a more qualitative analysis may be appropriate. Exchange controls and prices and incomes policies are two notable policy instruments that fit this mould.

The second component of policy analysis is concerned with the formulation of a policy or a policy rule that is expected to achieve objectives that are optimal by some criterion.

Unlike policy evaluation, policy design prerequires a 'suitable' econometric model. 'Suitable' here means that it explains all the variables relevant to the chosen criterion and it includes all the relevant policy instruments amongst its determinants. Broadly speaking therefore this second component of the subject uses the output from the first.

It is not always the case, however, that policy design is entirely conditioned on an econometric model and its coefficient estimates. In a stochastic world there will be a high degree of uncertainty concerning the model's coefficient estimates and if this is explicitly recognised then moments of these estimates (typically their variance) may enter the criterion function in some way. There is a possibility therefore that an instrument will be set not only to achieve a desired effect on the targets but also to reveal information about the model's coefficients and reduce their variance. As an example consider a policy plan for future interest rates meant to influence the subjective mean and variance of the demand for money. An optimal strategy may be to make interest rates highly variable in the early part of the plan. The resulting information will yield a better estimate of the interest rate elasticity of demand for money and so achieve money demand targets with greater certainty in the latter part of the policy plan. This elaborate procedure involves expensive Kalman filter estimation and as a result is rarely adopted. For the most part, policy design is carried out conditional on an econometric model and its parameter estimates.

It is clear from this overview that the subject is severely constrained by the amount of relevant historical experience. This is most acute when the policy under consideration is entirely new.

Such a situation was encountered when exchange controls were lifted in the U.K. in 1979. The Thatcher government responsible for the move had only a priori, largely untested economic theory to advise them as to the consequences of their actions. Whilst the development of a priori theory is very worthwhile in this respect we consider it to be outside the scope of the subject and, therefore of this thesis. We are forced to acknowledge that the constraint of data bounds the usefulness of our analysis.

To sum up this overview a number of key stages in the policy process have been defined. The first is the development of an econometric model. Estimation and hypothesis testing are the prime methodological components of this. The estimation of policy multipliers follows and for most models their size requires a numerical method. This gives us another methodological component, namely simulation. Moving further along the policy process we enter the policy optimisation stage. The numerical and analytical tools used here (including simulation) come under the methodological umbrella of optimal control theory.

This compartmentalisation of the subject of policy analysis may not be unanimously accepted. Nonetheless we feel that it identifies the key methodological elements of policy analysis and these elements form well defined topics with which to deal in this thesis.

1.2 The scope of the thesis

Tinbergen's pioneering work for the League of Nations in the 1930's was the foundation stone upon which the subject of policy analysis was built. Since that work no theoretical development has had a more profound effect on the subject than that of the rational expectations hypothesis. Directly, the hypothesis raised a myriad of methodological problems. All stages in the policy process from estimation through to policy simulations had to be (and are still being) rethought and reworked. Indirectly, the hypothesis threatened to undermine the usefulness of the subject itself. During the 1970's a number of rational expectations models emanated from the New Classical School of Macroeconomics. Members of this school, notably Lucas, Sargent and Wallace advanced theoretical economic models containing rational expectations in which the mean value of real variables was wholly independent of policy. It was suggested therefore that policy makers should abandon traditional policy tools and adopt simple, predictable policy rules. The view that the existence of rational expectations in an economic structure made policy impotent became increasingly popular. Whilst there is by no means a concensus on this view, there does seem to be a concensus on the adoption of the hypothesis itself. The widespread use of the hypothesis makes the reformulation of policy analysis in the light of rational expectations an urgent issue and it is this issue to which this thesis is devoted.

Chapter two casts a critical eye over the methodology that was orthodox before the rational expectations revolution. Because this methodological apparatus came under challenge with the advent of the hypothesis chapter three is devoted to a critical analysis of this challenge. Citing the important contributions of Buiter(1979),Pagan (1982) and others, two fundamental propositions (or rather, counter propositions) are established. The first states that closed loop rules, that is rules that use or feedback off current stochastic information in the economy (are superior to) open loop, fixed rules. The second asserts that the existence of rational expectations itself is neither a necessary nor sufficient condition for the failure of policy to exist. It is concluded that the impotence of policy is a direct property of the model and its construction rather than of the R.E. hypothesis.

Having established this the rest of the thesis devotes itself to the reworking of methodology that we alluded to above. Estimation, testing and simulation of economic models containing rational expectations are dealt with in these chapters because these are the fundamental stages of policy analysis. Focus here falls firmly on the problems raised by expectations of future variables since this is the area where work appears most scant. As a corollary of this it is also the area where most of the original work in this thesis lies.

Chapter four focuses on the problem of multiple solutions in models containing forward rational expectations. The use of terminal conditions is standard practice in the face of this problem and the validity of this practice is examined. An alternative procedure is advanced which involves estimating the solution jointly with the parameters of the model. This problem of estimation is taken up further in chapter six where first order conditions for maximising the likelihood function of a set of endogenous variables from a general linear model containing forward rational expectations are derived. The extensive use of analysis as opposed to numerical method sets this work aside from that of Fair and Taylor who have proposed an estimation method of their own. In particular we believe our method to be more computationally efficient as a natural consequence of this greater input of analysis. The first order conditions for a special class of rational expectations models are programmed into the routine "CLARE". CLARE is used to evaluate and compare, via a small and simple Monte Carlo study, full information estimation with a popular limited information method due to McCallum. The results sound a cautionary warning to would be users of the latter method although as with any Monte Carlo study it is difficult to assess how general these results are.

Chapter five revives the suggestion advanced by Anderson and Fair that standard models can be simulated under the hypothesis of rational expectations without having to re-estimate the model's parameters. An enticing possibility such as this deserves, we believe, closer inspection than that granted it by the original authors. Because the method uses 'old' parameter estimates policy multipliers are inconsistent and it is the seriousness of this asymptotic bias that we analyse. We conclude that the method may in many circumstances yield relatively accurate results (relatively small multiplier biases).

No thesis of a practical bent would be complete without some resort to data to support either the results or the basis of the thesis. Considering the profound nature of the hypothesis it is astonishing that its widespread acceptance has been accompanied by so little formal testing of its validity. In chapter seven then, a test of the hypothesis is carried out in the context of a simple model of financial asset demands. As well as verifying the hypothesis the chapter emphasises how full information estimation and testing must go hand in hand as the first stage in policy analysis when the hypothesis is invoked.

1.3 The framework of the analysis

It should be apparent from this by now that our framework is the standard linear dynamic model. Passing reference only is made to nonlinear models. The arguments for and against confining attention to linear models should be well known. We must stress however, that most of the large macro-models in the UK are nonlinear. Exactly how far our results carry over to nonlinear models is not clear but more and more modelling schools are maintaining condensed forms of their models as well as the models themselves. These are generally linearised minature versions of the larger model and are used to explore model properties and model policy responses in less detail but greater depth. Our methods and analysis can be directly applied to these condensed forms. Other areas of application can be found in small empirical models. These are usually developed to examine either a particular hypothesis or to focus on a particular aspect of policy. They arise from partial equilibrium theoretical analysis and are often just descriptions of a particular sector of the economy. One familiar example is the wage-price models of the labour market employed by for example, Lipsey and Parkin (1970), and Desai (1975). In any event we should not dwell too long on justifying what is the adopted framework of 90% of the profession.

Nonstationarity receives the same cursory treatment. We are on more solid ground here. Whilst it is clear that the world is nonlinear it is by no means clear that it is nonstationary. Again only a small segment of the profession troubles itself with the implications of nonstationarity.

For most of the thesis it is assumed that expectations at period 't' are formed in period 't-1' rather than in period 't'. This is purely our own preference as we view a lagged information base as a more palatable and realistic assumption. Fortunately the analysis is equally valid for either assumption although the exact form of the relevant algebraic expressions will of course differ.

1.4 Notation and conventions adopted for the thesis

Notation is not always uniform between chapters. Where special reference is made to existing literature we have adopted within limits the author's conventions so that the reader may compare our analysis with that of the original work more easily. Therefore, to maintain absolute clarity the notation adopted in each chapter is explained within the chapter itself. Certain conventions have been adhered to throughout however. In particular the rational expectation of y_t based on an information set dated at time $t-1$ (denoted Ω_{t-1}) is written in longhand as $E(y_t/\Omega_{t-1})$ and (where this longhand becomes too cumbersome) in shorthand as $y_t^e|_{t-1}$. Symbols denoting vectors are broadly in keeping with the literature. For example $\underset{\sim}{y}_t$, $\underset{\sim}{x}_t$ or $\underset{\sim}{z}_t$ and $\underset{\sim}{u}_t$ or $\underset{\sim}{v}_t$ denote vectors of endogenous, exogenous (or policy and predetermined) and white noise variables respectively and matrices are distinguished from scalars by using upper case symbols.

Multivariate R.E. models have three basic representations. The first form is where the basic economic relationships of the model (supply and demand curves, etc.) are simply stacked into a simultaneous equations system. This is the familiar structural form. Solving this structural form completely so that each endogenous variable is written purely in terms of observable exogenous and predetermined variables gives us our second representation, the reduced form. The third form the model can take lies between these two. If the structural form is partially solved by premultiplying by the inverse of the parameter matrix on the endogenous variables then each of the variables will be represented purely in terms of exogenous, predetermined and expectations variables. We have termed this the 'quasi reduced form' because unlike the reduced form the structure has only been partially solved. Other terminology used is either in general accordance with the literature or is explained where used.

2. POLICY ANALYSIS BEFORE THE RATIONAL EXPECTATIONS REVOLUTION

2.1 A review of the policy design literature

We open this chapter with a brief overview of the policy design literature prior to the advent of R.E. in macroeconomics.

A formal framework for policy design and evaluation was laid down by Tinbergen as long ago as 1952. In his pioneering work for the League of Nations he considered the possibility of influencing the course of a static economy of the form [1]

$$y = Rx + s \tag{1.1}$$

where y is a gx1 vector of endogenous variables, x is a kx1 vector of policy instruments, s is a gx1 vector of exogenous influences not subject to control and R is a gxk matrix of parameters which in Tinbergen's formulation were known.

The policy maker's objective function was assumed to be

$$Q = \Sigma(y_t-y^*)'(y_t-y^*) \tag{1.2}$$

where '*' denotes a targetted or desired value and a lower value of Q reflects a more successful policy.

A few notes on the form chosen for Q are pertinent. (1.2) does not allow us to weight each targetted variable differently. For example if two of the elements of y are unemployment and the trade deficit then these would have equal impacts on welfare. Further, positive and negative deviations from targetted values are treated symmetrically and this is clearly not satisfactory. It is probable that (1.2) was chosen by Tinbergen because of the emphasis of his study on stabilisation policies. In minimising (1.2) we may be attempting for example to minimise the

---

[1] The original analysis was in terms of scalars. We have generalised it here to vectors.

variations of output around its targetted level, very relevant in Tinbergen's time where apart from the great depression economic experience had been one of regular trade cycles. (1.2) also ignores society's (the decision maker's) intertemporal tastes as reflected by a discount rate. The form of welfare function is taken up in our discussion of dynamic economies where we generalise (1.2) in an attempt to eliminate some of the deficiences.

The problem then was to minimise

$$Q = y'y + y^{*\prime}y^* - 2y^{*\prime}y$$

subject to $y = Rx+s$

Substituting (1.1) into (1.2) gives us the unconstrained maximisation problem

$$\min Q = x'R'Rx + 2x'R's + s's + y^{*\prime}y^* - 2y^{*\prime}Rx - 2y^{*\prime}s \tag{1.3}$$

for which the K first order conditions are

$$dQ/dx = 2R'Rx + 2R's - 2R'y^* = 0 \tag{1.4}$$

We may distinguish three cases

Case (i); $K = g$

This is the case where the number of instruments equals the number of targets. In this case

$$\text{(a) } \hat{x} = R^{-1}(y^*-s) \tag{1.5}$$

$$\text{and (b) } \hat{Q} = 0 \tag{1.6}$$

where a '^' denotes an optimal value. We see then that in this case all targets are met exactly.

Case (ii) $K < g$

In this case $R'R$ is of full rank so that

$$\text{(a) } \hat{x} = (R'R)^{-1}R'(y^*-s) \tag{1.7}$$

$$\text{and (b) } \hat{Q} = (y^*-s)'M(y^*-s)$$

$$\text{where } M = (I-R(R'R)^{-1}R')$$

Note that $\hat{x}$ is simply an OLS estimate in the regression of (y^*-s) (the desired value of $R\hat{x}$) on the matrix R and so its interpretation is clear.

Case (iii) $K > g$

Here there are an excess of intruments over targets and obviously K-g of the targets cannot be set using the first order conditions (1.4) and have to be set arbitrarily. The remaining g instruments take on values in a similar way to those of case (i)

$$\hat{x}_g = R_g^{-1}(y^*-s) \tag{1.8}$$

where x_g is the g subvector of x containing the chosen elements and R_g contains the appropriate g columns of R. Note that it doesn't matter which g-k subset of x is set arbitrarily as (1.2) does not reflect instrument costs.

Subsequently Tinbergen's work was generalised chiefly by Theil (1968) and Brainard (1967) to incorporate various forms of uncertainty.

Theil considered the case of fixed R but random s. If instead of minimising welfare loss (Q) we now minimise its expectation

$$E(Q) = x'R'Rx+2x'R'E(s)+E(s's)+y^{*\prime}y^{*}-2y^{*\prime}Rx-2y^{*\prime}s \tag{1.9}$$

then it is clear that our previous results are unchanged providing that s is replaced by what Theil called its 'certainty equivalent' namely its expectation. A comparison of (1.9) with (1.3) confirms this.

Finally before we leave this very restrictive scenario of a static economy we note the case of Brainard where both R and s are allowed to be random and correlated. In particular joint normality is assumed. The uncertainty may either be in the structure of the economy or may reflect ignorance of the true values of R and s. Minimising expected loss in this case gives [2]

$$\hat{x} = (\Sigma_{RR}+\bar{R}'\bar{R})^{-1}\, \overset{E[R'(y^{*}-s)]}{\cancel{(R'(y^{*}-E(s))-P_{RS}(\Sigma_{RR}\Sigma_{SS}))}} \tag{1.10}$$

where $R = \{r_{ij}\}$, $\bar{R} = E(R)$, $\Sigma_{RR} = (\sigma^{R}_{ij}) = E(R-\bar{R})'(R-\bar{R})$ is a variance covariance matrix with elements $\sigma_{ii} = \sum_{k=1}^{K} \mathrm{var}(r_{ik})$ and $\sigma_{ij}=\sum_{k=1}^{K}\mathrm{cov}(r_{ik}r_{jk}), j\neq i$

In showing this result for scalar x, r, y and s Brainard noted the radical departure from the previous literature of the setting for the instruments. In practice, the coefficients (parameter estimates) of an econometric model are assumed to be certain when a control exercise is undertaken and the analysis in our simple case gives some indication of how complex policy optimisation may become when this assumption is dropped. As a further example consider the case where we explicitly recognise the fact that all the elements in (1.10) bar y* have to be estimated then the situation is very complex indeed. In this case the

---

[2] Note that quantities such as σ_{RR} may not be known and replacing them by their estimates may or may not provide a good approximation.

settings of the instruments actually affect the precision of the estimate of R. In turn the precision of the estimate of R is likely to have an important effect on $\hat{x}$ and so on expected loss (E(Q)) just as the variance covariance matrix Σ_{RR} (representative to some degree of the uncertainty in R) has in (1.10). Assessing the truly optimal policy in this case is a difficult task which we do not delve into here but touch on below in our discussion of dynamic economic systems.

Although Tinbergen's framework dates back to 1939 it was not until relatively recently that the more general problem of controlling stochastic dynamic economic systems was tackled. In his book Chow (1976) applied the then long established tools of engineering's optimal control theory to this very issue and it is to this that we now turn.

In the discussion above our economy was static and we were able to discuss policy unambiguously as a set of independent settings for $x_i (i=1,T)$. In our discussion of dynamic models however we have to distinguish two forms of policy. The first is open-loop policy where the time paths of policy variables are set at the beginning of the planning period without any regard to future events. This is the form that our policy took in the static world above. The second form is a specification of policy variables as functions of observations yet to be made. This function is called a feedback control rule and the policy a feedback policy. ('Feedback' indicates that results of current policy will in turn determine future policy). Note that there is obviously no role for feedback policy in static econometric models.

Consider now the general dynamic economic model considered by Chow

$$y_t = A_1 y_{t-1} + \dots A_m y_{t-m} + C_0 x_t + \dots . C_n x_{t-n} + b_t + U_t \qquad (1.11)$$

where y_t and x_t are as above, the A_i are gxg and the C_i gxk matrices respectively which are assumed known. Again b_t represents a vector of uncontrollable exogenous influences the form of which is unimportant to the analysis that follows. Finally u_t is a vector of disturbances representing structural innovations and to a limited extent capturing misspecification of the system.

Once again we consider minimising a quadratic welfare function subject to (1.11). Many other forms of welfare function could have been specified as 'better' approximations to that of the decision maker. We cannot guarantee in more general cases however that an analytically tractible control rule will emerge from the control process whereas as we see below our quadratic function gives us analytical and computational simplicitly. Further, if the disturbances and exogenous influences are normally distributed random variables then providing that x_t is subject to open loop or linear feedback control y_t will also be normal. Thus our quadratic welfare function contains all the relevant moments of the distribution of y, namely means, variances and covariances.

Before we proceed any further we reduce the order of (1.11) to one by redefining the vector y_t. In particular we may rewrite (1.11) in companion form as

$$
\left[\begin{array}{c} y_t \\ \hline y_{t-1} \\ \vdots \\ y_{t-m+1} \\ \hline x_t \\ \hline x_{t-1} \\ \vdots \\ x_{t-n+1} \end{array}\right]
=
\left[\begin{array}{c:c:c:c}
A_1 A_2 \;\cdots\; A_{m-1} & A_m & C_1 \;\cdots & C_n \\ \hline
 & 0 & & 0 \\
I & & 0 & \\
 & 0 & & 0 \\ \hline
0 \;\cdots\cdots\; 0 & 0 & & 0 \\ \hline
 & 0 & & 0 \\
0 & \vdots & I & \vdots \\
 & 0 & & 0
\end{array}\right]
\left[\begin{array}{c} y_{t-1} \\ \\ \\ y_{t-m} \\ \hline x_{t-1} \\ \vdots \\ x_{t-n} \end{array}\right]
+
\left[\begin{array}{c} C_0 \\ \hline 0 \\ \vdots \\ \hline I \\ \hline 0 \\ \vdots \\ 0 \end{array}\right] x_t
+
\left[\begin{array}{c} b_t \\ \hline 0 \\ \vdots \\ 0 \end{array}\right]
+
\left[\begin{array}{c} u_t \\ \hline 0 \\ \vdots \\ 0 \end{array}\right]
\qquad (1.12)'
$$

or as

$$y_t = Ay_{t-1} + Cx_t + b_t + u_t \qquad (1.12)$$

where the vectors are redefined according to (1.11). Because we have redefined our system in terms of a new y-vector we have to redefine our welfare function accordingly. In particular we write (1.2) as

$$Q = \sum_{t=0}^{T}(y_t - y_t^*)'K(y_t - y_t^*) \qquad (1.13)$$

where K may be specified appropriately with elements of unity and zero to give (1.2) as a special case. We go further however and specify K generally as a full matrix to allow for instrument costs (y_{t+1} as redefined contains the vector x_t) and to allow intertemporal covariances of the y's to affect welfare (these are no longer zero in our dynamic model

Our problem is finally specified; minimise (1.13) or rather its expectation subject to (1.12) and this we do using Bellman's method of dynamic programming.

If we look at the time structure of our economy we see that y_T is independent of future events (in periods T+i). Dynamic programming exploits this causal or recursive time structure found in all economic and physical systems by optimising backwards in time conditional on the past. Consider the problem of the decision maker in the final period T. It is

$$\min \quad Q_T = E(y_T - y_T^*)'K(y_T - y_T^*) \text{ w.r.t.} x_T \qquad (1.14)$$
$$\text{st. } y_T = Ay_{T-1} + Cx_T + b_T + u_T$$

This is obviously the case since in period T x_{T-i} (i>0) will already have been set and so optimisation proceeds conditional on (without regard to) their values.

Denoting an optimal value by '^' (1.14) gives a solution for x_T in terms of the past, i.e.,

$$\hat{x}_T = \hat{x}_T(y_{T-1}, x_{T-1}, b_{T-1}) \qquad (1.15)$$

and corresponding to this will be a minimum cost in T

$$\hat{Q}_T = \hat{Q}_T(\hat{x}_T, y_{t-1}, b_T) \qquad (1.16)$$

The problem in T-1 is now

$$\min Q_{T-1} = E(y_{T-1} - y_{T-1}^*)'K(y_{T-1} - y_{T-1}^*) + \hat{Q}_T(\hat{x}_T, y_{T-1}, b_T) \qquad (1.17)$$
$$\text{s.t.} \quad y_{T-1} = Ay_{T-2} + Cx_{T-1} + b_{T-1} + u_{T-1}$$

We may substitute out for $\hat{x}_T$ using (1.15) since we know that this is its truly optimal value in terms of y_{T-1}, x_{T-1} and b_{T-1} from (1.14) so that once again our problem is to minimise a function (conditional on) given past decisions. We can ignore x_{T-1}'s influence on $\hat{x}_T$ and in turn on Q_T as this has already been taken account of by the substitution for the optimal value of $x_T(\hat{x}_T)$ in terms of y_{T-1}, x_{T-1} and b_{T-1} according to (1.15). This procedure continues until the initial period is reached by which time a complete stream of policies $\hat{x}_1 \ldots \hat{x}_T$ will have been determined in terms of lagged values of y, b, and initial $x(x_0)$. The method will become clearer as we apply it explicitly to our problem.

Consider the first stage of our problem in (1.14).

Expanding gives [3]

$$\text{(a)}\quad Q_T = E_{T-1}(y_T'H_Ty_T-2y_T'h_T+c_T) \qquad (1.18)$$
$$\text{where (b) } H_T=K,\ h_T=Ky_T^* \text{ and } c_T=y_T^{*\prime}Ky_T^*$$

Substituting into (1.18) the constraint and differentiating with respect to x_T gives us the optimal instrument setting as

$$\text{(a)}\quad \hat{x}_T = G_Ty_{T-1} + g_T$$
$$\text{where}\quad \text{(b) } G_T = -(C'H_TC)^{-1}C'H_TA \qquad (1.19)$$
$$\text{and}\quad \text{(c) } g_T = -(C'H_TC)^{-1}(C^1H_Tb_T-C'h_T)$$

Substituting (1.19) into (1.18) gives $\hat{Q}_T$ as

$$\hat{Q}_T = y_{T-1}'(A+CG_T)'H_T(A+CG_T)y_{T-1}+2y_{T-1}'(A+CG_T)'(H_Tb_T-h_T)+(b_T+Cg_T)'$$
$$H_T(b_T+Cg_T)-2(b_T+Cg_T)'h_T+c_T+E_{T-1}(u_T'H_Tu_T) \qquad (1.20)$$

As we would expect this unwieldy expression is independent of any events in time T and we may now move on to the second stage of our procedure as in (1.17). Using (1.20) the problem is now

$$\text{(a)}\quad \min Q_{T-1} = E_{T-2}(y_{T-1}'H_{T-1}y_{T-1}-2y_{T-1}'h_{T-1}+c_{T-1})$$
$$\text{where (b) } H_{T-1} = K+(A+CG_T)'H_T(A+CG_T)\ ,$$
$$\text{(c) } h_{T-1} = Ky_{T-1}^*+(A+CG_T)'(h_T-H_Tb_T)\ , \qquad (1.21)$$
$$\text{and (d) } c_{T-1} = y_{T-1}^{*\prime}Ky_{T-1}+(b_T+Cg_T)'H_T(b_T+Cg_T)$$
$$-2(b_T+Cg_T)'h_T+E_{T-1}(u_T'H_Tu_T)$$

The important thing to note from this mass of algebra is that the problem in T-1 is of exactly the same form as in T. (Compare (1.18) with (1.20)). Because of this (1.21) (b) to (1.21) (d) form a set of difference equations in H, h and c with initial conditions (1.18) (b). These equations are recursive from (b) (determining H_t,Ψ_t) to (c) (determining the h_t given H_t,Ψ_t) to (d) (determining the c_t given H_t, h_t, Ψ_t).

We may ask at this stage under what circumstances does (1.21) (b) yield a steady state or long run value for H (and thus for G)? The steady state values must satisfy (using (1.21))

$$\text{(a)}\quad H = K + R'HR \Rightarrow H=\sum_{i=0}^{\infty} R'^iKR^i \quad \text{where } R=A+CG \qquad (1.22)$$
$$\text{and (b) } G = -(C-HC)C'HA$$

We see then that the steady state exists only if the summation on the right-hand side of (1.22) (a) converges. This is guaranteed if and only if

---

[3] The expressions use Chow's notation.

the eigenvalues of R are less than one.

Note, however, that under no circumstances regarding the system's parameters will there exist a corresponding steady state (long run) value for the $g_t \Psi_t$. This is because they depend on the time varying target values y_t^* and exogenous processes b_t.

The important result to emerge from this section may be summarised. If a quadratic objective function such as (1.13) is accepted as a good approximation to the decision maker's preference function then the optimal policy is of the feedback form (1.19). These feedback equations will have constant slopes if the planning period is long enough providing that the parameters satisfy the eigen value condition above. The intercept terms, however, do not have long run constant values. Their role is to react to accomodate changes in the exogenous influences b_t and target values y_t^*.

In an entirely analagous manner we may derive the optimal open loop policy, that is the policy that is 'best' given information only at the start of the planning horizon.

Again the problem is to

$$\text{(a)} \quad \min Q_0 = \sum_{t=0}^{T} E_0(y_t - y_t^*)'K(y_t - y_t^*) \tag{1.23}$$

$$\text{s.t.} \quad \text{(b)} \ y_t = Ay_{t-1} + Cx_t + b_t + u_t$$

It is more useful to write (1.23) (b) as

$$\text{(c)} \quad y_t = A^t y_0 + \sum_{i=0}^{t-1} A^i u_{T-i} + \sum_{i=0}^{t-1} A^i b_{t-i} + \sum_{i=0}^{\infty} A^i C^{i+1} x_{t-i}$$

and to substitute (c) into (a) and apply our dynamic programming technique.

Because the open loop rule is based on less information than the feedback rule it always does worse than (is dominated by) the optimal feedback rule. Basically the latter allow the x's to respond to shocks and events occuring up until the previous period whereas the former rule does not.

Before we close this section we note a few extensions which are easily accommodated within our framework.

Firstly we may allow the system to have time varying parameters by simply subscripting these parameters with appropriate time indices in (1.19) and (1.21). The difference equations for G, H etc., have the same form but obviously no steady state exists for the G_t, Ψ_t.

Secondly allowance may be made for unknown parameters that have to be estimated.

(1.19) (a) → (c) become

$$\begin{aligned}
&(a) \quad \hat{x}_T = G_T y_{T-1} + g_T \\
&(b) \quad G_T = -(E_{T-1}G'H_T G)^{-1}(E_{T-1}G'H_T A_T) \qquad (1.24) \\
\text{and } &(c) \quad g_T = -(E_{T-1}C_T'H_T C_T)^{-1}[(E_{T-1}C_T'H_T b_T) - (E_{T-1}C_T')h_T]
\end{aligned}$$

all notation as above. On the face of it (1.24) (a) is again a linear feedback rule in y_{T-1}. This is not however the case because in general the expected values of the coefficients will be functions of y_{T-i}, x_{T-i} and these functions themselves will be very hard to derive to provide optimal estimates. If however we are prepared to accept that the expected values of the coefficient matrices in all periods $t=0,\ldots,T$ may be approximated by their estimated values at the beginning of the planning horizon then the situation is essentially that of fixed coefficients and treated as above. In practice it is unlikely that the information revealed during the planning horizon if incorporated into the estimates each time period would have a significant impact on welfare. For one thing this information is often small relative to that accrued up to and including the start of the planning period.

Finally another form of uncertainty may be incorporated into our analysis.

It is often the case that certain elements of the y_t(target) vector are observed with an error through another variable say S_t. It has been argued for example that using money supply figures as a measure of liquidity in the economy is subject to error. A more obvious example perhaps is the permanent level of income or consumption which is observed only through actual levels. We have then observations on S_t where

$$S_t = M_t \, y_t + \eta_t \qquad (1.25)$$

where M_t is assumed a known matrix.

Providing that η_t is normally distributed and independent of the variables in the model then the optimal rule for t based on information at t-1 is

$$\hat{x}_t = G_t E_{t-1} y_{t-1} + g_t$$

where G_t and g_t are derived in exactly the same way as above so that they obey (1.21) (b) → (d). (The proof of this result is given by Chow (1976) ch. 8). The problem then separates naturally into two parts. Firstly calculate $E_{t-1}y_{t-1}$ by some means and secondly calculate G_t and g_t (as above). The first part is accomplished via use of the Kalman Filter (Kalman (1960)). Essentially this is an updating formula for $E(y_t|s_t)(=E_t y_t)$ of the form

$$E(y_t|s_t) = E(y_t|s_{t-1}) + D_t[s_t - E(s_t|s_{t-1})] \tag{1.26}$$

where D_t is a matrix whose ith row contains the regression coefficients of the ith element of y_t on the elements of s_t.

D_t by definition will be a function of the conditional expectation of the variance covariance matrix of s_t and so (via (1.25)) of that of y_t. An updating formula for the latter and for $E(s_t|s_{t-1})$ may be derived using (1.25) and the linear model for y_t in (1.12). Details are given in Chow (1976) pp. 186-191.

This completes our discussion of policy design prior to the R.E. revolution.

2.2 A review of the policy evaluation literature

In this section we deal with early attempts to estimate the response of endogenous variables (targets) to policy variables. These policy variables may be discrete (of the policy-on, policy-off type) or continuous in nature and we begin with the former.

The early policy literature had focused on discrete policy and in particular on assessing the effects on a wage-price structure of prices and incomes policies (see Bodkin (1966) for a survey of this extensive literature). In these studies time series data was split into two periods according to whether the policy was thought to be on or not and wage-price equations for both periods were compared. Possibly the most noteworthy of these policy-on policy-off studies was that of Lipsey and Parkin (1970), henceforth referred to as L-P.

Previous to L-P's study focus of the two period policy-on policy-off comparison had been on the significance of an intercept dummy added for the policy-on subsection of the data. The main force of their paper however was to show that intercept dummies may not be sufficient to capture fully the effects of the policy on wage-price relationships. In their analysis therefore they allow for policy effects on slope as well as intercept terms. Rather than explicitly add dummies however, the wage and price equations were estimated separately for the two periods and a test for coefficient stability (Chow or F test) was performed.

Now it is simple but tedious to show that in the model

$$y_t = X_t B_1 + u_t \qquad t=0,\ldots r-1$$
$$y_t = X_t B_2 + u_t \qquad t=r+1,\ldots T$$

the regression

$$\begin{matrix} y_{t1} \\ y_{t2} \end{matrix} = \begin{bmatrix} X_{t1} & X_{t2} \\ 0 & X_{t2} \end{bmatrix} \begin{bmatrix} B \\ \delta \end{bmatrix} + \begin{bmatrix} u_{t1} \\ u_{t2} \end{bmatrix}$$

gives values for $\hat{B}$ and $\hat{\delta}$ as $\hat{B}_1$ and $(\hat{B}_2-\hat{B}_1)$ respectively (hats denoting OLS estimates). It follows then that an F (Chow)-test based on $(\hat{B}_1-\hat{B}_2)$ from the separate regressions for $t=0,\ldots r$ and $t=r,\ldots T$ is identical to the (F-) test of joint significance of slope and intercept dummies in the complete regression for $t=0,\ldots T$. The wage-price model used there and in the preceeding literature was

$$\begin{aligned} &(a) \quad \dot{p}_t = a_1+a_2\dot{w}_t+a_3\dot{m}_{t-1}+a_4\dot{q}_t+u_{1t} \\ &(b) \quad \dot{w}_t = a_5+a_6U_t+a_7\dot{p}_t+a_8N_t+u_{2t} \end{aligned} \qquad (2.1)$$

where p, w, m, q, U and N are the price level, the nominal wage rate, the price of imports, a labour productivity variable, the unemployment rate and union aggressiveness variables respectively. Superscripted dots denote rates of change with all variables in logs. The slope coefficients for the policy on period were found to be significantly lower than for the policy-off period with the intercepts dramatically higher. L-P's broad conclusion then was that (focusing on the wage equation) although the incomes policy effectively 'broke' the wage-price relationship (the Phillips curve) the constant term therein was substantially augmented.

This it was argued represented the establishing of a norm leading to wage settlements well above that which would come about in the absence of policy at least for a wide range of values of the exogenous variables.

In his comment on the study Wallis (1972) pointed to a number of noteworthy deficiencies in L-P's work. The innapropriate use of O.L.S. on (2.1) for example. Of special interest to us here however is Wallis' closing suggestion that "the decision to implement the policy may not have been independent of the current values of endogenous variables themselves". Indeed if this were so L-P's Chow test is invalid. To see this consider a simplified version of (2.1) with policy dummies.

$$\begin{aligned} &(a) \quad \dot{p}_t = a_1\dot{w}_t + a_1'\delta_t\dot{w}_t + a_2\dot{m}_t + a_2'\delta_t\dot{w}_t + u_{1t} \\ &(b) \quad \dot{w}_t = a_3\dot{p}_t + a_3'\delta_t\dot{p}_t + a_4u_t + a_4'\delta_t u_t + u_{2t} \end{aligned} \tag{2.2}$$

where $\delta_t=0$ if $t \in T_0$ (the policy-off period)
and $\delta_t=1$ if $t \in T_1$ (the policy-on period).
Considering an identical procedure to L-P's Chow test we say that the policy has an impact if a_1', a_2', a_3' and a_4' are significantly different from zero. Now one interpretation of Wallis' policy regime is

$$\begin{aligned} &\delta_t=0 \text{ if } \dot{p}_t<\dot{p}^* \\ &\delta_t=1 \text{ if } \dot{p}_t > \dot{p}^* \end{aligned} \tag{2.3}$$

where $\dot{p}^*$ is the critical rate of inflation above which an incomes policy is enforced.

The implication of (2.3) is that even if appropriate instruments are used in (2.2) for $\dot{p}_t$ and $\dot{w}_t$ (2SLS instruments for example) δ_t is correlated with the error term violating the standard assumptions upon which most estimators are based. Estimates of all the parameters in this case are subject to simultaneous equations bias. Whilst an attempt is made in section 1.4 below to evaluate this bias and suggest an alternative procedure the simultaneous determination of the policy regime with the endogenous variables thwarts any such policy-on policy-off tests.

We now turn to the literature concerning continuous policy instruments. In the case where we have a structure containing continuous policy variables such as the level of government spending or money supply then two ways of estimating policy multipliers are discernable from the early literature. One is by means of solving a structural form to make the endogenous variables appear naturally on the left-hand side of each equation and then estimate the multipliers directly as the relevant O.L.S. coefficients. This 'reduced-form' method was proposed by Anderson and Jordan (henceforth A-J). Referring to equation (1.1) in the previous section this amounts to regressing the variables in y on the vector x providing multiplier estimates $\{r_{ij}\}$. These estimates will be consistent providing that the menu of left-hand side variables is complete.

The second method is to estimate the structural coefficient using say 2SLS and then to solve for the $\{r_{ij}\}$ either through a policy simulation or as the reduced form coefficients.

This then was the state of the art at the beginning of the 1970's. A critique of the (then) current methodology by Goldfield and Blinder (1972) (henceforth G-B) formed a milestone in policy evaluation history. Following Wallis' note on the L-P study an extensive analysis of the problem of simultaneous equations bias that arises when Federal or Government behavioural rules are ignored during structural or reduced form estimation is highlighted and forms the focus of their paper.

In the context of a simple example G-B derive the expected biases arising when a fiscal and monetary authority are estimating multipliers with varying degrees of astuteness using the A-J method. The model they use is (G-B's notation)

$$\text{(a)}\quad Y_t = K + \alpha F_t + BM_t + u_t \tag{2.4}$$

where Y_t is nominal income, F_t is government spending, M_t is money supply and K and u are a constant and error term respectively.

Reaction functions for F_t and M_t were specified so that their values deviate from a normal or equilibrium value (F_t^* and M_t^*) only when income is not at its targetted value (Y_t^*).

They are

$$\text{(b)}\quad F_t = F_t^* - F(Y_t - Y_t^*) + v_{1t} \qquad (2.4)$$

$$\text{(c)}\quad M_t = M_t^* - m(Y_t - Y_t^*) + v_{2t}$$

The degree of astuteness with which multipliers are estimated by the two authorities is measured as the ratio of estimated residual variance to the actual and so is for the fiscal and monetary authorities respectively

$$\gamma^2 = u^2_F/\sigma^2_u \quad \text{and} \quad \delta^2 = u^2_m/\sigma^2_u$$

where u^2_F, u^2_m and σ^2_u are the error variance of the fiscal authority, monetary authority and the actual model respectively.

The ratio of actual to estimated multipliers (the latter obtained by O.L.S. in (2.4)(a)) are for α and B respectively

$$\begin{aligned} &\text{(a) } R^{\alpha} = 1 + \delta(\rho\gamma - \delta)/\Delta \\ &\text{(b) } R^{\beta} = 1 + \gamma(\rho\delta - \gamma)/\Delta \end{aligned} \qquad (2.5)$$

where $\Delta = \gamma^2 + \delta^2 - 2\rho\gamma\delta + \gamma^2\delta^2(1-\rho^2) > 0$

and ρ is the correlation coefficient between u_F and u_m representing the degree to which the two authorities utilise the same information. Note that if

$$\rho < \delta/\gamma < \frac{1}{\rho}$$

a condition which is satisfied for negative ρ and probably satisfied for small positive ρ, then both multipliers are biased towards zero with the more serious bias attributing to the more astute authority! These findings were confirmed by Monte Carlo experiments which were undertaken using the Moroney-Mason (1971) model. For these experiments the standard version of the model was augmented by reaction functions like (2.4) (b) and (c) and used to generate data. Using this data they then proceeded to estimate multipliers in the A-J reduced form fashion and through structural simulation ignoring the reaction functions. In the former case multipliers were biased by a factor of between 10% and 120% for structural estimates however the implications of treating policy variables (F and M in (2.4) (a)) as exogenous when they are in fact endogenous according to (2.4) (b) and (c) are not so grave. Intuitively this is for two reasons. Firstly the policy variables concerned are likely to be few in number and will probably enter only a few equations in a large macro model. It is only these equations that suffer inconsistent estimates. Secondly the 'menu' of exogenous variables used as instruments in 2SLS is likely to be large in most cases and the mistaken addition of (say) two more (F and M) is not likely to be serious. When, however the structure is solved or simulated omitting policy rules such as (2.4) (b) and (c) then serious multiplier biases could be expected.

To make this clear consider G-B's simple model in (2.4) but exclude (2.4) (c) so that M is truly exogenous whilst F is subject to endogenous feedback as in (b). If the Fed was estimating responses to money supply whilst ignoring the reactions of the fiscal authorities (the government) then (2.4) (a) provides them with an estimate of B ($\hat{B}$) and this would be interpreted as the multiplier. In fact the true model is (2.4) (a) and (2.4) (b) so that

$$\text{(a)} \quad Y_t = K_t' + B'M_t + u_t' \tag{2.6}$$

where $K' = (K + \alpha F_t{}^* + \alpha FY_t{}^*) \,/\, (1 - \alpha F),$

$$u_t' = \alpha v_t + u_t$$

with policy multiplier (2.6) (b) $B' = B \,/\, (1 - \alpha F) = \delta Y_t / \delta M_t$

(2.6) (b) shows that even if all the structural estimates were consistent the Fed's estimate of the monetary multiplier ($\hat{B}$) is less than its true value B'.

The Monte-Carlo results for the structural estimation and simulation method showed multipliers to be overestimated by a factor of 2 contrasting strongly with the downward biases from the A-J method.

Finally we note that the class of reaction function of Chow (1976) or Buiter (1980) are excluded from the G-B critique as far as estimation goes. Clearly feedback rules that use predetermined variables do not cause correlation between the current error term and any regressors. Again however ignoring the reaction functions at the simulation stage has serious implications for multiplier estimates.

We conclude then as G-B do that

"there seems to be no substitute for specifying reaction functions and estimating (and simulating) the complete structure".

2.3 The policy experiments of Fair.

In this section we discuss some simulation techniques discussed and used by Fair (1975) in the context of his theoretical work "A Model of Macroeconomic activity". The study is an interesting one because policy responses derived using these techniques from a complex theoretical model are used to guide the specification of its empirical counterpart. As a result of this, policy multipliers from the empirical model are constrained by the theory and this is in stark contrast to the A-J methodology where policy multipliers are entirely based on the data. Before we discuss the simulation techniques we provide a very brief outline of the theoretical model.

The model's key features lie in its strong microeconomic foundations. The economy is divided into four sectors and the government. There is a bond dealer, a financial sector (banks), a corporate sector (firms) and a household sector. Every time period each sector solves a deterministic control problem subject to its binding quantity constraints and its available information (information set). The model is recursive with a flow of information and decisions running from a bond dealer to banks, from banks to firms and then on to households.

The fundamental difference between this model and its theoretical predecessors is the absence of forced equilibrium conditions which are usually achieved through a tâtonnement process. Indeed recontracting is not allowed anywhere in the model and this gives rise to its neat recursive structure. The model describes an economy in a state of fundamental disequilibrium there being no guarantee that 'desired' or targeted values of current decision variables are in fact met.

Expectations are non rational and are in fact very naive bearing in mind the model's overall level of sophistication. For example banks are assumed to expect the short rate not to change from its current value and this information is related to the bond dealer who then sets stock prices accordingly.

The sequence of decisions is fairly easy to follow. Each period a bond dealer sets current bond and bill rates according to his perceptions of current demands and supplies and communicates this information to firms and banks. Armed with this knowledge plus information from the government on tax rates and the reserve requirement ratio and having formed expectations about the level of current (and future) deposits available the banks set a loan rate, decide on the value of bills and bonds to buy and on the maximum amount of money that they will lend in the period.

In turn, firms receive information on the banks' decisions and on the current profit tax rate before they set their main decision variables. These are the volume of funds borrowed from banks, the maximum amount of labour hired from households, the amount of investment to undertake and the current level of output and the prices of goods.

Finally at the end of this decision sequence are households who decide how many hours to work subject of course to the upper bound set by firms.

As a final comment the model ignores distributional and search questions because there is perfect information within the structure described and aggregation over agents being coarse and over goods total.

It is not our purpose here to evaluate the model's microeconomic foundations or to establish the type of (dis)equilibrium described by the system. Instead we focus on the simulation methodology employed in an attempt to both 'condense' the model into a simple log linear form (henceforth this is referred to as the "condensed model") and to assess the affects of policy instruments on the dynamic path of key endogenous variables.

The simulation procedure involves six basic steps

Step 1:

Taking each sector separately write down the set of equations governing expectational behaviour the identities and flow of funds constraints facing the sector and its objective function.

Step 2:

Using an algorithm that searches over the parameter space find a set of 'parameter values' [4] such that when the sector maximises its objective function the sector's decision variables are in a steady state. [5]

Step 3:

Using this set of 'parameter values' (hence-forth 'steady-state

----

[4] 'Parameter values' here means parameters, initial conditions and exogenous variable values.

[5] The condition for the endogenous variables' time profile is that they are "broadly flat" rather than in an exact steady state.

parameters') change either one of the initial conditions or one of the exogenous variables and calculate the (proportional) response of key endogenous variables to this in subsequent periods.

Step 4:

Calculate the average response from two such experiments for each subsequent period. Substitute these average elasticities into a log linear dynamic model which explains the key current endogenous (decision) variables in terms of predetermined variables and variables determined outside the sector.

Step 5:

Simulate this condensed form to provide a 'base run'. This involves the setting of previously undetermined intercepts included in the condensed form in such a way that a steady state is once again achieved for the key endogenous variables.

Below we have set out a simple two sector two period analogy to Fair's model. Performing steps 1 to 5 on this will help to illuminate the method.

Our model is similar to one of temporary equilibrium where households are rationed on the labour they may sell so that there is involuntary unemployment and firms in turn are rationed on the goods supply side. Both firms and households then are off their unconstrained supply and demand functions (respectively) and so this is allied to the temporary equilibrium interpretation of a Keynsian system.

Households decide on consumption 'today' (C_1) versus consumption 'tomorrow' (C_2) given an initial endowment of money (m_0) and ration of labour (l_1 and l_2). This decision is encapsulated in their holdings of money in the current period (m_1). Note that money has no direct effect on utility.

Faced with their goods supply ration firms simply decide on the least cost combination of capital and labour at the current wage rate (W_i) and at the current rental price of capital services (r_i). To make the model truly recursive, it is not the actual goods supply ration that affects employment (c_i) but the perceived or expected ration (C_i^e). A simple lagged value proxies these expectations and so the current sequence of decisions is recursive with firms deciding on employment independently of the household's consumption decision.

The control problems of firms and households and the first order conditions which determine optimal values for $m_1(\hat{m}_1), l_1(\hat{l}_1)$ etc are laid out below in Table 2.1.

Having derived optimal settings for our key endogenous variables (step 1) we now emulate step 2. Taking each sector separately we find a set of 'parameter values' that yields a steady state solution for the decision variables. In the firm sector this implies

$$\hat{l}_1=\hat{l}_2(=L,\ \text{say})\ (\text{and}\ \hat{k}_1=\hat{k}_2=k) \tag{3.6}$$

and so (for example)

$$r_1c_0/Aw_1 = r_2c_1/Aw_2 \quad \text{or} \tag{3.7}$$

$$r_1/r_2 = (w_1/w_2)(c_1/c_0)$$

A steady state in the household implies

$$c_1=c_2=c_0=c \tag{3.8}$$

and so condition (3.7) reduces to

$$r_1 = (w_1/w_2)\ r_2 \tag{3.9}$$

Further, current money holdings must equal initial endowments so that

$$m_1=m_0(=m,\ \text{say}) \tag{3.10}$$

Using (3.8) and (3.10) with optimal household decisions gives

$$(w_1-w_2)l=m \tag{3.11}$$

Combining (3.6) - (3.11) we may write (for example)

$$\alpha_2 = \alpha_1(2w_2l + m)\ /\ m \tag{3.12}$$

Counting equations that our second stage delivers with those from the first order conditions in the optimisation stage we see that we have 12 equations (5 from the former and 7 from the latter) in 7 decision variables, 2 initial conditions, 4 exogenous variables and 6 parameters (19 'parameter values' and decision variables). Clearly there are an excess of 'parameter values' to solve for (5 equations in 12 'parameter values' once the decision variables have been solved for) and what values our unknowns take depend upon the ones for which solution values are sought or the way in which the equations are solved. We have provided two examples above where we have chosen to express r_1 in terms of w_1, w_2 and r_2 in (3.9) and α_2 in terms of α_1, w_2, l and m in (3.12).

Table 2.1

Households

$$\operatorname*{Max}_{(m_1)} \quad U = \alpha_1 c_1^2 + B_1 l_1^2 + \alpha_2 c_2^2 + B_2 l_2^2$$

s.t. $c_1 = m_0 - m_1 + w_1 l_1$

and $c_2 = w_2 l_2 + m_1$

with $l_1 = \bar{l}_1$ and $l_2 = \bar{l}_2$

$B_1, B_2 < 0$; $B_2 > B_1$; $\alpha_1, \alpha_2 > 0$; $\alpha_2 < \alpha_1$.

Firms:

$$\operatorname{Max}_{\{l_1, l_2\}} \Pi^e = c_1^e + \rho c_2^e - w_1 l_1 - w_2 l_2 - r_1 k_1 - r_2 k_2$$

s.t. $c_1 = A l_1 k_1 = \bar{c}_1$ and $c_2 = A l_2 k_2 = \bar{c}_2$ and $c_i^e = c_{i-1}$

with $0 < \rho < 1$, ρ being a discount rate.

First order conditions:

Households: $$\left.\frac{\delta U}{\delta m_1}\right|_{l_1, l_2} = 2\alpha_2(w_2\bar{l}_2 + m_1) - 2\alpha_1(m_0 - m_1 + w_1\bar{l}_1) = 0$$

$$\Rightarrow \quad \hat{m}_1 = (\alpha_1 m_0 + \alpha_1 w_1 \bar{l}_1 - \alpha_2 w_2 \bar{l}_2)/(\alpha_2 + \alpha_1) \tag{3.1}$$

$$\Rightarrow \quad \hat{c}_1 = m_0 - \hat{m}_1 + w_1 \bar{l}_1 \tag{3.2}$$

and $$\hat{c}_2 = w_2 \bar{l}_2 + \hat{m}_1 \tag{3.3}$$

Firms: $$\left.\frac{\delta \Pi^e}{\delta l_1}\right|_{c_1, c_2} = r_1 \bar{c}_0 l_1^{-2}/A - w_1 = 0$$

$$\left.\frac{\delta \Pi^e}{\delta l_2}\right|_{c_1 c_2} = r_2 \bar{c}_1 l_2^{-2}/A - w_2 = 0$$

$$\Rightarrow \quad \hat{l}_1 = (r_1 c_0 / A w_1)^{0.5} \tag{3.4}$$

and $$\hat{l}_2 = (r_2 \bar{c}_1 / A w_2)^{0.5} \tag{3.5}$$

Parameters: $\alpha_1, \alpha_2, B_1, B_2, \rho, A$.

Exogenous variables: w_1, w_2, r_1, r_2.

Initial conditions: c_0, m_0.

Decision variables: m_1, c_1, c_2, l_1, l_2 (and k_1, k_2).

$\hat{k}_1$ and $\hat{k}_2$ are residuals from the production function given $\hat{l}_1$ and $\hat{l}_2$.

Note: to simplify the problem and to maintain recursivity a nash type of equilibrium is described where agents assume no response from others to their actions.

Because our 'parameter values' are underdetermined we would expect that given start values for the former we would require very few iterations to achieve our steady state.

We do not however wish to labour the point that in our example the 'parameter values' are underdetermined by step 2. In fact in Fair's model 30 periods with 30 equations per decision variable have to solve for the unknowns in step 2 (in addition to the first order conditions). This number of periods may provide a just sufficient or over sufficient number of conditions to solve for the unknowns. If the latter is the case then only an approximate steady state can be derived with decision variables having only a 'broadly flat' not totally flat time profile in most cases.

The key point to note from this analysis is that whether over or underdetermined the values for the 'parameters' are entirely arbitrary. If for example we rearrange (3.12) we get

(a) $\alpha_2/\alpha_1 = 1 + 2w_2l/m$ (3.12)

Because positive start values for w_2 and m_0 will have been chosen (3.12) (a) states that α_2 will exceed α_1 so that referring back to the utility function we see that consumption 'tomorrow' is valued more highly than consumption today. This is obviously unacceptable.

Pressing on regardless to steps 3 and 4 we would estimate multipliers by perturbation of exogenous variables or initial conditions. In our model for example the multiplier response of the demand for money to its previous value (a dynamic multiplier) may be written as

$$M(m_1m_0) = \delta m_1/\delta m_0 = (2c\alpha_1 - \alpha_2w_2l)/(2c\alpha_1-\alpha_2w_2l+2c\alpha_2)$$

We see that the multipliers in our model are functions of parameters initial conditions and other variables in the model as indeed we would expect from a nonlinear model. Because all these elements are 'chosen' arbitrarily the multipliers are correspondingly arbitrary.

To sum up then we could say that multipliers delivered by the Fair routines may or may not be unique depending on whether there is a unique setting of the 'parameter values' that forces a steady state on the system. Either way there is no guarantee without inspection that the 'parameters' and multipliers so derived are sensible.

It is difficult to see why the criterion of steady state was chosen to derive multipliers to guide the specification of a macro model which is designed to explain short run and disequilibrium phenomena. Why not for example impose the criterion that outcomes from the simulation exhibit certain time series features such as those corresponding to observable data series? Alternatively, since it could be argued that to a large extent models merely formalise a modeller's a priori beliefs, why not choose 'parameters' that appeal to intuition?

This brings us to the final stage in the procedure. Given multiplier responses from steps 1 to 6 we must note how they are used to guide specification of the empirical counterpart.

The fundamental constraints imposed on the latter took the form of exclusion restrictions on the structure as is usual in economic modelling since exclusion restrictions allow parameters of interest to be identified from the data. Whilst most of the lag structures in the model were chosen for goodness of fit there are certain areas of the model where the data are a poor discriminant of dynamic structure. Such is the case of the response of various decision variables to tax rates;

"There is unfortunately much uncertainty regarding both the short run and the long run response of the economy to various tax law changes. The data do not appear to be very good at descriminating amongst different lag structures ..."

Thus certain equations contain lag polynomials in tax rates whose order was constrained by the predictions of the theoretical model. The simulations of the condensed form in step 5 guided all of those specification decisions.

We conclude this section by noting that simulation of complex theoretical models is undoubtedly the best way of exploring their responses. The problem of choosing numerical parameter values to input into this procedure is not trivial. Obviously the investigator's judgement plays an important part in this. However, whatever criterion is chosen should be explicitly stated and supported rather than left submerged in some complex numerical procedure.

2.4 Wallis' note on the Lipsey-Parkin study; an example

We noted in section 1.2 that policy evaluation studies of the policy on policy off nature may be subject to simultaneous equations bias if the decision to implement the policy is contemporaneously linked with an endogenous variable (G-B's critique). In this section then we conclude the paper by taking a closer look at the form of this bias in the context of a simplified version of L-P's wage-price model.

Omitting constants and two exogenous variables L-P's model reduces to

$$\dot{w}_t = \alpha_{11}\dot{p}_t + B_{11}U_t + v_{1t} \tag{4.1}$$

$$\dot{p}_t = \alpha_{21}\dot{w}_t + B_{21}\dot{m}_t + v_{2t} \tag{4.2}$$

all variables as above with all coefficients positive apart form B_{11}.

Taking the simplest hypothesis we assume that the effects of a prices and incomes policy falls on α_{11} and α_{21}.

Because we wish to focus on bias arising purely from endogenous policy we assume that appropriate instruments are used for $\dot{p}_t$ and $\dot{w}_t$ ($\hat{\dot{p}}_t$ and $\hat{\dot{w}}_t$ respectively) so that an L-P type exercise would amount to testing the significance of α_{12} and α_{22} in the regressions

$$\dot{w}_t = \alpha\hat{\dot{p}}_t + \alpha_{12}\delta_t\hat{\dot{p}}_t + B_{11}U_t + v_{1t} \tag{4.3}$$

$$\dot{p}_t = \alpha_{21}\hat{\dot{w}}_t + \alpha_{22}\delta_t\hat{\dot{w}}_t + B_{21}\dot{m}_t + v_{2t} \tag{4.4}$$

where δ_t is as above.

We would expect α_{12} and α_{22} to be negative as they represent a dampener on the impact of $\dot{p}$ on $\dot{w}$ through the influence of the incomes policy on trade union bargaining and of $\dot{w}$ on $\dot{p}$ through the effect of the prices policy on firms' price-cost margins.

How would we expect the significance of α_{12} and α_{22} to be affected if δ_t was subject to feedback of the form

$$\begin{aligned} &(a) \quad \delta_t=1 \text{ if } \dot{p}_t>\dot{p}^* \\ &(b) \quad \delta_t=0 \text{ if } \dot{p}_t\leq\dot{p}^* \end{aligned} \tag{4.5}$$

where $\dot{p}^*$ is a target level of inflation?

The two equations in (4.5) endogenise the variables $\delta_t\hat{\dot{p}}_t$ and $\delta_t\hat{\dot{w}}_t$. Asymptotic biases takes the standard form (when O.L.S. is applied to (4.3) and (4.4));

$$\text{(a)}\quad \begin{bmatrix} B_{\alpha_{11}} \\ B_{\alpha_{12}} \\ B_{\beta 11} \end{bmatrix} = \begin{bmatrix} \sigma_{\hat{\dot{p}}\hat{\dot{p}}} & \sigma_{\hat{\dot{p}}\delta\hat{\dot{p}}} & \sigma_{\hat{\dot{p}}u} \\ & \sigma_{\delta}{}^{2}\hat{\dot{p}}^{2} & \sigma_{\delta\hat{\dot{p}}u} \\ & & \sigma_{uu} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \sigma_{\delta\hat{\dot{p}}v_1} \\ 0 \end{bmatrix} \tag{4.6}$$

$$\text{(b)}\quad \begin{bmatrix} B_{\alpha_{21}} \\ B_{\alpha_{22}} \\ B_{\beta 21} \end{bmatrix} = \begin{bmatrix} \sigma_{\hat{\dot{w}}\hat{\dot{w}}} & \sigma_{\hat{\dot{w}}\delta\hat{\dot{w}}} & \sigma_{\hat{\dot{w}}m} \\ & \sigma_{\delta}{}^{2}\hat{\dot{w}}^{2} & \sigma_{\delta\hat{\dot{w}}m} \\ & & \sigma_{mm} \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ \sigma_{\delta\hat{\dot{w}}v_2} \\ 0 \end{bmatrix}$$

where B denotes bias in the relevant subscripted variables, σ_{xy} denotes $\lim_{T\to\infty} 1/T \sum_{i=1}^{T} x_i y_i$ and the R.H.S. matrices are symmetric. Focussing on $B\alpha_{12}$ and $B\alpha_{22}$ we see that

$$\text{(a)}\quad B_{\alpha_{12}} = (\sigma^{p}{}_{\delta^2\hat{\dot{p}}^2})\, \sigma_{\delta\hat{\dot{p}}v_1} \tag{4.7}$$

$$\text{and (b)}\quad B_{\alpha_{22}} = (\sigma^{w}{}_{\delta^2\hat{\dot{p}}^2})\, \sigma_{\delta\hat{\dot{m}}v_2}$$

where σ^{w} and σ^{p} are the 2,2 elements of the inverse of the relevant right hand side matrices. Because these terms are diagonal elements of the inverses of positive definite matrices they are positive. Both $\sigma_{\delta\hat{\dot{p}}v_1}$ and $\sigma_{\delta\hat{\dot{w}}v_2}$ are positive so that

$$B_{\alpha_{12}} > 0 \text{ and } B_{\alpha_{22}} > 0$$

This result suggests that if a prices and incomes policy successfully broke the wage-price relationship in (4.1) and (4.2) so that α_{12} and α_{22} are negative then endogenous policy of the form of (4.5) may lead to an inference of policy impotence.

A suggestion then is to use an instrument for δ. In particular we suggest

$$\delta_t = 1 \text{ if } \hat{\dot{p}}_t > p^*$$

$$\text{and } \delta_t = 0 \text{ if } \hat{\dot{p}}_t \leq p^*$$

where $\hat{\dot{p}}_t$ is again a valid instrument for $\dot{p}_t$.

P_t^* however is unobserved and unless it can be inferred from the data or from extraneous information such as the often invaluable accounts of historical policy, then clearly the method fails. We may assert then that knowledge of $\dot{p}^*$ is a necessary condition for identification of the true effects of policy on (4.1) and (4.2) as written in (4.3) and (4.4). This is not surprising since (4.3) and (4.4) is obviously an incomplete description of the generation of wages and prices.

Finally by an exactly analagous argument the observed impact of policy is reduced if intercept rather than slope dummies are used to represent the effects of policy. It is not however, possible to make such definitive statements about the nature of the biases if both slope and intercept dummies are included (the full L-P method) as the forms for biases are complex and uninformative.

3. POLICY ANALYSIS IN THE LIGHT OF R.E.

3.1 Lucas' critique

Possibly the most important and powerful critique of traditional policy analysis was made by Lucas (1976). In his influential paper he argues that unlike standard models R.E. models have structural parameters that vary directly with economic policy. His argument is easily seen in the context of a simple two equation model explaining a price level (p) and income (y);

$$\begin{aligned} &(a) \quad p_t = V + \gamma y_t + Bm_t + u_{1t} \\ &(b) \quad y_t = \alpha(p_t - p_t^e|_{t-1}) + y_o + u_{2t} . \end{aligned} \qquad (1.1)$$

y_o is a constant representing a natural level of income and m_t is the money stock. u_{1t}, u_{2t} are white noise errors and all variables are in logs.

For simplicity assume that the money stock is under the direct control of the authorities and that a simple open loop policy is in operation so that

$$m_t - m_{t-1} = m_o + \varepsilon_t \qquad (1.2)$$

m_o is obviously the rate of growth of the money stock and ε_t is white noise.

The quasi reduced form of (1.1) is

$$\begin{aligned} &(a) \quad P_t = \Pi_{11} + \Pi_{12}P_t{}^e|_{t-1} + \Pi_{13}m_t + v_{1t} \\ &(b) \quad y_t = \Pi_{21} + \Pi_{22}P_t{}^e|_{t-1} + \Pi_{23}m_t + v_{1t} \end{aligned} \qquad (1.3)$$

where $\Pi_{11} = (\gamma y_0+V)/(1-\alpha\gamma)$, $\Pi_{12} = -\alpha\gamma/(1-\alpha\gamma)$, $\Pi_{13} = B/(1-\alpha\gamma)$,

$\Pi_{21} = y_0+\alpha(\gamma y_0+V)/(1-\alpha\gamma)$, $\Pi_{22} = -\alpha/(1-\alpha\gamma)$, $\Pi_{23} = \alpha B/(1-\alpha\gamma)$,

$v_{1t} = (\gamma u_{2t}+u_{1t})/(1-\alpha\gamma)$, $v_{2t} = (\alpha u_{1t}+u_{2t})/(1-\alpha\gamma)$

The model contains a static version of the Lucas supply function (1.1)(b) where output deviates from its natural level (y_o) only when there are mistakes in anticipating the price level (ignoring the structural disturbance term). If $\beta=1$ and $\gamma=-1$ then (1.1)(a) is simply a quantity theory of money equation where the velocity of circulation, (V) is constant.

Solving (1.3) for p_t and y_t in terms of observables gives the reduced form

$$\text{(a)}\quad p_t = [\Pi_{12}(\Pi_{13}m_0+\Pi_{11})/(1-\Pi_{12})]+[\Pi_{12}\Pi_{13}/(1-\Pi_{12})]m_{t-1}+\Pi_{13}m_t+v_{1t}$$

$$(1.4)$$

$$\text{(b)}\quad y_t = [\Pi_{22}(\Pi_{13}m_0+\Pi_{11})/(1-\Pi_{12})]+[\Pi_{22}\Pi_{13}/(1-\Pi_{12})]m_{t-1}+\Pi_{23}m_t v_{2t}$$

Immediately we see that the reduced form parameters vary directly with monetary policy.

In chapter two we discussed traditional policy analysis and saw that a standard technique of the pre R.E. era was the reduced form multiplier approach of Anderson and Jordan. Adopting this approach we would regress y_t and p_t on m_t, m_{t-1} and m_{t-2} as our theory suggests that these are the relevant reduced form variables. We would have, in this case the following model

$$\text{(a)}\quad p_t = \hat{a} + \hat{b}m_t + \hat{c}m_{t-1}$$

$$(1.5)$$

$$\text{(b)}\quad y_t = \hat{d} + \hat{e}m_t + \hat{f}m_{t-1}$$

where a '^' denotes an OLS estimate.

If the rate of monetary growth was constant over the period of the regression sample and equal to (say) $\overline{m}_0$ then

$$\hat{a} \;\tilde{c}\; (\Pi_{12}\Pi_{13}\overline{m}_0+\Pi_{11})/(1-\Pi_{12})$$
$$\hat{b} \;\tilde{c}\; \Pi_{13}$$
$$\hat{c} \;\tilde{c}\; \Pi_{12}\Pi_{13}/(1-\Pi_{12})$$
$$\hat{d} \;\tilde{c}\; y_0 - \alpha\Pi_{13}\overline{m}_0(= y_0-\Pi_{23}\overline{m}_0)$$
$$\hat{e} \;\tilde{c}\; \alpha\Pi_{13}(= \Pi_{23})$$
$$\hat{f} \;\tilde{c}\; -\alpha\Pi_{13}(= -\Pi_{23})$$

$\tilde{c}$ denotes "is a consistent estimate of"

To evaluate the effects of an announced and permanent monetary contraction to (say) $\overline{m}_1$, we would simulate (1.5) substituting values for m_t

and m_{t-1} from the planned (tight) money trajectory. However at the instant the policy was implemented our estimated model (1.5) becomes invalid and redundant. The policy change of $\bar{m}_0$ to $\bar{m}_1$ will shift the intercept term and a structural break will be observed. The policymaker following this procedure will be continually revising monetary targets as his estimates of policy multipliers are continually falsified. This was the main force of Lucas' paper that became the 'Lucas critique'.

As a straightforward answer to this problem Wallis (1980) suggested that parameters associated with policy (and with exogenous variable processes) be separated from those associated with the economic structure. The latter, because they are behavioural parameters of private economic agents should be termed 'structural' parameters. A structure so defined is invariant to changes in policy and to changes in the structure of exogenous variables so that identifying them

> "... allows policy evaluation to proceed in the traditional manner."
>
> Wallis (1980)

In concluding his paper Lucas remarked that an implication of his analysis was that if policy was varied in a discretionary manner from period to period then we may never hope to infer anything about private economic behaviour. Put another way, unless policy makers confide with private agents to make policy forecastable, expectations of policy will be ill defined and the economic structure hard to identify. This would make policy analysis a vague and uncertain procedure. We certainly share this view. However, such a constraint in no way binds policy makers to a particular form of policy. In particular the constraint does nothing to advance the case against closed loop policy rules in favour of open loop, fixed rules. A closed loop rule may make policy just as predictable as an open loop rule as long as it is announced and explained. Further if agents see that the rule is adhered to then the policy can be made just as credible as a simpler fixed rule.

We can conclude then by noting that if Lucas' closing remark

> "The preference for "rules versus authority" in economic policy making suggested by this point of view is not ...

> based on any optimality properties of rules in general. ... The point is rather that this possibility {of authority} cannot in principle be sustained empirically." (p.141 of the paper. Brackets denote our words).

was meant to suggest that because of his critique the scope for closed loop discretionary policy was severely limited and that as a result an open loop rule must be adopted we must reject this insinuation outright.

3.2 Time inconsistent policies

In their highly influential paper Kydland and Prescott (henceforth K-P) consider the problem of policy optimisation in a model which contains forward R.E. (K-P(1977)). The conclusions they draw have very serious implications for policy design. Their basic claim is that whereas in a non rational world there is scope for economic stabilisation using optimal control theory in a world where future R.E. are present the scope for such policy disappears and control theory is useless. They argue further that

> "... stabilisation may well be dangerous and it is best that it is not attempted. Reliance on policies such as a constant growth in the money supply and constant tax rates constitute a safer course of action."
>
> K-P (1977) p.487.

These conclusions constitute a powerful attack on discretionary (closed loop) policy.

Other economists, notably Sargent and Wallace (1972) and Lucas (1975) have constructed and advanced models in which discretionary policy is useless but as we shall see in the next section the force of their criticisms rests on the validity of the particular models that they advance. It is because K-P's proposition is independent of such considerations that it forms with Lucas' critique one of the two most important attacks on standard policy analysis.

To expose their argument it is sufficient to use the simple model they present in section II of their paper (equations 2.1 to 2.3 below).

$$W = w(x_1, x_2, \Pi_1, \Pi_2) \tag{2.1}$$

$$x_1 = F(\Pi_1, \Pi_2^e), \quad (\Pi_2^e = \Pi_2) \tag{2.2}$$

$$x_2 = g(x_1, \Pi_1, \Pi_2) \tag{2.3}$$

w is a welfare function in terms of a policy instrument, Π and a state variable x. Subscripts 1 and 2 denote time periods 1 and 2 respectively. (2.1) states that not only are the targets of importance to the policymaker but also that there are welfare costs associated with the levels (and changes in the levels) of the instruments. Note that the system is deterministic so that the policy Π_2 is perfectly anticipated in period 1. The more interesting stochastic case is discussed below.

K-P go on to define a consistent policy solution to the model:

> "A policy Π is consistent if, for each time period t, Π maximises (2.1) taking as given previous decisions, ... and that future policy decisions are similarly selected".
>
> op cit p.475.

It is clear from this definition that consistency requires that the policy Π be derived by dynamic programming methods.

The consistent plan then must set Π_2 such that (2.1) is maximised given the past decisions Π_1 and given x_1. Explicitly, Π_2 must be the solution to

$$\left.\frac{dw}{d\Pi_2}\right|_{x_1\Pi_1} = \frac{dw}{dx_2}\frac{dg}{d\Pi_2} + \frac{dw}{d\Pi_2} = 0 \tag{2.4}$$

Because the consistent policy ignores the influence of 'tomorrow's' policy (Π_2) on 'today's ' state (x_1) it is suboptimal. In the second time period the policy maker will wish to abandon the plans for Π_2 that were made in period 1 because (to his surprise) his plan for Π_2 has changed period one's state. A truly optimal policy would take this into account. Explicitly, Π_2 should be the solution to

$$\frac{dw}{dx_2}\frac{dg}{d\Pi_2} + \frac{dw}{d\Pi_2} + \frac{df}{d\Pi_2}\left[\frac{dw}{dx_1} + \frac{dw}{dx_2}\frac{dg}{dx_1}\right] = 0 \tag{2.5}$$

The reason why (2.4) is adopted by policy makers in preference to (2.5) K-P argue, is that there is no mechanism to force future policy makers (the people responsible for setting Π_2 in period 2) to take into account the effects of their action (in period 2_2) on the outcome of the current period. In other words, period 1 will be history as far as the planners in period 2 are concerned and policy for the current and future periods may be revised away from the initial plan. The optimal policy (2.5) is therefore said to be time inconsistent.

We defer answers to K-P's critique to the next section. Suffice to say for the present that if expectations rely on policy announcements then the necessity to make these announcements would itself guarantee that in the future, policy makers would not cheat or renege on previously planned policy. For this reason the optimal policy in (2.5) will not be revised and the time inconsistency of the policy is removed.

To close this section we note that to have R.E. is not sufficient to induce time inconsistency. Referring to our simple example, the optimal value of Π_1 must satisfy

$$\frac{dw}{dx_1}\frac{df}{d\Pi_1} + \frac{dw}{d\Pi_1} + \frac{dw}{dx_2}\left[\frac{dg}{d\Pi_1} + \frac{dw}{dx_2}\frac{dg}{dx_1}\frac{df}{d\Pi_1}\right] = 0$$

If there are no direct instrument costs except in the terminal (second) period then $dw/d\Pi_1=0$. Further, if the function g was directly independent of Π_1 $(dg/d\Pi_1=0)$ so that the current state is not directly influenced by lagged policy, a not too unreasonable assumption then this condition becomes

$$\frac{df}{d\Pi_1}\left[\frac{dw}{dx_1} + \frac{dw}{dx_2}\frac{dg}{dx_1}\right] = 0$$

The term in square braces is the same as that in (2.5) and is equal to zero in this special case. This leaves the optimal solution for Π_2 from (2.5) identical to the consistent solution for Π_2 from (2.4). The problem of time inconsistency disappears. This result holds for the general n-period case (see for example Buiter (1979)).

3.3 An answer to K-P's critique

Buiter (1981) and Driffill (1980) have argued that the deterministic model used by K-P is uninteresting and even misleading.

It is uninteresting because

> "... it is uncertainty and the fact that the economy is continuously subjected to unexpected shocks which provides the need for stabilisation policy".
>
> Driffill (1980)

It is misleading because in the presence of unpredictable shocks there is scope for discretionary or interventionist policy. The suggestion advanced by Buiter and Driffill was that the policy plan should feedback from future unpredictable events. In this way, agents are unable to adjust current plans to (the stochastic part of) future policy in the way that K-P say they can simply because the feedback part of future policy is unpredictable.

To expose their arguments we consider the following model from Driffell's paper

$$y_t = Ay_{t-1} + B_0x_t + \sum_{i=1}^{\infty} B_i x^e_{t+i}|_{t-1} + u_t \tag{3.1}$$

y_t and x_t are vectors of state and control (instrument) variables and u_t is a vector of independently distributed disturbances. A, B_0 and B_1 are conformable and known parameter matrices and the information set includes all policy plans made at time t-1 and all variables dated t-1 and earlier.

The policy objective is to choose x_t so as to minimise

$$W_0 = [\sum_{t=1}^{T} y_t'K_ty_t]^e|_0 \tag{3.2}$$

The suggestion then is to plan the following policy for time τ

$$x_\tau = x_\tau|_0 + \sum_{t=1}^{\tau-1} \gamma_{\tau i}u_i \tag{3.3}$$

The first term in (3.3) is deterministic and represents the plan for period τ made at the start of the planning horizon. This policy cannot be determined recursively by dynamic programming. This is because it explicitly recognises the dependency of current outcomes on future policy thus violating the optimality principle, the time recursiveness that dynamic programming exploits. As we noted above credibility requires that this deterministic plan is rigidly adhered to. This component will not be revised therefore. The second part of (3.3) represents the response to innovations that occur between time 0 and τ. The γ coefficients have subscript τ to emphasise the fact that at each and every time period this response component is revised in the light of the new information that has become available.

To get a feeling for how this optimal policy is determined, using (3.3) we can solve (3.6) as follows

$$y_t = y_t^e|_0 + \sum_{\tau=1}^{t} \Delta_{t\tau} u_\tau \tag{3.4}$$

where

$$y_t^e|_0 = A^t Y_0 + \sum_{k=1}^{t} \sum_{i=0}^{\infty} A^{T-k} B_i x_{k+i}^e|_0$$

is the deterministic part of y_t itself the sum of a policy independent component and a deterministic policy induced component and where

$$\Delta_{t\tau} = \sum_{k=\tau+1}^{t} A^{t-k} \sum_{i=0}^{\infty} B_i \gamma_{k+i}^e|_\tau + A^{t-\tau} \text{ for } \tau=1 \rightarrow t-1$$

and

$$\Delta_{tt} = I$$

are the matrix weights for each time period assigned to the innovations in the stochastic innovation-dependent part of the policy.

Substituting (3.4) into (3.2) gives the unconstrained problem:

$$\max W_0 = \sum_{t=1}^{T} [y_t'^e|_0 K_t y_t^e|_0 + (\sum_{\tau=1}^{t} u'\Delta_{t\tau} u_\tau)^e|_0]$$

This problem is exactly analagous to that of chapter 2, section 1. As we did then we can minimise W in two stages dealing with its deterministic and stochastic parts in turn. The former problem is

$$\min \sum_{t=1}^{T} y_t'^e|_0 \, K_t \, y_t^e|_0$$

with respect to $x_1^e|_0, \ldots\; x_T|_0$

and the latter problem is

$$\min \sum_{t=1}^{T} \left(\sum_{\tau=1}^{t} u_\tau' \Delta_{t\tau}' k_t \Delta_{t\tau} u_\tau \right)^e |_0$$

with respect to the parameter sets $\gamma_{t\tau}$ ($t=1\ldots.\tau$, $\tau=1\ldots.t-1$).

The beauty of the stochastic component is that although it is determined in a mechanical way, the policy revisions are unpredictable before they occur. The policy makers are seen therefore to be acting in a rational manner, maintaining credibility and, of course the ability to revise plans in a discretionary manner.

Although this problem is not subject to traditional dynamic programming techniques it is a standard optimal control problem. As we might expect therefore, the standard result on dominance applies. The rule that allows reaction to stochastic news as it becomes apparent and that also controls the mean path of the economy dominates the rule that only controls the latter. Closed loop feedback rules achieve a lower expected loss than do fixed deterministic open loop rules. This result contradicts K-P's conclusions that

> "... policy makers should follow rules rather than have discretion ..."

and that

> "... there is no way control theory can be made applicable to economic planning when expectations are rational".

K-P were led to these conclusions by considering a deterministic model. In such a world there is no 'news' upon which policy may be revised.

3.4 The neutrality of anticipated policy

K-P have not been the only critics of discretionary policy. Other authors have advanced the view that closed loop stabilisation policy has no bearing on the real economy when this policy is anticipated. This assertion can be stated formally in the context of a partially solved reduced form of a linear R.E. model due to Barro (1977).

$$y_t = A_1 z_t + \sum_{i=0}^{T} B_i \left(x_{t-i} - x^e_{t-i|t-i}\right) + \sum_{i=0}^{T} C_i x_{t-i} + u_t \qquad (4.1)$$

y_t, x_t and z_t are vectors of endogenous variables, instruments and predetermined and exogenous variables respectively. u_t is a vector of white noise error terms.

(4.1) is the partial solution to a more general R.E. model. The question as to whether or not such a solution will exist in general is addressed below.

The effect of an anticipated increase of x_{t-i} on y_t is

$$\frac{dy_t}{dx_{t-i}} + \frac{dy_t}{dx^e_{t-i|t-i}} = C_i$$

The new classical school of economists assert that

$$C_i = 0 \quad \forall_i$$

that is that anticipated policy has no influence on real economic variables (y_t). In this section we wish to make the point that this impotence of anticipated policy has nothing whatsoever to do with the presence of R.E. Rather it is a feature of the way in which these models are constructed. More explicitly, elements of the matrices in (4.1) are subject to restrictions which are the implication of some a priori economic theory.

All this may at first seem obvious. However a proliferation of models emerged in the 1970's in which policy (primarily monetary policy) had no impact on real behaviour. The models of Barro (1977), Lucas (1975) and Sargent and Wallace (1975) are notable members of this class. This in turn led to counter examples where the scope for policy did exist and it was this fragmented debate that provoked Pagan (1980) to construct generalised and formal theories on the existence of economic policy when R.E. are present. The paper is general and rigorous in its approach and so we restate the essence of his argument below. For a rigorous treatment of the matter the reader is referred to the actual paper.

Focus is on deterministic models so that it is the expected mean path of the state variables that are the targets of policy. We start with models in which there are no predetermined variables and define the existence of policy to be the ability to achieve m targets by the setting of m instruments. The existence of policy in this context depends on the existence of a mapping from instrument space to target space and on the form of this mapping. In particular if x_t is a vector of m instruments and y_t is a vector of K targets and if y_t and x_t are related by the mapping

$$y_t = \Pi x_t + b_t \tag{4.2}$$

(b_t is a vector of exogenous variables), then policy exists

$$\text{iff} \quad \rho(\Pi) \geq K \tag{4.3}$$

where ρ denotes rank.

The question of the existence of such a mapping in the presence of R.E. (a question raised above) is answered for a general dynamic model in Theorem 8 on p.27 of his paper. Broadly speaking the proof states that such a mapping must exist in order that R.E. of the endogenous variables be defined.

The parameter matrix Π will be a matrix function of the structural parameter matrices. The existence of policy then depends on the restrictions imposed on these structural parameter matrices.

To take a concrete example, consider the model

$$y_t = Ax_t + By_t^e|_{t-1} \tag{4.4}$$

(we have suppressed the exogenous variables as it adds to simplicity without affecting the argument). The solution to (4.4) is a simple problem (see Wallis (1980)) and is achieved as follows:

$$y_t = Ax_t + By^e_t|_{t-1} \Rightarrow y^e_t|_{t-1} = Ax^e_t|_{t-1} + By^e_t|_{t-1}$$

$$\Rightarrow y^e_t|_{t-1} = (I-B)^{-1} Ax_t|_{t-1}$$

$$\Rightarrow y_t = Ax_t + B(I-B)^{-1} Ax^e_t|_{t-1}$$

Because the model is deterministic, policy is likewise deterministic so that

$$x_t{}^e|_{t-1} = x_t$$

and

$$y_t = (A+B(I-B)^{-1}A)x_t = (I-B)^{-1}Ax_t \tag{4.5}$$

is the solution to (4.4).

Note that we have assumed that $(I-B)^{-1}$ exists. If it does not the R.E. vector $y_t{}^e|_{t-1}$ is not defined. Note also the nonsingularity of (I-B) guarantees the existence of a mapping from x_t to y_t. Assuming that the whole vector y_t is targetted then the condition (4.5) is

$$\rho[(I-B)^{-1}A] = \rho(A) \geq K$$

and this is independent of whether or not R.E. enter (4.4) (that is, independent of whether B is a matrix of zeros or not).

To press his argument further Pagan shows that a simplified version of McCallum's wage-price-output model (McCallum (1979)) can be written in the form of (4.4) and then proceeds to show that the restrictions McCallum places on the counterpart of our matrix A is the root cause of the failure of monetary policy to exist. In his model McCallum's a priori restrictions on A make it rank deficient.

The existence of policy in dynamic models with (forward) R.E. is a more difficult concept. Instead of being able to achieve a particular target as in the static case the existence of policy is defined in terms of being able to achieve a target path. Again existence of policy depends on the existence of a mapping of the targets onto the instruments (this is guaranteed by theorem 8) and on the rank of the matrix premultiplying the vector of instruments in this mapping. Again the mere presence of R.E. on its own has no bearing on this condition and is therefore neither a necessary nor sufficient condition for the failure of policy to exist.

The introduction of unpredictable stochastic terms opens a new dimension to policy. Even if models are constructed in such a way that anticipated policy fails to achieve an expected target path, in a stochastic world the variance of the target may be influenced by policy. This policy will always exist even in the class of models proposed by the New Classical School because such policy is related to unpredictable shocks and is therefore always unanticipated.

Finally even when (4.2) holds for the class of models in (4.1) policy may still exist. The argument is advanced by Buiter (1980) and relies on asymetric information. Broadly speaking, if economic agents are slower to perceive a policy than the authorities are at setting it then policy will exist. For full details of this argument the reader is referred to the original paper.

3.5 Conclusion

We have shown in this chapter that despite vigorous claims to the contrary the superiority of discretionary rules over fixed (open loop) rules is unaffected by the existence of R.E. forward or current. The case advanced by the New Classical Macroeconomics School in favour of open loop policy rests purely on the a priori restrictions placed on their models by their economic theory. The existence of R.E. in itself has no bearing on the matter.

This is all very comforting in terms of the usefulness of the following chapters. We can now proceed to grapple with the methodological issues raised by R.E. and in particular the problems posed for estimation, (chapter six), testing (chapter seven) and numerical simulation (chapters four and five).

4. MULTIPLE SOLUTIONS AND CURRENT PRACTICE IN SIMULATING MODELS CONTAINING FORWARD R.E.

4.1 The problem of multiple solutions.

The problem of multiple solutions in models containing forward R.E. is now well known. Efficient estimation, the use of powerful tests and policy simulations, three crucial procedures in policy analysis all require a unique solution in terms of observables for all the endogenous variables. The existence of a continuum of solutions in R.E. models then undermines the whole feasibility of policy analysis against the background of the hypothesis.

In this section we review the problem in the light of Blanchard (1978), and Gourieroux, Laffont and Monfort (1982). (Henceforth the latter shall be referred to as G.L.M.).

The analysis in this literature focuses on the single equation

$$p_t = \alpha p^e_{t+1} + m_t + v_t \qquad (1.1)$$

where p is the price level, m the outstanding nominal money stock and 'e' denotes an expectation. The usual R.E. assumption employed is

$$p^e_{t+1} = E(p_{t+1} | \Omega_t)$$

where Ω_t is an information set containing current exogenous variables, past values of p and m and the model in (1.1). The assumption of a current information set is made purely in accordance with the literature and the analysis is equally valid with a set dated at (say) t-1.

(1.1) is the reduced form of a system of structural equations and it describes euilibrium in the money market. In accordance with the literature our analysis shall also focus on (1.1) but there are more compelling reasons for its consideration. It is easy to show that many linear models with forward R.E. can be reduced to a form or forms such as (1.1). As an example consider the system

$$Y_t + Ay_{t-1} + By^e_{t+1|t} + Cx_t + u_t = 0 \tag{1.2}$$

where A and B are nxn matrices of full rank, C is an nxk matrix and u_t a vector of (quasi) reduced form errors.

(1.2) may be written as

$$z_{it} = \lambda_i z^e_{it+1|t} + \bar{c}_i x_t + \bar{u}_{it} \quad i = 1, \ldots n$$

where λ_i are the diagonal elements of Λ in the matrix decomposition

$$P\Lambda P^{-1} = \begin{bmatrix} I_n & A \\ I_n & 0 \end{bmatrix}^{-1} \begin{bmatrix} -B & 0 \\ 0 & I \end{bmatrix} = \begin{bmatrix} 0 & I \\ -A^{-1}B & -A^{-1} \end{bmatrix}$$

z_{it} are the elements of the vector

$$P^{-1} \begin{bmatrix} Y_t \\ Y_{t-1} \end{bmatrix} \qquad \text{(P is a matrix of eigenvectors)}$$

and $\bar{c}_i x_t$ and $\bar{u}_{it}$ are the respective elements of

$$P^{-1} \begin{bmatrix} 0 & 0 \\ -A^{-1}C & 0 \end{bmatrix} \begin{bmatrix} X_t \\ 0 \end{bmatrix} \quad \text{and} \quad P^{-1} \begin{bmatrix} 0 & I_n \\ A^{-1} & -A^{-1} \end{bmatrix} \begin{bmatrix} -u_t \\ u_t \end{bmatrix}$$

We now set about solving (2.1). Blanchard (1978) uses the method of undetermined coefficients due to Muth (1960) in solution of (1.1).

Writing a general solution gives

$$P_t = \sum_{i=1}^{\infty} a_i m_{t-i} + \lambda m_t + \sum_{i=1}^{\infty} c_i m_{t+i|t} + u_t \tag{1.3}$$

Leading (1.3) one period and taking expectations gives

$$P_{t+1|t} = \sum_{i=1}^{\infty} a_i m_{t-i+1} + \lambda m_{t+1|t} + \sum_{i=1}^{\infty} c_i m_{t+i+1|t} \tag{1.4}$$

Substituting (1.3) into (1.1) and equating coefficients with those in (1.3) gives

$$\begin{aligned} &\text{(i)} \quad a_1 = (\lambda-1)/\alpha \\ &\text{(ii)} \quad a_i = \alpha a_{i+1} \qquad i=1,\ldots\infty \\ &\text{(iii)} \quad c_1 = \alpha\lambda \\ &\text{(iv)} \quad C_i = \alpha^{-1} C_{i+1} \qquad i=1,\ldots\infty \end{aligned} \tag{1.5}$$

and we see immediately that the general solution is indeterminate. There are a continuum of solutions and all of them are parameterised by λ.

Heuristically this indeterminacy arises because from (1.1) the market at time t has to determine (the expected path of) current and future prices so that however far forward the time horizon is taken there will always be one market (clearing condition) too few to determine a unique price path. We return to this question of the source of the indeterminacy below.

Writing the general solution in (1.2) as

$$p_t = \lambda p_t^F + (1-\lambda) p_t^B \tag{1.6}$$

where

$$p_t^F = m_t + \sum_{i=1}^{\infty} \alpha^i m_{t+i|t}^e + u_t \tag{1.7}$$

$$p_t^B = - \sum_{i=1}^{\infty} \alpha^{-i} m_{t-i} + u_t \tag{1.8}$$

we see that the set of solutions in (1.3) can be expressed as a linear combination of the well-known "forward" and "backward" solutions.

The approach (above) of Blanchard has been criticised in the subsequent work of G-L-M. Couching the problem in a more general framework they expose the deficiencies of this solution method. In particular it is noted that when expectations of more than one step ahead occur then both "forward" and "backward" solutions may fail to exist invalidating (1.3) or (1.6) as a solution. In return we must note that models of this kind may have representations in terms of canonical variables in the form of (1.1) and at least one solution must exist. The point is taken however that the solution method lacks generality because it does not necessarily provide the complete set of solutions. In response to this G.L.M. propose a general method of solution based on standard difference equation techniques. We briefly outline the method here. For a full treatment the reader is referred to G.L.M.'s paper itself.

The general solution for P_t may be written as

$$P_t = P_t^P + \alpha^{-t} M_t$$

where P_t^P is a particular solution to (1.1) and M_t is any martingale process (see Doob (1953)). G.L.M. provide specific examples when the x-process is of AR, MA and mixed ARMA form. As an immediate example consider the AR(1) process for m_t

$$m_t = \rho m_{t-1} + \varepsilon_t \quad |\rho|<1, \ \varepsilon_t \text{ is white noise.}$$

The particular solution in this case is

$$P_t^P = Z_t/1-\rho\alpha \qquad \rho\alpha \neq 1$$

and this exists for all α and ρ. Any martingale forms a solution to the homogenous equation

$$P_t = \alpha P_{t+1|t}$$

so that this time the solution is not parameterised by a constant but by the specification of the martingale process. The crucial point here is that each of the solutions implies different behaviour. Focussing on the solutions in (1.7) and (1.8) we see that under (1.7) current price is independent of current unanticipated changes in the nominal money stock whereas the opposite is true in (1.8).

At this stage we might suspect that the model (1.1) is poorly specified and in some sense incomplete. We may even be tempted to assert that well specified models and in particular those built on a choice theoretic base would not suffer such problems. To examine this assertion we turn to describe one such choice theoretic model due to Mirrlees (1968).

Under an assumption of perfect foresight (an assumption which coincides with R.E. in a non-stochastic economy), profit and utility maximisation, and full arbitrage in assets markets, Mirrlees derives a set of (two) behavioural relationships describing an equilibrium growth process in a two sector model in discrete time.

The model is one of overlapping generations where consumers leave and receive no bequests, work for a period, live for one further period and (for further simplicity) have homothetic preferences. The two sectors produce capital and consumption goods respectively. The two equations describing equilibrium growth are

$$W_t S(F_k(P_{t+1}^e)) = P_t(F_p(P_t,K_t)+K_t) \tag{1.9}$$

$$(1+n)K_{t+1} = F_p(P_t,K_t) \tag{1.10}$$

with $P_{t+1}^e = P_{t+1}$ (perfectly foreseen capital gains) and where

W_t is the wage rate per worker,

S is the proportion of income saved,

n is the rate of population growth,

P is the price of a unit of capital relative to the consumption good whose price is unity

K is the real capital stock per man, and

$F = F(P,K) = \text{Max}\{p.y:(y,-k)\in Y\}$

where Y is a convex and bounded per capita production set.

The first equation describes equilibrium in the assets market with savings (on the L.H.S.) being just sufficient to finance next period's captial stock. The average propensity to save, s, is a function of the marginal product of capital which in equilibrium growth is equal to the rate of return on assets. Note that as a result of there being no bequeathals or bequests saving can only take place out of current wages earned during the first period of life. The second equation is the equilibrium constraint for the real side of the economy. Note that F_p is the output of the captial good sector because at a profit maximising optimum

$$\max Y = F = P(\overline{K}) + 1.C$$

(where C is the consumption good whose price is normalised to unity, and $\overline{K}$ is production of K) and so this second equation is a technological constraint on the evolution of the capital stock.

Although the model is in descrete time a phase diagram is still a useful illustrative (though not rigorous) tool for tracing the equilibrium paths for K and P. The stationery locus for K ($K_{t+1}=K_t$) is defined from (1.9) as

$$nk = F_p \qquad (1.11)$$

and that for p from (1.10) as

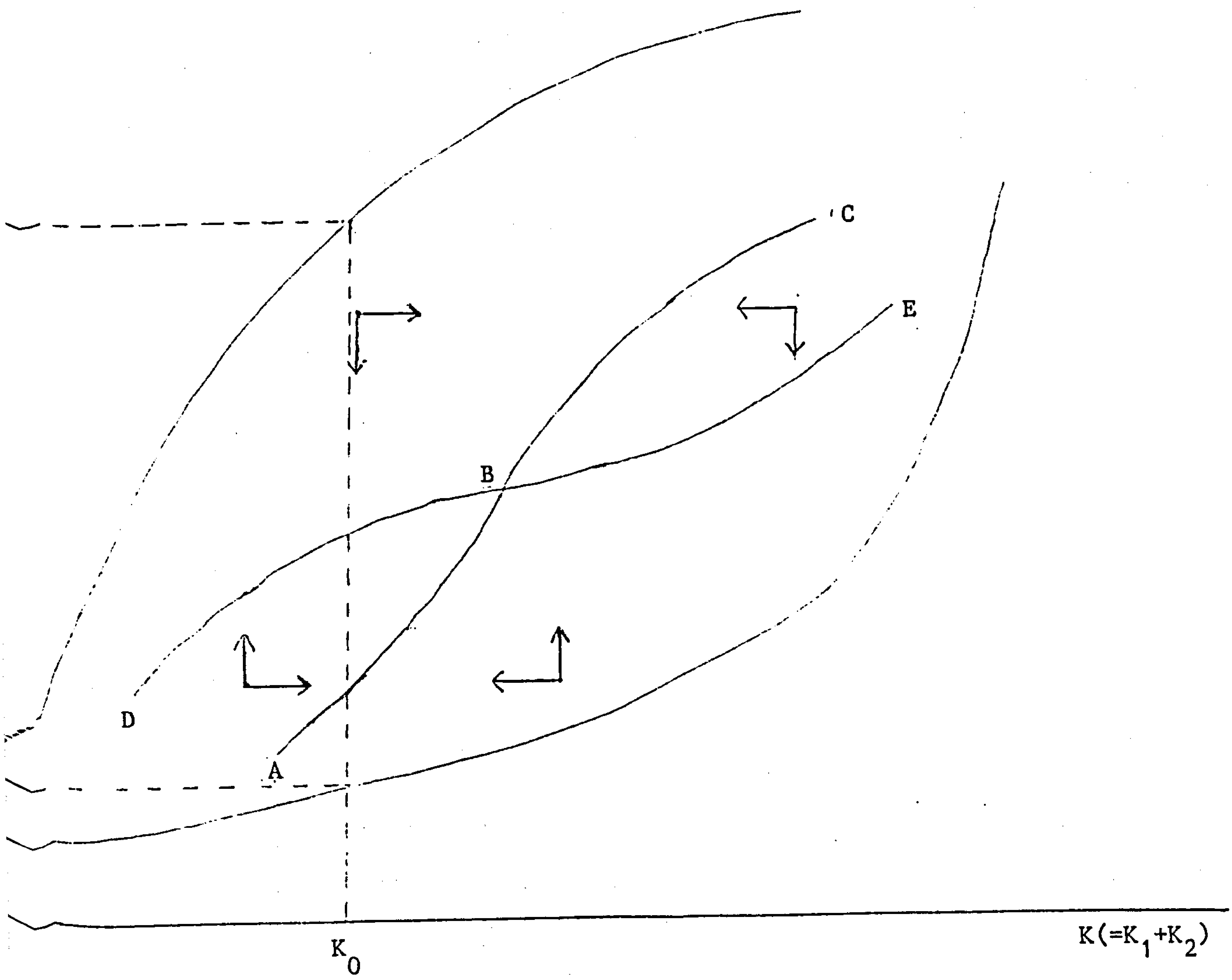
$$(F-KF_k)s(F_k) = p(F_p+K) \qquad (1.12)$$

(The first term in (1.12) is the difference between the value of output per head and rental returns on capital per head and is therefore just w).

We shall see below that the growth process depicted in Figure 4.1 is possible. In the diagram K_1 is capital in sector 1 (consumption goods) and K_2 is capital in sector 2 (capital goods). From (1.12) if (p_t,k_t) lies above ABC (the stationery locus for capital) then the capital stock is growing ($K_{t+1}>K_t$). Providing that

FIGURE 4.1

REPRESENTATION OF AN EQUILIBRIUM GROWTH PROCESS



$$F_{kp} < -1 \text{ and } S_{f_k} > 0 \text{ [6]}$$

then from (1.11) if (p_t, K_t) lies to the left of DBE (the stationery locus for prices) then prices are rising (i.e. $p_{t+1} > p_t$). The second condition in (1.13) requires that the savings propensity rise with the rate of interest (equal to f_k) and is thus reasonable. The first condition in (1.13) however requires that output of capital must be lower the higher its price and this requires some justification. There will be two forces at work here. The higher its own price the more profitable capital production will be. On the other hand the higher the price is the greater is the cost of the capital input for any extra production (given an exogenous labour force). If the amount of captial per man employed is already relatively great so that its marginal product is relatively low then this second effect may outweigh the first by a sufficient margin to give the desired inequality. Finally then the relative slopes of the two stationery loci as drawn holds providing that

$$1 + sF_{pk} > 0 \text{ or } F_{pk} > -1/s$$

and from our arguments above this is a clear possibility.

Having established that Figure 4.1 is a plausible description of equilibrium growth derived from a choice theoretic model we must reject out earlier assertion that non-uniqueness in R.E. models arises only in poorly specified models. From the diagram given an initial capital stock K_0 we see that any path with initial price in the range p_0' to p_0'' satisfies the equilibrium conditions. Further, many other well specified economic models have the same property. (See for examnle, Calvo (1977)).

We must conclude therefore that the problem lies not with the models but at the very roots of the R.E. hypothesis itself. It is the latter that is "poorly specified" and incomplete. Without further investigation of the process by which R.E. are formed (the so called 'learning process') this incompleteness will remain as will the problem of multiple solutions. [7] The literature in this area is very scant and due to the difficulty of formulating a model for (unobservable) expectation formation it is likely to remain so in the near future. The seriousness

---

[6] Derivation of these conditions is tedious and uninformative and so is not presented here.

[7] One possible direction for such an investigation would be the description of the evolution of R.E. in a Bayesian context.

of the problem for policy analysis has already been noted above so that in the absence of such detailed research described in the preceeding paragraph some medium term answers must be found.

In the next section therefore we analyse some current practices and some solutions to the problem in the context of policy analysis and in the conclusion to that section we propose an alternative procedure based on the data.

4.2 Current practice and terminal conditions.

Current practice in policy analysis in large macro models in this context is dominated by the methodology of Minford (1978), Holly (1980), Fair (1979), etc. which involves the imposition of a terminal condition on a numerical solution to the model. Their procedures form the focus of this section.

Before we turn to this we must first revise some non-numerical suggestions put forward to identify a unique solution in such models.

If we invoke stationarity in (1.1) [8] whilst maintaining the assumption of an AR(1) process for m_t we can distinguish two distinct possibilities dependent on the value of a.

(i) $|a|<|$; in this case the only stationary solution is the forward solution which we may recall is

$$p_t^{\,f} = p_t = 1/(1-\alpha\rho)\; m_t + u_t \qquad (2.1)$$

In terms of the general solution of G-L-M the martingale in this case is forced by the stationority condition to be a sequence of zeros.

(ii) $|\alpha|>|$; in this case the general solution is an infinite one-sided MA process and the complete set of stationary solutions is provided by a linear combination of two particular solutions (G-L-M). Taking the forward and backward solutions (which may be represented as one-sided infinite M.A.'s) as our two particular solutions gives the set of solutions as

---

[8] Note that stationarity is a pre-requisite for likelihood estimation.

$$p_t = \frac{\lambda}{1-a\rho} m_t + \frac{(1-\lambda)}{a-L} m_{t-1} + u_t \tag{2.2}$$

so that λ parameterises a continuum of solutions.

We see then that a priori information about a may or may not be sufficient to give a unique solution when stationarity is invoked.

Consistency with a priori economic theory has been put forward by Sargent and Wallace (1973) as a way out of the problem. Citing our money market example they note that in the forward solution

> "(it is) the price level at each moment (which) adjusts instantaneously in order to ensure that the real balances people hold equal the amount they would like to hold"

and thus argue that the forward solution is more satisfactory at least in this case.

A related argument advanced by Burmeister, Flood and Turnousky (1978) says that because only current and future dated variables enter (1.1) none of which necessarily depend on the past, the forward solution may be chosen.

Whilst arguments such as these are at best contentious (see Blanchard (1978) for a full exposition) they have led to undue consideration of the forward solution in the R.E. literature. Wallis (1980), Ravenkar (1981) and Schmidt (1982) for example focus exclusively on the latter in their discussion of forward R.E.'s.

Finally we note one further solution criterion, namely that of predictive power. Using the results in G-L-M we note that for $|a|>|$ the k-step ahead prediction error satisfies the equality

$$e_t(k) = \lambda e^f{}_t(k) + (1-\lambda)e^B t(k)$$

where e^f and e^B are the respective prediction errors of the forward and backward solutions. The suggestion put forward in ad hoc forms by earlier authors [9] but generalised by G-L-M is to choose the solution that minimises

---

[9] See for example Taylor (1977).

$$\sigma^2(k) = \mathrm{Var}(e_t(k)) = \mathrm{Var}(\lambda e_t^f(k) + (1-\lambda)e_t^B(k)) \quad (2.3)$$

The λ that minimises (2.3) is given as

$$\lambda = \sigma_B^2(k) - \sigma_{Bf}(k)/\sigma_f^2(k) + \sigma_f^2(k) - z\sigma_{FB}(k)$$

where $\sigma_F^2 = \mathrm{Var}(e_t{}^F(k)), \sigma_B^2 = \mathrm{Var}(e_t{}^B(k))$ and

$$\sigma_{FB}(k) = \mathrm{Cov}(e_t{}^F(k), e_t{}^B(k))$$

This general selection procedure incorporates as a special case that of Taylor (1977) who suggests choosing the price path with minimum variance. This is because

$$\lim_{k \to \infty} \sigma^2(k) = \mathrm{Var}\ (p_t)$$

and minimising the L.H.S. in this limiting case minimises the R.H.S.

Rather than impose one of the analytical solutions above the common practice in current policy analysis is to obtain a set of consistent structural estimates using a limited information method and then simulate the model with a terminal condition imposed on one or more of the endogenous variables dated outside the forecast horizon. This procedure has been proposed and used notably by Minford (1978) and Holly and Beenstock (1980) in the U.K. and Fair (1979) in the U.S.A. against the background of the L.I.T.P., L.B.S. and Fair (1976) models respectively.

To solve the indeterminacy problem, numerical solution paths are forced through a point at some time t+N. The terminal value chosen varies. In arguing that terminal conditions in macro models are exactly analagous to transversality conditions in micro models where dynamic optimisation is involved, Minford proposes that the terminal condition(s) for the (expectation(s) of the) endogenous variable(s) be their equilibrium solution from the model. (We interpret 'equilibrium solution' to mean 'steady state solution' in line with Minford and Holly).

> "... the terminal condition is exactly analagous to the transversality conditions in micro-economics which provide the necessary condition for an optimum".

Whilst accepting Minford's analogy H-B see difficulties in practice in obtaining a steady state solution from typical nonlinear models and so propose the condition that at some future date t+N the expected change in the endogenous variables will be zero.

We shall return to the analogy between terminal and transversality conditions below but presently we shall focus on the affect of these two terminal conditions on the solution to our model in (1.1).

We assume that m follows a simple AR(1) process with parameter ρ and that the world is stationary. Recall that there are two distinct cases here; $|a|<|$ in which case prices follow (2.1) and $|a|>|$ where prices follow (2.2).

Minford's terminal condition in either case would be

$$P^e_{t+N|t} = \overline{P} = 0 \quad \text{(say)} \tag{2.4}$$

Using this in (1.1) and solving backwards gives

$$P^e_{t+N-1|t} = aP^e_{t+N|t} + m^e_{t+N-1|t} = m^e_{t+N-1|t}$$

$$P^e_{t+N-2|t} = am^e_{t+N-1|t} + m^e_{t+N-2|t}$$

$$P^e_{t+n-n|t} = P_t{}^e{}_{|t} = P_t - v_t = \sum_{i=1}^{N} a^i m^e_{t+i|t}$$

$$\text{or} \qquad P_t = \frac{1-(a\rho)^N}{1-a\rho}\, m_t + v_t \tag{2.5}$$

Taking the case $|a|<1$ first (in (2.1)), although (2.5) violates (2.1) if N is sufficiently large they are approximately the same.

H-B would use the terminal condition

$$P^e_{t+N|t} = P^e_{t+N+1|t}$$

using (1.1) gives

$$P^e_{t+N|t} = aP^e_{t+N+1|t} + m^e_{T+N|t}$$

and so

$$P^e_{t+N+1|t} = P^e_{t+N|t} = m^e_{t+N|t}/1-a$$

Using this in (1.1) and substituting backwards gives

$$P^e_{t+N-1|t} = (a/1-a)\ m^e_{t+N|t} + m^e_{t+n-1|t}$$

$$P^e_{t+N-2|t} = (a^2/1-a)\ m^e_{t+N|t} + am^e_{t+N-1|t} + m^e_{t+N-2|t}$$

$$P^e_{t+N-N|t} = P_t^e = P_t - v_t = (a^N/1-a)\ m^e_{t+N|t} + \sum_{i=1}^{N} a^i m^e_{t+i|t}$$

$$\text{or } P_t = \left[\frac{(a\rho)^N}{1-a} + \frac{1-(a\rho)^{N-1}}{1-a\rho}\right] m_t + v_t \qquad (2.6)$$

Again if $|a|<1$ this approximates (2.1) well for a sufficiently large value of N.

So we see that in the case where invoking stationarity yields a determinate solution for sufficiently large N the two terminal conditions described give this (forward) solution fairly accurately.

Turning to the more interesting case where invoking stationarity fails to give a determinate solution, the relevant comparison is now with (2.2). A glance at (2.2) reveals that there is no value for λ for which (2.5) or (2.6) is an exact solution. Even if $\lambda=1$, the most favourable case, because $|a|>1$, $|\rho|$ would have to be very small indeed for (2.5) or (2.6) to approximate (2.2) well. For $\lambda\neq 1$, the situation is even worse since not only would the impact multipliers be wrongly estimated but the infinite distributed lag response of prices to money in (2.2) would be totally ignored.

Forcing an absolute steady state N periods in the future then would typically truncate a distributed lag response (for $\lambda\neq 1$ that is) and an 'equilibrium' type response is observed. This is interesting since it implies that whilst a model with (say) a long run neutrality property may have the potential for exhibiting fluctuating economic activity in the short run, the imposition of a terminal condition will suppress this potential.

Further, we see no reason why this result should not carry over to models with either more than one forward R.E. term and/or more than one endogenous variable (i.e. simultaneous models). In either case the system is representable in terms of canonical variables as a set of separable equations like (1.1) and forcing the endogenous variables into on absolute steady state at t+N also forces these canonical variables into a steady state since they are just linear combinations of the former.

The above analysis also provides an explanation of Minford's numerical result that simulated endogenous variable paths are relatively insensitive to the choice of N the terminal date. A glance at (2.5) and (2.6) above shows this insensitivity clearly.

We now turn to the theoretical basis of terminal conditions namely that they are the macro analogy of transversality conditions and to this end we briefly review the role of the latter in microeconomics.

Consider the intertemporal optimisation problem of a utility maximising agent. For an individual starting a T-period life in period 1 and earning income from the sale of an exogenous labour endowment.

$$\max \bar{U} = \sum_{t=1}^{T} U_t(y_t, q_t) \text{ s.t. } y_{t+1} = y_t - q_t + w_{t+1} \text{ and } y_1 = \bar{y}$$

Or in Lagrangean form this is

$$\max L = \sum_{t=1}^{T} [U_t(y_t, q_t) + \lambda_t(y_{t+1} - y_t + q_t - w_{t+1})]$$

where y_t is income at time t consisting of savings and constant wage income (w_t) and q_t is the value of consumption.

The F.O.C. are

$$\text{(i)} \quad \frac{dL_t}{dq_t} = \frac{dU_t}{dq_t} + \lambda_t = 0 \quad (t = 1 \rightarrow T)$$

$$\text{(ii)} \quad \frac{dL_t}{dy_t} = \frac{dU_t}{dy_t} - \lambda_t + \lambda_{t-1} = 0 \quad (t = 2 \rightarrow T)$$

$$\text{(iii)} \quad \frac{dL_t}{d\lambda_t} = y_{t+1} - y_t + q_t - w_{t+1} = 0 \quad (t = 1 \rightarrow T) \qquad (2.7)$$

with $\quad \text{(iv)} \quad y_1 = \bar{y}$

We now have only 3T equations to solve for 3T+1 variables (namely $\lambda_1 \rightarrow \lambda_T, q_1 \rightarrow q_T$ and $y_1 \rightarrow y_{T+1}$). The final or terminal value of income (y_{T+1}) is indeterminate and so the intuitively plausible restriction is invoked that the consumer must plan to have no income after his death. That is that

$$y_{T+1} = 0$$

This is called a transversality condition since it allows us to 'transverse' individual decisions and aggregate over many agents born at different dates. If the utility function is of a 'convenient' form an aggregate closed form expression may be derived describing the evolution of (say) aggregate income or savings over time independent of any generation. It is often the case however that such optimisation problems have only numerical solutions. In this case a macro equation (aggregate decision rule) corresponding to the situation in (2.7) may be interpreted as some form of approximation. Either way transversality conditions are 'used up' in the derivation of such macro rules (implicitly in the case of the approximation and explicitly in the case of the closed form solution) and are not available for invokation at some later stage of the analysis, Minford's analogy then is somewhat surprising.

As a specific example for the R.E. case consider the growth model in section 1. Recall that the individual's problem was to maximise utility for the two periods of his life working in the first but not in the second. The assumption of no bequests gives us an initial condition like (2.7) and the assumption of no bequeathals gave us a transversality condition like (2.8). This latter with the initial condition enabled us to write the generation independent equation (1.9) describing the evolution of savings and thus capital through time in equilibrium growth. (Recall that this equation simply said that all current outstanding financial assets needed for the current capital stock had to come from current savings, a direct consequence of the transversality condition of no bequeathals). Referring to the diagram illustrating the growth process we conclude that the transversality condition was required to draw it in the first place and invoking it again is of no consequence for the solution. We must note however that invoking a terminal condition such as

$$P_{t+N} = \overline{P}$$

<u>is</u> of consequence as it gives us a determinate path in the diagram. This terminal condition is not the transversality condition here and in general

will not be in any model. Indeed if it were it would be impotent as a device to obtain a determinate solution in models containing forward R.E.

Clearly then justification of terminal conditions may not be given by an analogy with transversality conditions. The interpretation of terminal conditions that we are left with is vacuous. They simply force the model into a steady state at some point in time t+N and so arbitrarily choose a solution.

This conclusion is hardly surprising since we saw in section 1 that the indeterminacy in R.E. models lies at the roots of the hypothesis and not the model. Invoking a condition for a determinate solution that pertains to the formulation of the latter and not the former is therefore an arbitrary way to proceed.

What possibilities are we left with then? We saw at the beginning of this section that one possibility is to invoke a solution on the grounds of consistency with a priori economic reasoning. Such is Taylor's procedure of imposing a minimum variance condition on a price path. We have noted however that any such a priori reasoning is eminently open to challenge. What is not subject to controversy however is the data itself. Since it is the data that are a product of economic behaviour a solution consistent with the former must be consistent with the latter and we therefore propose that the solution (if identified) be jointly estimated with the parameters of the model.

We consider for this purpose identification and estimation of the system

$$y_t = Ay^e_{t+1}|_t + Cx_t + v_t \tag{2.8}$$

where A and C are gxg and gxk matrices and v, y and x are gx1, gx1 and kx1 vectors respectively. For simplicity we assume that A is of full rank although the analysis is equally valid when this is not so. We also assume separate AR(p) processes for each element of the exogenous variable vector x, with lag polynomials $\varphi_1 \ldots \varphi_k$.

We may write the quasi reduced form in (2.8) in terms of observables by parameterising all the solutions to (2.8) by the gxg diagonal matrix λ_g to get

$$y_t = [B\lambda_g G(L) + B[I_g - \lambda_g]LH(L)]x_t + v_t \qquad (2.9)$$

where

$$H(L) = [q_{ij}(a_i - L)^{-1}] \quad i=1\to g,\ j=1,k,$$

$$\text{with } [q_{ij}] = B^{-1}C \text{ and } \lambda_g = \begin{matrix} \lambda_1 & 0 \\ 0 & \lambda_g \end{matrix}$$

$B = [b_{ik}]$ is the matrix of eigen vectors and a_i the eigenvalues of A and C(L) is a lag polynomial matrix representing the 'forward solution' component and therefore of order p-1.

The issue of FIML estimation is taken up in full in chapter 6. For the present we discuss an estimation procedure that may be more computationally simple.

Conceivably we could estimate the parameters of the likelihood function corresponding to (2.9) i.e.

$$\max \quad L(y_t,\ \lambda_g, a_i, b_{ij} | x_t, q_m) \qquad (2.10)$$

$$\text{s.t. } \lambda_i = I \text{ iff } |a_i|<| \quad i=1\to g,\ j=1\to g \text{ and } m=1\to k$$

Because of the strange restriction in (2.10) this likelihood function is highly discontinuous in the neighbourhood of $\lambda_i=|a_i|=1$ and in practice this may cause numerical problems.

We therefore suggest a two stage procedure. Firstly, to obtain estimates of the elements the elements of A and C ($\hat{A}$ and $\hat{C}$) through a limited information procedure such as the e.v.m. of Wickens (1978). Secondly conditional upon $|\hat{a}_i|<|$ impose $\lambda_i=1$ and $|\hat{\hat{a}}_i|<1$ and so remove the discontinuity in (2.11) when it is estimated in the second stage. We note a few points here.

Under this scheme the estimates $\hat{\hat{a}}_i$, $\hat{\hat{b}}_{ij}$, $\hat{\hat{c}}_{ij}$ and $\hat{\lambda}_i$ have a complex distribution of the form

$$F_0(\hat{\hat{a}}_i,\ \hat{\lambda},\ \hat{\hat{b}}_{ij},\ \hat{\hat{c}}_{ij}) = Gh_f(\hat{\hat{a}}_i, \ldots) = (1-G)h_B(\hat{\lambda}_i,\ \hat{\hat{a}}_i) \qquad (2.11)$$

where $G = {}_{-1}^{+1}Sg(\hat{a})d\hat{a}$ [10]

---

[10] A '^' denotes an estimate from the first stage and '$\hat{\hat{}}$' denotes one from the second stage.

g is the marginal distribution of the e.v.m. estimator of $\hat{a}_i$ (in the first stage), h_f is the distribution of the estimates from the forward solution under the initial hypothesis (subscript '0') and h_B is that from the mixed solution.

The method obviously provides consistent estimates since in infinite samples the first stage provides the 'correct' hypothesis upon which to condition the second stage. Obviously the final estimates are efficient if the information accrued at the first stage is considered as true in some sense and forms the maintained hypothesis for the second stage.

Because the parameters of the quasi reduced form (2.8) enter the final 'observable' form (2.9) in such a highly non-linear fashion sufficient conditions for identification are in general not available, (see for example Pesaran (1982)). We therefore only discuss necessary conditions for the identification of the parameters in (2.8) and the solution.

If our first stage reveals that h of the a_i's lie within the unit circle then we have $g-h+g^2+gk$ parameters to solve for from the equations [11]

$$[\hat{\Pi}(L) - B\lambda' gC(L) + B[I_g-\lambda_g']LH(L)] = [0] \tag{2.12}$$

where $\lambda_g' = \begin{bmatrix} 1 & & & 0 \\ & 1 & & \\ & & \lambda_{h+1} & \\ 0 & & & \lambda_g \end{bmatrix}$ and other symbols are as before.

For (2.12) to have a non trivial solution there must be at least as many independent equations as this. The equations in L^{p+i} $(\forall i>0)$ provide only g-h equations because $\hat{\Pi}(L^{p+i})$ $i>0$ are restricted by the lag polynomial in H(L). A necessary condition for identification then is

$$g + k \leq kp \tag{2.13}$$

So we see that if there are more exogenous variables than endogenous variables and if $p \geq 2$ (neither of which is very restrictive) then this

---

[11] Note that our first stage does not aid identification since the a priori information this latter yields, reduces the number of equations (in (2.12)) and parameters (λ's) by the same amount.

'order' condition is satisfied.

The existence of lagged exogenous varibles worsens the situation, however. For example when DL is added to C in (2.8), (2.13) becomes

$$g + 2k \leq kp$$

On the other side of the balance sheet however an information set dated at t-1 [12] aids identification as (2.13) becomes

$$g \leq Kp$$

Further the elements of A and C are very likely to be subject to restrictions because most macro models are in fact heavily overidentified. Such restrictions may aid identification of (2.8) and the solution. We believe therefore that not only are such models likely to be identified but that the R.E. hypothesis in this case may yield overidentifying restrictions which could form the basis of a test of the hypothesis.

In practice identification beyond some order condition is likely to be a numerical issue focussing on the rank of the information matrix. There are problems however in using Newtonian optimisation techniques where the latter is rank deficient since most of these routines require the inverse of this matrix.

4.3 Conclusion

In this paper we have shown that existing practices in policy analysis in the context of models with forward R.E.'s are deficient in one important respect, namely that at some stage they impose a solution using criteria which are at best contentious. Numerical conditions were found to have no firm theoretical basis and so provide a completely arbitrary way of choosing a solution. In response to this we proposed and described joint estimation of the parameters and solution of simultaneous models. Finally we derived some necessary conditions for identification of this model and its solution.

----

[12] Certain authors consider this assumption the more plausible of the two e.g. Wallis (1980).

5. Rational Forecasts from Nonrational Models

5.1. Introduction

It is now widely accepted that the existence of Rational Expectations (henceforth R.E. has profound implications for economic policy analysis. The so-called 'R.E. revolution' has returned the whole question of the effectiveness of monetary and fiscal stabilisation policy to the centre of economic debate. For example the well known policy trade off between inflation and unemployment (the Phillips curve) disappears when the hypothesis is invoked (Lucas, 1973). However imposing the hypothesis on a simultaneous equations econometric model gives rise to complex within and cross equation nonlinear parameter restrictions and to acknowledge these fully requires FIML estimation. Perhaps this computational complexity has inhibited the widespread adoption of R.E.'s in the estimation of the large scale macroeconometric models that are used for forecasting and policy analysis. Nevertheless interest in the implications of the R.E. hypothesis for policy analysis has provoked investigation of the possibility of extracting 'rational policy responses from nonrational models'.

In his influential paper Anderson[13] describes and uses an algorithm for

"simulating standard models under the assumption of R.E. when re-estimation under that assumption is considered too costly"[14]

[13] The simulation algorithm was also developed simultaneously by Fair but for use in a different context to that of Anderson.

[14] All quotes are taken from Anderson (1979) and brackets denote our own words.

By means of a simple alteration of the coding of standard simulations, it is claimed that solution paths may be obtained from existing non-rational structures which approximate those that would have been obtained from models that incorporate R.E. as a maintained hypothesis.

Whilst this method is not put forward as a long term alternative to re-estimation under the maintained hypothesis of R.E. it is suggested that it provides a cheap interim method of simulating under R.E.

"The estimation of a large R.E. macroeconometric model is a costly time-consuming project. While research on the R.E. hypothesis continues it seems useful to develop a method for simulating existing models under the added assumption that expectations adjust in an optimal manner. This chapter presents one such possible simulation method which incorporates a rationality postulate."

Anderson uses this method to derive monetary and fiscal policy multipliers from the St. Louis and FMP models. These latter are compared with multipliers obtained from the existing models and are found to be substantially smaller in magnitude (in fact policy multipliers were halved when the method was invoked).

Because of the quantitative and qualitative significance of these results and of the potential attractiveness of the method to macro modellers the method deserves closer scrutiny. In this chapter we examine the method and evaluate its claims.

The next section provides an outline of the method in the context of a simple two equation model also used by Anderson. Section 3 briefly reviews ad hoc extrapolative proxies in common usage and their implementation in macro modelling. Adaptive expectations and data determined schemes are considered because of their overwhelming popularity. An unbiased and optimal (in the sense of minimum forecast error variance) extrapolative predictor is also described and used in the analysis because it provides a useful benchmark as the 'best' extrapolative proxy available. Section 4 provides some analytical results on multiplier biases using the method whilst in section 5 some numerical examples of these biases are tabulated and discussed. The numerical study uses a simple two equation macro model that bears a close resemblance to the two main equations of the condensed St. Louis model described by Anderson and this with its linearity and simplicity makes it an ideal structure in which to house the analysis. In section 6 we examine further properties of the algorithm. Focus here is on the seriousness of ignoring 'mistakes' during simulation by substituting actual outcomes for expectations and on problems raised by Lucas' critique of policy evaluation (Lucas, 1976). Section 7 provides a summary and conclusion.

5.2 Outline of the Fair-Anderson method

In his influential paper Anderson (1979) undertakes dynamic deterministic ex post policy simulations on the FMP and St. Louis models with a general aim of deriving dynamic policy multipliers. The method is easily exposited in the context of the following simple two equation model explaining a price level and nominal income

$$p_t = \alpha m_t + u_{1t} \tag{2.1}$$

$$y_t = \beta m_t + \gamma p_t^e + u_{2t} \tag{2.2}$$

where superscript 'e' denotes an expectation formed at time t-1 and p_t, y_t, m_t and u_{1t} and u_{2t} are the price level, nominal income, money supply and structural disturbances respectively (all variables in natural logs). The model is incomplete without an assumption about expectation formation and until the R.E. revolution macro models such as this were typically estimated incorporating extrapolative schemes such as

$$p_t^e = \sum_{i=1}^{n} \delta_i \, p_{t-i} \tag{2.3}$$

The R.E. of prices (p_t^*) however is defined as

$$p_t^* = E(p_t | \Omega_{t-1})$$

where Ω_{t-1} is an information set containing the model ((1.1) and (1.2) in this case) and all variables up to time t-1 and so

$$p_t^* = \alpha\hat{m}_t \tag{2.4}$$

where a hat denotes an optimal (minimum forecast error variance) extrapolative prediction.

The R.E. compares starkly with any extrapolative scheme such as (2.3) since it involves optimal prediction of the exogenous variables (in this case, the money supply) and the form of this predictor will obviously depend on the exogenous variable process.

Having obtained estimates of the structural parameters of (2.1) and (2.2) by incorporating (2.3) and imposing an arbitrary restriction on the δ's (such as that they sum to one) to identify γ one might proceed to a simulation exericse. This would normally involve setting all structural errors at their means of zero (for deterministic solution) and numerically solving (2.1) to (2.3) under different money supply settings with a general aim of deriving policy multipliers such as

$$\frac{\partial p_i}{\partial m_j} \text{ and } \frac{\partial y_i}{\partial m_j} \; (i \geqslant j) \quad .$$ [15]

Our simulated values would thus satisfy

---

[15] Obviously in our linear example the multipliers are independent of initial conditions. This is not so in nonlinear models where multiplers such as

$$\left.\frac{\partial p_i}{\partial m_j}\right|_{y_0, \ldots} \text{ are calculated.}$$

$$p_t^s = am_t^s \tag{2.5}$$

$$y_t^s = bm_t^s + c \sum_{i=1}^{n} \delta_i \, p_{t-i}^s = bm_t^s + ca \sum_{i=1}^{n} \delta_i \, m_{t-i}^s \tag{2.6}$$

where superscript 's' denotes a simulated value. Numerical values for multipliers are easily calculated by comparing the time paths of endogenous variables (given the same initial conditions) under different money supply settings.

To repeat the exercise under a 'maintained' hypothesis of R.E. Anderson suggests simulating under the expectations scheme

$$p_t^e = p_t \tag{2.7}$$

in place of (2.3) so as to make 'expectations consistent with the predictions of the model'. Justification of this comes from the fact that R.E.'s differ from actual values only by a stochastic error consisting of current structural disturbances via the reduced form and current innovations in the exogenous variables. In our example the error in the R.E. is

$$p_t - p_t^* = \alpha m + u_{1t} - \alpha\hat{m}_t = v_t + \alpha\varepsilon_t$$

where ε_t is one step ahead prediction error of the money supply. In deterministic simulations however the model is solved with structural disturbances set at their expected values of zero so that imposing (2.7) it is claimed provides approximate rationality. For example in our model such 'consistent' expectations are given as

$$p_t^e = p_t^s = am_t^s \tag{2.8}$$

as an approximation to the R.E. in (2.4). Using (2.2) and (2.8) the substitution for income is then

$$y_t^s = (b + ca)m_t^s \tag{2.9}$$

If the hypothesis of R.E.'s is maintained from the beginning, comparable forms for price and income are obtained by solving (2.1), (2.2) and (2.4) to give

$$p_t = \alpha m_t + u_{1t} \tag{2.10}$$

$$y_t = (\beta + \gamma\alpha)\hat{m}_t + \beta\varepsilon_t + u_{2t} \tag{2.11}$$

By comparing (2.8) and (2.9) with (2.10) and (2.11) ignoring differences due both to structural disturbances and money supply innovations (we shall take up these differences later) we see that the essential distinction rests on the estimates a, b and c. Making expectations consistent with the predictions of the model as in (2.8) and (2.9) is not imposing R.E.'s because the structural parameter estimates used therein were obtained under a different expectations hypothesis (in our example the ad hoc scheme in (2.3)). As we show in section 4 below purely extrapolative proxies are always inadequate for providing consistent estimates of the parameters of an R.E. model. Indeed Anderson himself acknowledges this problem although he sees it purely as a problem of identifying the coefficients on the expectational term ('c' in our example);

"... as equation ((2.6)) is usually estimated by time series methods the coefficient 'c' is not identified econometrically ... (and so) one's judgement must come into play". As we shall see however the problem is not strictly one of pure identification, moreover bias is not just restricted to the coefficient on the expectational term. Three key questions arise at this stage. Firstly how serious are the implications of this for multiplier estimates in practical situations? Secondly, are some extrapolative proxies better than others in the sense that (under fairly general circumstances) they result in better multiplier estimates when the method is used? Finally because R.E.'s are generated from past information does the substitution of a current outcome for an R.E. give rise to any peculiar properties during simulation, either desirable or undesirable?

The first of these three are taken up in sections 4 and 5 whilst the third is dealt with in section 6. Before we set about tackling these questions however it will be useful to briefly review extrapolative proxies in common use in existing macroeconometric models.

5.3 Extrapolative proxies in common use

Of the extrapolative proxies typically used in macro modelling we may distinguish those of finite order from those of infinite order. Infinite schemes obviously require a restriction on the lag structures for practical implementation, the exponentially declining lag of adaptive expectations providing a classic example. Finite schemes may or may not be further restricted; in an Almon lag the weights follow a polynomial in terms of the lag operator whereas the scheme in (2.3) allows the data to determine the weights up to the imposed truncation point (n). We consider estimation incorporating each scheme in turn in the context of the income equation (2.2) above. Using the finite but otherwise unrestricted scheme of (2.3) we would substitute (2.3) into (2.2) and estimate freely by O.L.S.

$$y_t = bm_t + c \sum_{i=1}^{n} \delta_i p_{t-i} + u_t \qquad (3.1)$$

The restriction that the δ's sum to one would be imposed afterwards to identify c. The maximum lag n may be chosen 'sufficiently' large to capture the bulk of the distributed lag, leaving the error largely free of autocorrelation although it has often been set equal to one giving a simple lagged variable proxy. The advantage of this procedure lies in allowing the data to determine the form of the distributed lag where cues from the underlying theory are weak. Its weakness lies in the arbitrary restriction required to identify the structural parameter. In dynamic models further such restrictions are required.

Among all extrapolative schemes available those of the Almon lag and adaptive expectations stand out through frequency of use.

Using the former would mean estimating

$$y_t = bm_t + c \sum_{i=1}^{n} \delta(i)\, p_{t-i} + u_t$$

where $\delta(i) = K_0 + K_1 i + K_2 i^2 + \dots K_m i^m$

with additional 'degrees of freedom' requirement that n be greater than $m + 1$ (m is commonly set equal to two giving a quadratic form) and this may be achieved using restricted least squares methods.

Invoking adaptive expectations gives

$$y_t = bm_t + c \left[\frac{(1 - \delta)}{(1 - \delta L)} \right] p_{t-1} + u_t \qquad (3.2)$$

where L is the lag operator.

A Koyck transform is often applied to give

$$y_t = bm_t - b\delta m_{t-1} + c(1-\delta)p_{t-1} + \delta y_{t-1} + (1-\delta L)u_t \qquad (3.3)$$

and this is typically estimated by nonlinear least squares methods incorporating the relevant nonlinear parameter restrictions but approximating the induced first order M.A. error term with an autoregressive error scheme. [16]

The next three sections describe the implications for estimates derived using these proxies when expectations are really rational. In any linear dynamic simultaneous model which includes exogenous variables so generated there always exists a univariate representation for each endogenous variable (Prothero and Wallis, 1976) from which may be obtained a predictor that is unbiased and which has minimum one-step-ahead forecast error variance among the class of purely extrapolative predictors. This optimal extrapolative predictor (henceforth O.E.P.) is to our knowledge, never purposefully used in non-rational models although it may by chance coincide with an ad hoc scheme. It nevertheless provides a useful benchmark in the analysis because, like an R.E., it is a conditional expectation, but it differs from the R.E. by being conditioned on a smaller information set, namely the past values of the variable itself and as a result it is less efficient than the R.E.

5.4 Some analytical results on multiplier biases

We stated in section 2 that using an ad hoc extrapolative proxy in place of the R.E. results in inconsistent parameter estimates. We show this below and then turn our attention to

---

[16] A programme by Osborne recently written would incorporate the M.A. and so provide an exact ML estimation procedure.

the seriousness of this problem for the multiplier estimates.

Consider the general linear model

$$B\underset{\sim}{y}_t + \Gamma\underset{\sim}{x}_t + C\underset{\sim}{y}_t^* = \underset{\sim}{u}_t \tag{4.1}$$

where B is a $g \times g$ matrix of full rank, Γ and C are $g \times k$ and $g \times g$ matrices respectively, y_t is a $g \times 1$ vector of endogenous variables with R.E. y_t^*, x_t is a $k \times 1$ vector of exogenous and predetermined variables (henceforth x-variables) and u_t is a vector of non-autocorrelated structural disturbances.

(4.1) has reduced form

$$\underset{\sim}{y}_t = \Pi_1\underset{\sim}{x}_t + \Pi_2\underset{\sim}{y}_t^* + \underset{\sim}{v}_t \tag{4.2}$$

where $\Pi_1 = -B^{-1}\Gamma$, $\Pi_2 = -B^{-1}C$, $\underset{\sim}{v}_t = B^{-1}\underset{\sim}{u}_t$

and $\underset{\sim}{x}_t = \hat{\underset{\sim}{x}}_t + \underset{\sim}{\varepsilon}_t$

Before we discuss x-variable multipliers we must first draw an important distinction between responses to the anticipated component of the x-variables and their unanticipated components. The latter is the response to shocks or innovations in the exogenous variables and occurs only when current x-variables enter the structure in their own right.

The anticipated components of the x-variables have their effects both through the current x-variables in the structure and through the rational expectation and these effects represent responses to changes in the x-processes themselves. Solving (4.2) gives

$$\underset{\sim}{y}_t = \Pi_1\underset{\sim}{x}_t + \Pi_2(I - \Pi_2)^{-1}\Pi_1\hat{\underset{\sim}{x}}_t + \underset{\sim}{v}_t \tag{4.3}$$

$$= \Pi_1\underset{\sim}{\varepsilon}_t + (I - \Pi_2)^{-1}\hat{\underset{\sim}{x}}_t + \underset{\sim}{v}_t$$

The first of these multiplier components then is

$$\text{(a)} \quad Mu = \left(\frac{\partial y_t}{\partial \varepsilon_t}\right)' = \Pi_1 \tag{4.4}$$

and the second is

$$\text{(b)} \quad Ma = \left(\frac{\partial y_t}{\partial \hat{x}_t}\right)' = (I - \Pi_2)^{-1}\Pi_1 \tag{4.4}$$

Although often not explicitly stated focus falls on the responses to anticipated x-variables (the Ma's). Indeed such were the multipliers that Anderson derives in his deterministic simulations. (There are no shocks and thus no Mu's in deterministic simulations.)

Note that the substitution of an extrapolative proxy (y_t^e) for the R.E. augments the error term in (4.1) by

$$\text{(a)} \quad \underset{\sim}{\xi}_t^{\,s} = C(\underset{\sim}{y}_t^{\,e} - \underset{\sim}{y}_t^*) = C(\underset{\sim}{y}_t^{\,e} - \underset{\sim}{y}_t + \Pi_1\underset{\sim}{\varepsilon}_t + \underset{\sim}{v}_t) \tag{4.5}$$

and in (4.2) by

$$\text{(b)} \quad \underset{\sim}{\xi}_t^{\,r} = \Pi_2(\underset{\sim}{y}_t - \underset{\sim}{y}_t^{\,e} - \Pi_1\underset{\sim}{\varepsilon}_t - \underset{\sim}{v}_t) \tag{4.5}$$

The term in brackets is just the error in the proxy minus that in the R.E. and in general is correlated both with all the exogenous variables in the model and (in general) the proxy itself invalidating standard structural and full information estimation methods. The multiplier biases resulting from this inconsistency have complex uninformative analytical forms even in the simplest of models of the form of (4.1) incorporating the simplest proxies and so some numerical examples are discussed in the next section. However, because of its position as 'best' extrapolative proxy available we examine the case where the proxy coincides with the O.E.P. from the model.

In this case (4.5) becomes

$$\begin{aligned} \underset{\sim}{\xi}_t^{\,s} &= C(\Pi_1 \underset{\sim}{\varepsilon}_t + \underset{\sim}{v}_t - \mu_t) \\ \underset{\sim}{\xi}_t^{\,r} &= \Pi_2(\underset{\sim}{\mu}_t - \Pi_1 \underset{\sim}{\varepsilon}_t - \underset{\sim}{v}_t) \end{aligned} \qquad (4.5')$$

where $\underset{\sim}{\mu}_t$ (equal to $\underset{\sim}{p}_t - \hat{\underset{\sim}{p}}_t$)[17] is the vector of innovations in the univariate representations of the elements of $\underset{\sim}{y}$.

Taking first the case where (4.1) is just identified then simple O.L.S. yields direct estimates of the Π's. Denoting Y and X as $t \times g$ data matrices respectively we now write the estimated equations in stacked form as

$$Y = X\Pi_1' + \hat{Y}\Pi_2' + v = \begin{bmatrix} XY \end{bmatrix} \begin{bmatrix} \Pi_1' \\ \Pi_2' \end{bmatrix} + E$$

where $E = \xi^r + v$ $\left(\xi^r = \begin{bmatrix} \xi_1^{\,r} \ldots \xi_t^{\,r} \end{bmatrix}^{-1}\right.$ and

$$\left. v = \begin{bmatrix} v_1 \ldots v_t \end{bmatrix}' \right)$$

---

[17] A '^' denotes O.E.P.

Now denoting

$$\sigma_{y'y} = \text{plim}T^{-1}(Y'Y) \quad \sigma_{x'y} = \text{plim}T^{-1}(X'Y)$$

we may write asymptotic forms for the biases in Π_1' and Π_2' (B_1 and B_2 respectively) as

$$\begin{bmatrix} B_1 \\ B_2 \end{bmatrix} \quad \begin{bmatrix} \sigma_{x'x} & \sigma_{x'\hat{y}} \\ \sigma_{\hat{y}'x} & \sigma_{\hat{y}'\hat{y}} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{x'E} \\ \sigma_{\hat{y}'E} \end{bmatrix}$$

Now noting that the vector of O.E.P.'s $(\hat{Y})$ is uncorrelated with all current innovations being a function of only past information so that

$$\sigma_{\hat{y}'E} = 0$$

and premultiplying both sides by the R.H.S. moment matrix gives

$$\begin{bmatrix} \sigma_{x'x} & \sigma_{x'\hat{y}} \\ \sigma_{\hat{y}'x} & \sigma_{\hat{y}'\hat{y}} \end{bmatrix} \begin{bmatrix} B_1 \\ B_2 \end{bmatrix} = \begin{bmatrix} \sigma_{x'E} \\ 0 \end{bmatrix}$$

Rearranging the lower g equations gives

$$\sigma_{\hat{y}'\hat{y}}B_2 = -\sigma_{\hat{y}'x}B_1 \tag{4.6}$$

Now note that we can rewrite (4.3) in data matrix form as

$$Y = XMa' + \bar{E} \tag{4.7}$$

where $\bar{E} = v - \varepsilon\Pi_1'\left[I - \Pi_2' \right]^{-1} \Pi_2'$, and $\varepsilon = \left[\varepsilon_1 \ldots \varepsilon_t \right]'$

Premultiplying (4.7) by $\hat{Y}'$ and taking plims gives

$$\sigma_{\hat{y}'\hat{y}} = \sigma_{\hat{y}'x}Ma' + \sigma_{\hat{y}'\bar{E}} \qquad (4.8)$$

Now

$$Y = \hat{y} + \mu \Rightarrow \sigma_{\hat{y}'y} = \sigma_{\hat{y}'\hat{y}} \quad (\text{because } \sigma_{\hat{y}'\mu} = 0)$$

Further

$$\sigma_{\hat{y}'\bar{E}} = 0$$

so that after postmultiplication by B_2 (4.9) becomes

$$\sigma_{\hat{y}'\hat{y}}B_2 = \sigma_{\hat{y}'x}Ma'B_2 \qquad (4.9)$$

Combining (4.9) with (4.6) gives the equality

$$-Ma'B_2 = -B_1 \qquad (4.10)$$

It is easy to show that (4.10) is the unique condition that B_2 and B_1 must satisfy for Ma to be consistently estimated. Denoting the estimate of Ma' as $\hat{M}a'$ we have

$$\hat{M}a' = \left[\Pi_1' + B_1 \right]\left[I - \Pi_2' - B_2 \right]^{-1}$$

so that we require

$$\hat{M}a' = Ma' \Rightarrow \left[\Pi_1' + B_1\right]\left[I - \Pi_2' - B_2\right] = \Pi_1' \left[I - \Pi_2'\right]^{-1}$$

Postmultiplying both sides by $\left[I - \Pi_2' - B_2 \right]$ and rearranging gives

$$\Pi_1' + B_1 = \Pi_1'\left[I - \Pi_2'\right] \left[I - \Pi_2'\right] - \Pi_1'\left[I - \Pi_2'\right] B_2$$

or more succinctly

$$B_1 = -Ma'B_2 \qquad \text{Q.E.D.}$$

We thus have the rather surprising result that although all the estimated coefficients are biased when the proxy coincides with the O.E.P. the multipliers derived during deterministic simulations (the elements of Ma) are consistent. It is slightly more tedious to show that the above proof holds when the Π's are subject to overidentifying restrictions and so this is relegated to appendix A.

This proxy that is 'best' in the sense of minimum one step ahead forecast error variance is thus also 'best' in our current context. It is tempting to assert that the nearer the proxy is to the O.E.P. the less serious is the problem of multiplier bias using the method. This of course is impossible to prove. For one thing we have no rigorous definition of 'nearness'. However, if we define 'nearness' in terms of m.s.e. of prediction then the numerical experiments in the next section appear to bear out our assertion.

We must note that we have considered only asymptotic estimates. In finite samples the variance of the estimates of Ma are likely to be relatively large even when consistent. The choice of instruments for 2SLS estimation of (4.1) will no longer be optimal for example. Also, estimates of Mu obtained in stochastic simulations remain biased.

Finally, we have not considered an important class of models in our analysis namely those containing forward R.E.'s. Solution and analysis of these is difficult in all but the simplest of circumstances. However in Appendix B we prove in an analogous way to that above that our results hold true for the general linear model

$$\underset{\sim}{y}_t = A\underset{\sim}{y}^*_{t+1} + B\underset{\sim}{x}_t + \underset{\sim}{v}_t$$

and x process

$$\underset{\sim}{x}_t = P\underset{\sim}{x}_{t-1} + \underset{\sim}{\varepsilon}_t$$ [18]

[18] Most forward R.E. models may be written in this form and any finite order AR x-process may be written in first order form.

under the assumption that the eigen values of A lie within the unit circle. This assumption 'chooses' the widely implemented 'forward (looking)' solution as the unique stationary one given a stationary x-process.

5.5 Some numerical examples

Having established that when an O.E.P. proxy is incorporated during estimation in place of the R.E. the method yields consistent multiplier estimates in deterministic simulations we turn to examine our assertion that the nearer the proxy is to this 'best' case the less serious is the problem of multiplier bias. To this end we consider the following simple two equation linear macro model explaining a price level and real income.

$$\begin{aligned} &(a) \quad p_t = B_{12}y_t + \alpha p_t^e + u_{1t} \\ &(b) \quad y_t = B_{21}(m_t - p_t) + \gamma g_t + u_{2t} \end{aligned} \qquad (5.1)$$

with reduced form

$$\begin{aligned} &(a) \quad p_t = \Pi_{11}p_t^e + \Pi_{12}m_t + \Pi_{13}g_t + v_{1t} \\ &(b) \quad y_t = \Pi_{21}p_t^e + \Pi_{22}m_t + \Pi_{23}g_t + v_{2t} \end{aligned} \qquad (5.2)$$

where $\Pi_{11} = \frac{\alpha}{\Delta}$, $\Pi_{12} = \frac{B_{21}B_{12}}{\Delta}$, $\Pi_{13} = \frac{\gamma B_{12}}{\Delta}$, $\Pi_{21} = -\frac{B_{21}\sigma}{\Delta}$.

$\Pi_{22} = \frac{B_{21}}{\Delta}$, $\Pi_{23} = \frac{\gamma}{\Delta}$ with $\Delta = 1 + B_{12}B_{21}$

and where g_t is the level of government purchases and m_t is the nominal money stock (all variables in logs).

Specifying R.E.'s for p_t^e and processes for m and g completes the model:

$$m_t = \rho_1 m_{t-1} + \varepsilon_t$$

$$g_t = \rho_2 g_{t-1} + \xi_t$$

$$\text{cov}(\varepsilon_t, \xi_t) = 0$$

We consider this model specifically because it resembles the two main equations in Anderson's condensed version of the St. Louis model. (5.1b) is a static analogue of Anderson's total spending equation (equation 11 in his paper) and (5.2b) of Anderson's 'price equation' (equation 12 in his paper).

Below are tabulated the results of 8 numerical experiments using four sets of parameter values. Taking (5.2) as our data generation with

$$\rho_1 = \rho_2 = \rho$$

(merely a convenient simplification) we derived[19] asymptotic values for multiplier estimates obtained incorporating firstly an O.E.P. proxy and secondly a simpler extrapolative proxy.

[19] Taking the rational model as the data generation, asymptotic values of data moments (σ_{pp}, σ_{py} etc.) involved in 2SLS were derived in terms of the parameters of this data generation (B_{12}, α, σ_ε^2, σ_u^2 etc.) and these expressions were used to generate estimates for a set of model parameter values.

Because the system is overidentified we quote asymptotic values of 2SLS estimates as this would be a typical estimation procedure here. The simple proxy used is that in (3.1) with n set to unity i.e.

$$p_t^e = p_{t-1} \tag{5.3}$$

The O.E.P. is

$$\hat{p}_t = \rho p_{t-1} - \theta \mu_{t-1} \qquad |\theta| < 1 \tag{5.4}$$

where μ is the innovation in the univariate representation of p. (Full derivation of this predictor is given in appendix C.)

In four of the experiments, α assumed the value it took in Anderson's equation (10) namely 0.86. The remaining structural parameters were set to an order of magnitude thought 'reasonable' in the sense that they give rise to sensible multiplier responses. For the remaining four experiments α was chosen to be unity. This was thought interesting since when α is unity money is said to be 'neutral' i.e. a doubling of the money stock doubles prices leaving real output unaffected. Government expenditure is also 'neutral' in this case.

Each of these sets of four subdivide into a low value of the AR parameter (0.5) and a high value (0.9). The rationale for this comes from the form of the O.E.P. Referring to (5.4) it is trivial to show that the m.s.e. of our lagged value proxy differs from that of the O.E.P. by

$$\text{m.s.e. } p^e - \text{m.s.e. } \hat{p} = (\rho - 1)^2 \mathrm{var}(p) + 2(1 - \rho)\theta\sigma^2_\mu$$

The simple proxy (5.3) then is by our criterion close to the O.E.P. for values of ρ close to unity (say 0.9) but not for other values (say 0.5).

The multiplier estimates are tabulated in the table below with their corresponding parameter sets. The multipliers reported are

$$\begin{array}{lllll}
(\mathrm{I}) & M_1 = \dfrac{\partial p_t}{\partial \hat{m}_t} = \dfrac{\Pi_{12}}{1 - \Pi_{11}} \\[2ex]
(\mathrm{II}) & M_2 = \dfrac{\partial p_t}{\partial \hat{g}_t} = \dfrac{\Pi_{13}}{1 - \Pi_{11}} \\[2ex]
(\mathrm{III}) & M_3 = \dfrac{\partial y_t}{\partial \hat{m}_t} = \Pi_{21} Ma^{P_m} + \Pi_{22} \\[2ex]
(\mathrm{IV}) & M_4 = \dfrac{\partial y_t}{\partial \hat{g}_t} = \Pi_{21} Ma^{P_g} + \Pi_{23} \\[2ex]
(\mathrm{V}) & N_1 = \dfrac{\partial p_t}{\partial \varepsilon_t} = \Pi_{12} & & & (5.5) \\[2ex]
(\mathrm{VI}) & N_2 = \dfrac{\partial p_t}{\partial \xi_t} = \Pi_{14} \\[2ex]
(\mathrm{VII}) & N_3 = \dfrac{\partial y_t}{\partial \varepsilon_t} = \Pi_{22} \\[2ex]
(\mathrm{VIII}) & N_4 = \dfrac{\partial y_t}{\partial \xi_t} = \Pi_{23}
\end{array}$$

(5.5)(I) to (5.5)(IV) are the responses of p and y respectively to anticipated policy and (5.5)(V) to (5.5)(VIII) their stochastic policy response counterparts.

Referring to the table we see that when (by our criterion) the simple proxy is close to the O.E.P. the Ma multiplier estimates are nearly consistent biases averaging around 10%. This contrasts starkly with the cases where the proxy differs markedly from the O.E.P. ($\rho = 0.5$). Biases of between 50% and 70% are reported with one multiplier estimated at nine times its actual value.

These results bear out our assertion that the closer the proxy is to the O.E.P. the less the multiplier bias arising from the method in deterministic simulations. We must draw attention however to the results on the Mu's. All of these are seriously biased even when the O.E.P. is used. The method clearly does not give good estimates of responses to stochastic disturbances.

The conclusions drawn from our numerical experiments point towards using the method in deterministic simulations when it is reasoned that the proxy incorporated during estimation is close to the O.E.P. from the model. These conclusions are of course at best tentative as are any results derived from specific numerical experiments.

Parameter	Actual value		O.E.P. estimate		Simple proxy estimate	
B_{12}	0.5	0.5	4.4803	1.6659	4.525	1.6717
α	1.0	1.0	0.9367	0.9993	0.8552	0.3293
B_{21}	0.25	0.25	0.25	0.25	0.25	0.25
γ	0.25	0.25	0.25	0.25	0.25	0.25
Π_{11}	0.88	0.88	0.4418	0.7055	0.4013	0.2323
Π_{12}	0.11	0.11	0.5582	0.294	0.5308	0.2948
Π_{13}	0.11	0.11	0.5582	0.294	0.5308	0.2948
Π_{21}	-0.22	0.22	-0.1105	-0.1764	-0.1003	-0.0581
Π_{22}	0.22	0.22	0.1179	0.1765	0.1173	0.1763
Π_{23}	0.22	0.22	0.1179	0.1765	0.1173	0.1763
M_1	1.0	1.0	1.0	1.0	0.8866	0.384
M_2	1.0	1.0	1.0	1.0	0.8866	0.384
M_3	0.0	0.0	0.0	0.0	0.0284	0.154
M_4	0.0	0.0	0.0	0.0	0.0284	0.154
N_1	0.11	0.11	0.5582	0.294	0.5308	0.2948
N_2	0.11	0.11	0.5582	0.294	0.5308	0.2948
N_3	0.22	0.22	0.1179	0.1765	0.1173	0.1763
N_4	0.22	0.22	0.1179	0.1765	0.1173	0.1763
ρ	0.9	0.5	0.9	0.5	0.9	0.5
	SET(1)	SET(2)	SET(1)	SET(2)	SET(1)	SET(2)

Parameter	Actual value		O.E.P. estimate		Simple proxy estimate	
B_{12}	0.6	0.6	1.1275	1.9984	1.1534	2.2073
α	0.86	0.86	0.7387	0.5714	0.1516	0.3674
B_{21}	0.2	0.2	0.2	0.2	0.2	0.2
γ	0.25	0.25	0.25	0.25	0.25	0.25
Π_{11}	0.7679	0.7679	0.6028	0.4082	0.1232	0.2549
Π_{12}	0.1071	0.1071	0.184	0.2732	0.1874	0.3063
Π_{13}	0.1339	0.1339	0.23	0.3414	0.2343	0.3828
Π_{21}	-0.1536	0.1536	-0.1206	-0.0817	-0.0346	-0.0051
Π_{22}	0.1786	0.1786	0.1632	0.1429	0.9857	0.1387
Π_{23}	0.2232	0.2232	0.204	0.1786	0.2031	0.1734
M_1	0.4614	0.4614	0.4614	0.4614	0.2137	0.411
M_2	0.5769	0.5769	0.5769	0.5769	0.2672	0.5138
M_3	0.1077	0.1077	0.1077	0.1077	0.9783	0.1177
M_4	0.1346	0.1346	0.1346	0.1346	0.1939	0.1472
N_1	0.1071	0.1071	0.184	0.2732	0.1874	0.3063
N_2	0.1339	0.1339	0.23	0.3414	0.2343	0.3828
N_3	0.1786	0.1786	0.1632	0.1429	0.9857	0.1387
N_4	0.2232	0.2232	0.204	0.1786	0.2031	0.1734
ρ	0.5	0.9	0.5	0.9	0.5	0.9
	SET(3)	SET(4)	SET(3)	SET(4)	SET(3)	SET(4)

5.6 Further properties of the method

The preceding sections show that to impose R.E. during simulation may require a set of parameter estimates obtained under the same maintained hypothesis, since the set obtained using an ad hoc expectations hypothesis may be inadequate. This section exposes further properties of the method assuming that a set of parameter estimates has been obtained under the correct hypothesis of R.E.

Properties associated with Lucas' critique

Because an R.E. differs from the actual outcome only by an innovation uncorrelated with the 'past' as contained in the information set then substitution of the actual value for the R.E. will ensure immunity from the structural variation noted by Lucas (1976), since no varying parameters associated with optimal prediction enter the simulation. Indeed this was very much the motivation for the method.

To make this clear recall the model as in (5.2)

$$\begin{aligned} &(a) \quad p_t = \Pi_{11} p_t^e + \Pi_{12} m_t + \Pi_{13} g_t + v_{1t} \\ &(b) \quad y_t = \Pi_{21} p_t^e + \Pi_{22} m_t + \Pi_{23} g_t + v_{2t} \end{aligned} \tag{5.2}$$

Specifying R.E.'s and general AR MA (p, q) processes for m and g gives

$$p_t^{e} = p_t^{*} = \frac{\Pi_{12}}{1 - \Pi_{11}} \hat{m}_t + \frac{\Pi_{13}}{1 - \Pi_{11}} \hat{g}_t \qquad (6.1)$$

$$= \frac{\Pi_{12}}{1 - \Pi_{11}} L^{-1} \left[1 - \frac{\phi_1(L)}{\theta_1(L)} \right] m_{t-1} + \frac{\Pi_{13}}{1 - \Pi_{11}} L^{-1} \left[1 - \frac{\phi_2(L)}{\theta_1(L)} \right] g_{t-1}$$

where $\phi_i(L)$ $i = 1,2$ are the AR(p) lag polynomials

and $\theta_i(L)$ $i = 1,2$ the MA(q) lag polynomials on m and y respectively.

Substitution of (7.1) into (5.2a) and (5.2b) gives a reduced form in terms of observables which includes the parameters of the x-processes (the ϕ's and the θ's). Changes in the latter bring about changes in this reduced form and these must be accounted for when such changes are simulated. Making expectations consistent with the predictions of the model we would simulate

$$\text{(a)} \quad p_t^{s} = \frac{\Pi_{12}}{1 - \Pi_{11}} m_t^{s} + \frac{\Pi_{13}}{1 - \Pi_{11}} g_t^{s} = Ma^{p_m} m_t^{s} + Ma^{p_g} g_t^{s} \text{ [18]}$$

$$\text{(b)} \quad y_t^{s} = \left[\frac{\Pi_{21}\,\Pi_{12}}{1 - \Pi_{11}} + \Pi_{22} \right] m_t^{s} + \left[\frac{\Pi_{21}\,\Pi_{13}}{1 - \Pi_{11}} + \Pi_{23} \right] g_t^{s} \qquad (6.2)$$

$$= Ma^{y_m} m_t^{s} + Ma^{y_g} g_t^{s}$$

and changes in the x-processes do not raise a problem during simulation. In effect then the method keeps separate the parameters of the x-processes from those of the economic structure (a distinction drawn by Wallis (1980)) allowing policy simulation to proceed in the traditional fashion. This feature of the method combined with the removal of the need to solve the model to obtain an expression for the R.E. form its most attractive properties.

[18] $Ma^{y_x} = \frac{\partial y_t}{\partial \hat{x}_t}$

Properties associated with deterministic simulation

The key element of the method is to make expectations consistent with actual outcomes. At first sight this seems like imposing not R.E.'s but perfect foresight. However properly executed deterministic simulations yield paths for the endogenous variables that are themselves expectations (conditioned not on the previous period's information but on the information available at the first period of simulation). Making expectations consistent with outcomes in this deterministic setting then is an exact procedure in the R.E. context. To make this clear consider simulating our model in section 5. Simulated paths using the method must satisfy (6.2). However, the model's actual solution is

$$\begin{aligned} &(a) \quad p_t = Ma^{p_m}\hat{m}_t + Mu^{p_m}\varepsilon_t + Ma^{p_g}\hat{g}_t + Mu^{p_g}\xi_t + v_{1t} \\ &(b) \quad y_t = Ma^{y_m}\hat{m}_t + Mu^{y_m}\varepsilon_t + Ma^{y_g}\hat{g}_t + Mu^{y_g}\xi_t + v_{2t} \end{aligned} \qquad (6.3)$$

Now taking expectations of both sides conditioned on the information available at the first period of simulation (Ω_0) gives

$$\begin{aligned} &(a) \quad E(p_t|\Omega_0) = Ma^{p_m}E(m_t|\Omega_0) + Ma^{p_g}E(g_t|\Omega_0) \\ &(b) \quad E(y_t|\Omega_0) = Ma^{y_m}E(m_t|\Omega_0) + Ma^{y_g}E(g_t|\Omega_0) \end{aligned} \qquad (6.4)$$

Comparing (6.4) with (6.2) we see that in order to provide a path for the endogenous variables conditioned on Ω_0 (as a deterministic simulation should) we require the x inputs to be

$$m_t^s = E(m_t|\Omega_0) \tag{6.5}$$

and $$g_t^s = E(g_t|\Omega_0)$$

Using actual x's instead of those based on Ω_0 allows unforeseeable exogenous events to be foreseen. It is unfortunate that this latter option is much more convenient than (6.5).

Stochastic simulation and problems associated with perfect foresight in neutrality models

There are certain classes of model for which mean paths obtained from deterministic simulations, are very uninteresting and uninformative from a policy analysis point of view. For such models structural disturbances must be added to all the equations and actual exogenous variable values (as opposed to predicted or mean values) used i.e. stochastic simulation must be undertaken. In this situation setting outcomes equal to expectations is imposing not R.E.'s but perfect foresight since outcomes include all current stochastic disturbances.

One class of models that require stochastic simulation and for which the imposition of perfect foresight instead of R.E.'s has very serious consequences is the so called neutrality class of Lucas (1972) and Sargent and Wallace (1975). Broadly speaking these models are built to give a steady state in which the level of real output is independent of the money supply. Typically they have as a central feature a Lucas supply function determining output, and a quantity theory equation or a demand for money equation determining the price level. The former has real output deviating from a natural level only through the effect of current and past errors in predicting the price level. If expectations are rational these errors are unpredictable and consist of innovations in the exogenous variables and in the structure, both being uncorrelated with past events contained in the information set. In effect then mistakes in expectations consisting of these innovations drive a trade cycle, and if they are ignored by invoking perfect foresight then output will not deviate from its natural level for the period of simulation. As an illustration consider the simple neutrality model

$$y_t = y^n + \sum_{i=0}^{m} \gamma_i (p_{t-i} - p^*_{t-i}) + u_{1t} \qquad (6.6)$$

$$p_t = V + m_t - y_t + u_{2t} \qquad (6.7)$$

where y^n is a natural level of output. The reduced form is

$$y_t = y^n + (1 + \gamma_0)^{-1} \sum_{i=0}^{m} \gamma_i (\varepsilon_{t-i} + u_{2t} - u_{1t}) + u_{1t} \qquad (6.8)$$

$$p_t = V + m_t - y^n - (1 + \gamma_0)^{-1} \sum_{i=0}^{m} \gamma_i (\varepsilon_{t-i} + u_{2t} - u_{1t}) + u_{2t} - u_{1t} \qquad (6.9)$$

where ε_t is, as above, the current innovation in the money supply. Deterministic simulation of (6.6) and (6.7) or (6.8) and (6.9) using the method would yield solution

$$y_t^s = y^n$$

$$p_t^s = V + m_t^s - y^n$$

This is clearly useless for analysis of the short run behaviour of say, output and for this end we require stochastic simulation. However, making expectations 'consistent' in a stochastic simulation yields the solution

$$y_t^s = y^n + u_{1t}^s$$

$$p_t^s = V + m_t^s - y^n + u_{2t}^s - u_{1t}^s$$

Again this is totally uninformative about short run fluctuations in output, now because of the perfect foresight problem. Clearly we must reintroduce 'mistakes' in expectations in the simulation procedure. A more natural substitution of the R.E. comes from the identity

$$p_t^* \equiv p_t - \eta_t \qquad (6.10)$$

where η_t is the error in the R.E. In our simple model the substitution derived from (7.8) and (7.9)[19] is

[19] In linear models calculation of the error in the R.E. is relatively easy. In nonlinear models, however, a linear approximation may have to be taken first.

$$p_t^* = p_t^s - \mu_t^s = p_t^s - \varepsilon_t^s + \gamma_0(1 + \gamma_0)^{-1}(\varepsilon_t^s + u_{2t}^s - u_{1t}^s)$$

Because this substitution is exact, traditional policy evaluation may now proceed. It may not be feasible or easy however to obtain such a neat analytical form for the error in the R.E. We therefore propose a further suggestion which is exact in linear models but only an approximation[20] in nonlinear models. Undertake a deterministic solution using the method but this time instead of using (6.5) for the x variables use

$$\text{(a)} \quad m_t^s = E(m_t | t - 1) \qquad\qquad (6.11)$$

$$\text{(b)} \quad g_t^s = E(g_t | t - 1)$$

The simulated paths for p and y are now approximately $E(p_t | t - 1)$ and $(y_t | t - 1)$. Since these are just the R.E.'s p_t^* and y_t^* they may be used as an expectations series in a second simulation which adds stochastic disturbances (v_{it}'s) to each equation and uses actual exogenous outcomes, that is in a second stage full stochastic simulation. If the model is dynamic care must be taken to add to all predetermined variables their respective reduced form errors in the first stage. (These are the errors that are generated and added to the equations in the second stage, stochastic simulation.)

---

[20] It is well known that replacing exogenous and predetermined variables in nonlinear models with their expectations and solving does not yield exact expected values of the endogenous variables.

This has to be done because predetermined variables (and not their expectations) are contained in the information set upon which the R.E.'s are based. Again, obtaining a series as an approximation to the R.E. in this way allows policy evaluation to proceed in a traditional manner.

5.7 Summary and Conclusion

The analysis in section 4 and the numerical experiments in section 5 have led us to conclude that providing that the extrapolative proxies incorporated in existing models are close to the O.E.P. derived from the model's rational counterpart then the method will yield approximately consistent multiplier estimates in deterministic simulations, performed in the manner outlined in section 6.

Unfortunately a glance at our review of expectation proxies in common usage shows that the rather ad hoc mechanisms typically incorporated into existing macro models at the estimation stage may not correspond closely with the O.E.P. from the rational model. Where this is the case then our numerical experiments in section 5 warn of severe multiplier biases.

The simulation concept itself however is very useful provided that 'mistakes' in expectations are reintroduced into the simulation procedure where stochastic simulation is required. This latter point is especially relevant for policy analysis in so called neutrality models where deterministic simulation is uninformative about the short run behaviour of the endogenous variables.

Finally, the most attractive feature of the algorithm is that which motivated it in the first place, namely that it allows policy evaluation to proceed in a traditional manner. Making expectations consistent with the predictions of a macro model has the effect of separating the economic structure from the 'structure' of the exogenous processes and thus removes the source of the structural variation referred to by Lucas, namely these processes themselves.

APPENDIX A

Proof of multiplier consistency when a structure is overidentified.

Consider the general overidentified model written in data matrix form

$$Y\bar{\beta} = Y^{*}\alpha + X\Gamma + u$$

where α, $\bar{\beta}$ are $g \times g$ and Γ is a $k \times g$ parameter matrix (all other symbols are as in the main text).

We may rewrite this as

$$Y = Y\beta + Y^{*}\alpha + X\Gamma + u \qquad \text{(A1)}$$

where $\beta = I - \bar{\beta}$ has zeros on its diagonal.

The Ma matrix is

$$Ma' = \Gamma\Big[I - \beta - \alpha\Big]^{-1}$$

Consistency of Ma requires that the biases in 2SLS estimates of β, α and Γ (B_β, B_α and B_Γ) obey the relation

$$(Ma' =)\ \Gamma\Big[I - \beta - \alpha\Big]^{-1} = \Big[\Gamma + B_\Gamma\Big]\Big[I - \beta - \gamma - B_\beta - B_\alpha\Big]^{-1}$$

and tedious manipulation reduces this to

$$Ma'\Big[B_\beta + B_\alpha\Big] = -B_\Gamma \qquad \text{(A2)}$$

TSLS estimation of

$$Y = Y\beta + \hat{Y}\alpha + X\Gamma + z$$

where $z = u + \alpha\Big[\mu - (Y - Y^{*})\Big]$

gives the biases as

$$\begin{bmatrix} B_\beta \\ B_\alpha \\ B_\Gamma \end{bmatrix} = \begin{bmatrix} \sigma_{\hat{\hat{y}}'y} & \sigma_{\hat{\hat{y}}'\hat{y}} & \sigma_{\hat{\hat{y}}'x} \\ \sigma_{\hat{y}'\hat{\hat{y}}} & \sigma_{\hat{y}'\hat{y}} & \sigma_{\hat{y}'x} \\ \sigma_{x'\hat{\hat{y}}} & \sigma_{x'\hat{y}} & \sigma_{x'x} \end{bmatrix}^{-1} \begin{bmatrix} \sigma_{\hat{\hat{y}}'z} \\ \sigma_{\hat{y}'z} \\ \sigma_{x'z} \end{bmatrix} \qquad \text{(A3)}$$

where a '$\hat{\hat{\ }}$' denotes unrestricted reduced form predictions from the first stage of 2SLS as opposed to a '$\hat{\ }$' which again denotes O.E.P.

Premultiplying both sides of (A3) by the R.H.S. matrix, noting that

$$\sigma_{\hat{y}'z} = 0,$$

and rearranging the second block of equations gives

$$\sigma_{\hat{y}'\hat{\hat{y}}}B_\beta + \sigma_{\hat{y}'\hat{y}}B_\alpha + \sigma_{\hat{y}'x}B_\Gamma = 0 \qquad \text{(A4)}$$

From the properties of 2SLS we have

$$\sigma_{\hat{y}'\hat{\hat{y}}} = \sigma_{\hat{y}'\hat{y}}$$

Using this in (A4) and rearranging gives

$$\sigma_{\hat{y}'\hat{y}}\left[B_\beta + B_\alpha\right] = -\sigma_{\hat{y}'x}B_\Gamma \qquad \text{(A5)}$$

To show that (A5) is equivalent to (A2) rewrite (A1) as

$$Y = XMa + u - \alpha E$$

where $E = \left[u + (Y - Y^*)\right]\left[I - \beta - \alpha\right]^{-1}$

Premultiplying both sides by $\hat{Y}'$ and taking plims gives

$$(\sigma_{\hat{y}'y} =)\ \sigma_{\hat{y}'\hat{y}} = \sigma_{\hat{y}'x}Ma \tag{A6}$$

Postmultiplying (A6) by $\left[B_\beta + B_\alpha \right]$ and combining with (A5) gives

$$-B_\Gamma = Ma \left[B_\beta + B_\alpha \right]$$

which is just our consistency condition (A2)

Q.E.D.

APPENDIX B

Forward Expectations

Consider the general linear model

$$Y = Y^*_{+1}A' + XB' + v^{(1)} \tag{B1}$$

with x-process

$$X = X_{-1}P' + \varepsilon$$

On the assumption that all the eigen values of A lie inside the unit circle . the forward solution is uniquely stationary. A general form for this solution may be written as

$$Y = X_{-1}F' + XB' + v \tag{B2}$$

To get a form for F advance (B2) by one period and take expectations conditional on t-1 to get

$$Y^*_{+1} = X_{-1}\left[FP\right]' + X_{-1}\left[BP^2\right]' \tag{B3}$$

Substituting (B3) into (B1) gives

$$Y = X_{-1}\left[AFP + ABP^2\right]' + XB' + v \tag{B4}$$

---

(1) $Y_{+1} = \begin{bmatrix} y_1 \\ \vdots \\ y_t \\ y_{t+1} \end{bmatrix}$ and $\sigma_{y+1'y} = \text{plim } Y_{+1}'Y$

Equating coefficients in (B2) and (B4) gives

$$F = AFP + ABP^2$$

Vectorising both sides, applying the rule

$$\mathrm{vec}XYZ = (Z' \otimes X)\mathrm{vec}Y \tag{B5}$$

(for Z = I) and rearranging gives an explicit form for vecF as

$$\mathrm{vec}F = \left[I - P' \otimes A \right]^{-1} \mathrm{vec}ABP^2 \tag{B6}$$

Turning now to the multipliers in this model Ma is now

$$Ma = F + BP \tag{B7}$$

Taking vecs of (B7) applying the rule (B5) and using (B6) gives

$$\begin{aligned} \mathrm{vec}Ma &= \left[I - P' \otimes A \right]^{-1} \mathrm{vec}ABP^2 + \mathrm{vec}BP \\ &= \left\{ \left[I - P' \otimes A \right] (P' \otimes A) + I \right\} \mathrm{vec}BP \\ \mathrm{vec}Ma &= \left[I - P' \otimes A \right]^{-1} \mathrm{vec}BP \end{aligned} \tag{B7}$$

Once again the matrices of biases in A' and B' (B_A and B_B) that arise using an O.E.P. in place of the R.E. must satisfy

$$\begin{bmatrix} \sigma_{\hat{y}_{+1}'\hat{y}_{+1}} & \sigma_{\hat{y}_{+1}'x} \\ \sigma_{x'\hat{y}_{+1}} & \sigma_{x'x} \end{bmatrix} \begin{bmatrix} B_A \\ B_B \end{bmatrix} = \begin{bmatrix} 0 \\ \sigma_{x'E} \end{bmatrix}$$

where $E = A(\mu - \varepsilon P'B' - \varepsilon_{+1}B')$

Rearranging the upper equation block gives

$$\sigma_{\hat{y}'\hat{y}} B_A = -\sigma_{\hat{y}_{+1}'x} B_B \tag{B8}$$

Now, following earlier procedures we derive a form for Y_{+1} in terms of Ma. Advancing (B2) by one period gives

$$\begin{aligned} Y_{+1} &= XF' + X_{+1}B' + v \\ &= XF' + XP'B' + v + \varepsilon \\ Y_{+1} &= XMa' + v + \varepsilon \end{aligned} \tag{B9}$$

It follows straightforwardly that

$$\sigma_{\hat{y}_{+1}'\hat{y}_{+1}} = \sigma_{\hat{y}_{+1}'x} Ma' \tag{B10}$$

Combining (B8) and (B10) gives

$$\sigma_{\hat{y}_{+1}'x} Ma'B_A = -\sigma_{\hat{y}_{+1}'x} B_B \tag{B11}$$

Now we must verify that condition (B11) guarantees a consistent estimate of Ma.

Recalling (B7) the condition we require is

$$\left[I - P' \otimes A - P' \otimes B_A'\right]^{-1} \text{vec}\left[BP + B_B'P\right] = \left[I - P' \otimes A\right]^{-1} \text{vec}BP$$

Premultiplying both sides by the L.H.S. matrix gives

$$\text{vec}\left[BP + B_B'P\right] = \text{vec}BP - (P \otimes B_A')\left[I - P' \otimes A\right]^{-1} \text{vec}BP$$

Cancelling terms gives

$$\mathrm{vec}B_B{}'P = -(P' \otimes B_A{}')\mathrm{vec}Ma$$

Premultiplying by $(P'^{-1} \otimes I)$ and using (B5) gives

$$\mathrm{vec}B_B{}' = -(I \otimes B_A{}')\mathrm{vec}Ma$$

$$\mathrm{vec}B_B{}' = \mathrm{vec}B_A{}'Ma$$

so that we have

$$B_B = -Ma'B_A$$

and this verifies (B11)

Q.E.D.

APPENDIX C

Derivation of the O.E.P. for p_t in (5.2)

Solving (5.2a) gives an 'observable' equation

$$p_t = \frac{\Pi_{11}\Pi_{12}\rho}{1 - \Pi_{11}} m_{t-1} + \frac{\Pi_{11}\Pi_{13}\rho}{1 - \Pi_{11}} g_{t-1} + \Pi_{12}m_t + \Pi_{13}g_t + v_{1t}$$

To derive a purely autoregressive predictor we must substitute out m_t, m_{t-1}, g_t and g_{t-1} to get

$$p_t = \left[\frac{\Pi_{11}\Pi_{12}\rho}{1 - \Pi_{11}}\right] \varepsilon_{t-1} + \left[\frac{\Pi_{11}\Pi_{13}\rho}{1 - \Pi_{11}}\right] \xi_{t-1} + \Pi_{12}\varepsilon_t + \Pi_{13}\xi_t \; \frac{1}{1-\rho L} + v_{1t}$$

Multiplying throughout by $1 - \rho L$ gives

$$p_t = \left[\frac{\Pi_{11}\Pi_{12}\rho}{1 - \Pi_{11}} \cdots\right] + v_{1t} - \rho v_{1t-1} + \rho p_{t-1} \tag{C1}$$

The R.H.S. of (C1) is the sum of two independent MA (1) processes and thus has a representation in terms of a single innovation μ_t giving

$$p_t = \rho p_{t-1} + \mu_t - \theta\mu_{t-1} \tag{C2}$$

where $|\theta| < 1$ and where θ and the variance of μ_t are found by means of the cannonical factorisation of (C1). The O.E.P. in (5.4) follows trivially from (C2)

6.1 Existing Methods

The first and some would say most important step in policy analysis is estimation. In the R.E. arena this topic is most controversial. Academic argument on the matter appears to be polarised into two schools. The first regard FIML as an appropriate and indeed imperative tool for estimation and inference whilst the second school argue that the costs of this approach outweigh the benefits and put forward limited information methods. Because these methods are well known and documented we comment only briefly on them.

Broadly speaking there are two approaches to limited information estimation. The first proposes that actual values be substituted for expections and then the problem be treated as one of errors in variables (see for example, Wickens (1980)). The second proposes that expectations terms be proxied by fitted values from a regression of their corresponding actual values on all variables in the information set (see McCallum (1976) for more details of this). Both of these approaches have dominated econometric practice to date and it's not difficult to understand why. Using existing software consistent estimates are obtained simply, conveniently and most important of all, cheaply. Limited information methods do not however provide any framework at all within which a test of the R.E. hypothesis may be made. The method simply does not impose any testable restrictions that the R.E. hypothesis provides and so no inference may be made as to the validity of the expectations mechanism vis a vis some alternative. Such methods do of course provide the usual t-ratio tests on the significance of structural coefficients but such a test applied to coefficients of expectations terms has no clear interpretation. As an illustration, consider the simple model

$$y_t = \alpha y^e_{t+1|t-1} + Bx_t + u_t \qquad (1.1)$$

with $\quad x_t = \rho x_{t-1} + \varepsilon_t$ (1.2)

The t-test on α in (1.1) has the rather uninteresting null hypothesis

$$y_t = Bx_t + u_t$$

Further, suppose that the true model were (1.1) but with $y^e_{t+1|t-1}$ replaced with the naive expectations scheme of a lagged variable so that the true model is

$$y_t = \alpha y_{t-1} + Bx_t + u_t$$

Using (1.2) we may rewrite the true model as

$$y_t = 1/\alpha\ y_{t+1} - B/\alpha\ \rho x_t - B/\alpha\varepsilon_{t+1} - 1/\alpha\ u_{t+1} \tag{1.3}$$

Now because all variables in the information set (x_{t-1-i} and y_{t-1-i}, $i \geq 0$) are orthogonal to the composite error term in (1.3) a Wickens type e.v.m or a McCallum type procedure would yield consistent estimates of $-B\rho/\alpha$ rather than of α and B respectively. In any case the significance of the observed coefficient on the expectations term can clearly not be treated as a verification of the R.E. hypothesis in (1.1). Thus, to undertake limited information estimation the investigator must have total faith in the R.E. hypothesis since no meaningful diagnostic tests are available in this context.

Until very recently investigators wishing to undertake FIML estimation of models incorporating forward R.E. have had no software to help them. Faced with this they have confined their attention to simple models with analytically tractable parameter restrictions obtained by imposing the forward solution on the model. Then, using some numerical differentiation routines or packages have found the point in the structural parameter space that maximises the likelihood function (see Burmeister and Wall (1983) for an example of such empirical work). In their recent paper Fair and Taylor describe and use an algorithm that delivers exact FIML estimates of the parameters of a general linear R.E. model. For full details the reader is referred to Fair and Taylor (1983). Essentially the routine mimics the FIML procedure described above except the model solution is executed numerically rather than analytically: Given a set of parameters, expected exogenous variable paths and a sufficient set of terminal values for the endogenous variables the program sets all stochastic disturbances to their means (usually zero) and then iterates on endogenous variable paths until convergence is reached between these and their corresponding expectations in the model. By allowing calculation of the L.F. this expectations solution then forms the basis for its numerical maximisation with respect to the parameters of the model.

In the context of the multiple solutions debate the algorithm appears to be neat because terminal values for the endogenous variables which are normally set to identify a solution are not required. In fact the algorithm chooses an integer 'N' such that changes in the terminal values (chosen arbitrarily) for periods beyond t+N to solve uniquely for the endogenous variables in periods t+i,i=1,N, do not significantly affect these latter solution values. In simpler terms the solution that is chosen by this procedure is one where expectations far into the future have negligible influence on current outcomes (and therefore on current expectations). It was shown in chapter 4 that terminal conditions were needed to determine a unique solution. Therefore a solution chosen so that the terminal conditions dont 'bite' should raise our suspicions to say the least.

Consider as an illustration the simple model in (1.1) and (1.2). If $|\alpha|>|$ then solving backwards from a terminal conditon at t+n $(y^e_{t+n}|_{t-1})$ gives us

$$y_t^e|_{t-1} = \alpha^n y^e_{t+n}|_{t-1} + B\sum_{i=1}^{N}\alpha^i x^e_{t+n-i}|_{t-1}$$

Obviously the solution for the current value $y^e_{t+n-i}|_{t-1}$ is extremely sensitive to the choice of terminal condition y^e_{t+n} and extending the time horizon only makes matters worse. The programme would fail in this instance in that it would not find a solution satisfying its own criteria; the solution would most definitely depend on the terminal value $y^e_{t+n}|_{t-1}$. Note however that if $|\alpha|<|$ then the algorithm chooses the only stationary solution in this case namely the forward solution with n chosen sufficiently large to obliterate the effect of the terminal condition $(y^e_{t+n}|_{t-1})$.

Despite this shortcoming the programme does have the important property of being general enough to handle any linear or nonlinear model. However the computional cost it implies for even the smallest of models is likely to be inhibitive. For example in their paper Fair and Taylor use the programme to estimate the parameters of a model. To this end 28,000 passes through the model are required for convergence of FIML estimates. (A pass through the model consists of one Gauss-Seidel iteration).

The purpose of this chapter is to show that for linear models at least it is possible to use standard analysis to provide closed forms for the first order conditions for maximisation of the likelihood function without having to impose a solution prior to estimation. These first order conditions can be written in a general form to cover virtually all linear R.E. models so that they may be programmed and then used to solve for the M.L.E.'s by simple numerical methods.

Before we proceed we must pay heed to an alternative solution and estimation technique advanced by Chow (1981). Again for full details the reader is referred to the paper itself.

To obtain a solution Chow employs identities of the form

$$y_t^e|_{t-1} \equiv y_t + \xi_t$$
$$y_{t+1}^e|_{t-1} \equiv y_{t+1} + (1-CL)\xi_{t+1}$$

Then substituting out for the R.E. terms in a model gives a structure in terms of observables with M.A. error processes.

$$y_t = \alpha y_{t-1} + Bx_t + C_0\xi_{t+1} + C_1\xi_t + u_t$$

The suggestion then is to estimate this structure by standard maximum likelihood methods. It is claimed that this procedure does not require any knowledge of the exogenous variable processes but more importantly that it allows the data to reveal the form of the solution rather than it being imposed a priori. Closer examination of the method however reveals that it does not reveal the form of the solution and so its use for policy analysis is curtailed. This is not surprising considering that the method does not incorporate information from the exogenous variable processes. Furthermore, it does not allow us to impose the forward solution during estimation. This is important where the latter is suggested by the model's theoretical structure. For example, the model of Hall (1978) has a priori a 'unique stationery solution' (or 'saddlepoint') property and fully efficient estimation here would require imposing the restrictions of the forward solution only.

The method we propose below may be used in all cases regardless of whether or not the model solution is determinate.

6.2 An Alternative Approach to FIML Estimation

In this section we attempt to derive analytical forms for the first order conditions (henceforth F.O.C.) of the likelihood function of a gx1 vector of endogenous variables y_t with respect to the structural parameters of the model governing y_t written as [23]

$$Cy_t = \Omega y^e_{t+1} + \Gamma x_t + \Delta y_{t-1} + u_t \qquad (2.1)$$

where C, Ω and Δ are gxg matrices, Γ is a gxk matrix (possibly subject to exclusion restrictions) and u_t is a gx1 vector of normal structural disturbances where

$$E(u_t u_t') = S$$

It is assumed that Γ and the vector of exogenous variables can be written in such a way as to be able to represent the x-processes jointly as

$$x_t = Px_{t-1} + \varepsilon_t \qquad (2.2)$$

where ε_t is a kx1 vector of disturbances independent of u_t $\forall_t$.

We assume for the time being that Ω is of full rank. This assumption is quite severe in that it seriously restricts the number of models that can be written in the form of (2.1) and so we relax this at the end of the

---

[23] From now on superscript 'e' unambiguously denotes a R.E. conditional on Ω_{t-1}.

section where we consider how to overcome the complications raised by Ω being singular Another respect in which (2.1) is not perfectly general is that it doesn't deal with expectations horizons of two or more periods (y^e_{t+1},i>0). Whilst we believe most models conform to (2.1) this case is dealt with also at the end of the section.

We may write the quasi reduced form of (2.1) as

$$y_t = Ay^e_{t+1} + Bx_t + Dy_{t-1} + v_t \tag{2.3}$$

where $A = C^{-1}\Omega$, etc.

Following chapter 4 the exhaustive set of solutions to (2.3) may be parameterised by the gxg matrix

$$\Lambda = \begin{bmatrix} \lambda_1 & & 0 \\ & \lambda_2 & \\ 0 & & \lambda_g \end{bmatrix}$$

as follows

$$y_t = \Lambda y_t^0 + [I-\Lambda]\, y_t^1 \tag{2.4}$$

where y_t^0 and y_t^1 are two distinct solutions to (2.3). The solutions we choose for y_t^0 and y_t^1 are the intuitively appealing and empirically popular 'forward' or 'forward looking' solution (y_t^F) and a 'backward' or 'backward looking' solution (y_t^B). Analytical forms for the latter are easy to come by and are well documented in the literature (see for example Pagan (1982)). It is the former however that has been overwhelmingly and almost exclusively popular in the empirical literature but the absence of a general form for this solution for a model such as (2.1) has seriously restricted FIML estimation to simple one or two equation models such as that of Burmeister and Wall (1983). As an example of this difficulty we quote Wallis who in his discussion of FIML estimation of models with forward R.E. states that such estimation is possible

> "provided that the parameter restrictions resulting from the dependence of the R.E. on the predictions of all future values of the exogenous variables .... can be conveniently expressed"

(Wallis (1980))

The main force and novelty of this chapter then is the derivation of a neat analytical representation of the restrictions implied by the forward solution. This, together with the backward solution allows us to analytically derive and then solve the F.O.C. of the likelihood function of the y_t and so to obtain M.L.E.'s of the structural parameters in C, Ω, Γ, Δ and those describing the solution (the λ's in Λ).

We begin then by deriving the forward solution. To this end we use a method similar to that of undetermined coefficients.

The final form of y_t^F from (2.3) may be written as

$$y_t^F = F_1 x_{t-1} + F_0 x_t + F_D y_{t-1} + v_t \tag{2.5}$$

where F_1, F_0 and F_D are gxk, gxk and gxg matrices respectively. Lags of greater than one will not enter (2.5) for the simple reason that given (3.1) and (3.2) the variables dated at time t-1 are a sufficient basis for determining the forward expectations stream x_{t+i}^e and y_{t+i}^e which arise on forward solution of (2.5).

The next step is to advance (2.5) by one period and take expectations conditional on Ω_{t-1} to get

$$y_{t+1}^e = F_1 P x_{t-1} + F_0 P^2 x_{t-1} + F_D y_t^e \tag{2.6}$$

Now taking expectations of (2.5) gives an expression for y_t^e which when substituted into (2.6) gives

$$y_{t+1}^e = (F_1 P + F_0 P^2 + F_0 F_1 + F_D F_0 P) x_{t-1} + F_D^2 y_{t-1}$$

Substituting this expression for y_{t+1}^e in (2.3) and comparing coefficients therein with their counterparts in (2.5) gives the implicit relationships for F_1, F_0 and F_D as

$$\begin{aligned} &\text{(i)} && F_0 = B \text{ and so} \\ &\text{(ii)} && F_1 = A(F_1 P + BP^2 + F_D F_1 + F_D BP) \\ &\text{(iii)} && F_D = AF_D^2 + D \end{aligned} \tag{2.7}$$

Explicit forms for F_1 and F_D are not available (although a form for vecF_1 in terms of A, B, P and F_0 does exist) but (2.7) provides all the information we need to derive analytical derivatives of the likelihood function when the forward solution is imposed.

We turn now to the derivation of a 'backward looking' solution to (2.3). Rearranging (2.3) gives

$$Ay_{t+1}^e = y_t - Bx_t - Dy_{t-1} - v_t$$

or $\quad Ay_{t+1} = y_t - Bx_t - Dy_{t-1} - v_t - A\varsigma_{t+1}$

where $\quad \varsigma_{t+1} = y_{t+1}^e - y_{t+1}$

is the error in the forward R.E. Taking y_t to the L.H.S. and inverting the L.H.S. matrix lag polynomial gives

$$y_{t+1} = - [A-IL]^{-1}Bx_t - [A-IL]^{-1}Dy_{t-1} - [A-IL]^{-1}A\varsigma_{t+1} - [A-IL]^{-1}v_t$$

Taking expectations conditional on Ω_{t-1} again and substituting back into (2.3) gives

$$y_t = [I-A^{-1}L]^{-1}BPx_{t-1} - [I-A^{-1}L]^{-1}Dy_{t-1} + Bx_{t-1} + Dy_{t-1}$$
$$- [I-A^{-1}L]^{-1}u_t^e + [I-A^{-1}L]^{-1}\varsigma_{t+1}^e + v_t$$

Finally premultiplying both sides by $[I-A^{-1}L]$ gives a form for $y_t{}^B$ as

$$y_t{}^B = A^{-1}y_{t-1} - [BP+A^{-1}B]x_{t-1} - A^{-1}Dy_{t-1} + Bx_t \tag{2.8}$$
$$+ v_t - A^{-1}v_{t-1}$$

To see that $y_t{}^B$ is a solution to (2.3) we check that the two initial conditions for $y_t{}^e$ and y_{t+1}^e derived from (2.8) are consistent with (2.3).

From (2.3) it is obvious that

$$y_t{}^e = y_t - B\varepsilon_t - v_t$$

and it is immediately apparent that (2.8) yields the same form for $y_t{}^e$.

Advancing (2.8) by one period and taking expectations yields on initial condition for y_{t+1}^e as

$$y_{t+1}^e = A^{-1}y_t{}^e - A^{-1}Dy_{t-1} + A^{-1}BPx_{t-1}$$

and substituting this into (2.3) gives

$$y_t = y_t{}^e - Dy_{t-1} + BPx_{t-1} + Bx_t + Dy_{t-1} + v_t$$

or $$y_t = y_t{}^e + B\varepsilon_t + v_t$$

an initial condition for $y_t{}^e$ that we know satisfies (2.8) and (2.3) and so the condition for y_{t+1}^e from (2.8) is consistent with that from (2.3).

Now, combining $y_t{}^B$ from (2.8) with $y_t{}^F$ from (2.5) in (2.4) gives

(a) $$y_t = [\Lambda F_D+(I-\Lambda)A^{-1}]y_{t-1} - (I-\Lambda)A^{-1}Dy_{t-2} + Bx_t + [\Lambda F_1$$
$$- (I-\Lambda)(BP+A^{-1}B)]x_{t-1} + v_t - (I-\Lambda)A^{-1}v_{t-1} \tag{2.9}$$

or more compactly

(b) $$y_t = H_1y_{t-1} + H_2y_{t-2} + Bx_t + J_1x_{t-1} + v_t - (I-\Lambda)A^{-1}v_{t-1}$$

where H_1, H_2 and J_1 are so as to conform with (2.9)(a).

We have derived then useful analytical forms for the restrictions implied by two distint solutions to (2.3). This enables us to write the likelihood function for the y_t in (2.9) and so to derive the F.O.C.

Before we go any further we must recall some matrix algebra and calculus. Using the theorems and conventions of Neudecker (1969) the following is true for conformable matrices A, B, and C

(a) $\mathrm{vec}(ABC) = (C' \otimes A)\mathrm{vec}B$

(b) $dy_j/dx_i = [d_{ij}]$ (2.10)

where the elements y_j and x_i form the vector y and x respectively

(c) $\dfrac{d\mathrm{vec}A}{d\mathrm{vec}B} = \dfrac{d\mathrm{vec}C}{d\mathrm{vec}B}\ \dfrac{d\mathrm{vec}A}{d\mathrm{vec}C}$

(d) $d(\mathrm{vec}A) = \mathrm{vec}(dA)$

(e) $\delta(AB) = (\delta A)B + A(\delta B)$

and so

(f) $\dfrac{d\mathrm{vec}AB}{d\mathrm{vec}C} = \dfrac{d\mathrm{vec}A}{d\mathrm{vec}C}(B \otimes I) + \dfrac{d\mathrm{vec}B}{d\mathrm{vec}C}(I \otimes A')$

(g) $\dfrac{d\mathrm{vec}(A^{-1})}{d\mathrm{vec}A} = -(A^{-1} \otimes A'^{-1})$

(h) $\dfrac{d\log|A|}{d\mathrm{vec}A} = \mathrm{vec}A'^{-1}$

(i) $\dfrac{d\mathrm{tr}AB}{d\mathrm{vec}B} = (A)'$

Now to write the joint likelihood function of the y_{it}, $t=1,\dots,T$, $i=1,\dots,g$ replace the gx1 and kx1 vectors in (2.9) (y_t, y_{t-1}, y_{t-2} and x_t, x_{t-1} respectively) by the gxT and kxT data matrices Y', Y_1', Y_2' and X', X_1' respectively. The log liklihood function then becomes

$$L(\mathrm{vec}(Y')) = T/2 \log|\bar{\bar{\Sigma}}|^{-1} - 1/2[\mathrm{vec}(Y'-H_1Y'_1-H_2Y'_2-BX'-J_1X'_1)]' \bar{\bar{\Sigma}}^{-1}[\mathrm{vec}(Y'-H_1Y'_1-H_2Y'_2-BX'-J'_1X'_1)] \quad (2.11)$$

where

$$\bar{\bar{\Sigma}} = [(\Sigma-[I-\Lambda]A^{-1}\Sigma A'^{-1}[I-\Lambda])\otimes I] - [(I-\Lambda)A^{-1}\Sigma \otimes R]$$

where

$$R = \begin{bmatrix} 0 & 1 & & \\ & & 0 & \\ 1 & & & \\ & & & 1 \\ & 0 & & \\ & & 1 & 0 \end{bmatrix}$$

and where $\Sigma = C'SC$

Now using the relations in (2.10) it is reasonably straightforward to derive the F.O.C. for maximisation of (2.11) with maximands vecC, $vec\Omega$, $vec\Gamma$, $vec\Delta$, vecS and $vec\Lambda$. These F.O.C. may be written as products of their components as

$$(a)\quad \frac{dL}{dvec\Omega} = \frac{dvecA}{dvec\Omega}\cdot\frac{dvecA^{-1}}{dvecA}\cdot\left[\frac{dvec\bar{\Sigma}}{dvecA^{-1}}\cdot\frac{dvec\bar{\Sigma}^{-1}}{dvec\bar{\Sigma}}\cdot\frac{dL}{dvec\bar{\Sigma}} + \frac{dvecF_D}{dvecA^{-1}}\cdot\frac{dvecH_1}{dvecF_D}\cdot\frac{dL}{dvecH_1} + \frac{dvecF_1}{dvecA^{-1}}\cdot\frac{dvecJ_1}{dvecF_1}\,\frac{dL}{dvecJ_1} + \frac{dvecH_2}{dvecA^{-1}}\cdot\frac{dL}{dvecH_2}\right] = R\Omega\cdot\frac{dL}{dvec\Omega}$$

$$(b)\quad \frac{dL}{dvec\Gamma} = \frac{dvecB}{dvec\Gamma}\cdot\left[\frac{dvecF_D}{dvecB}\cdot\frac{dvecH_1}{dvecF_D}\cdot\frac{dL}{dvecH_1} + \frac{dvecF_1}{dvecB}\cdot\frac{dvecJ_1}{dvecF_1}\cdot\frac{dL}{dvecJ_1} + \frac{dvecJ_1}{dvecB}\cdot\frac{dL}{dvecJ_1}\right] = R\Gamma\cdot\frac{dL}{dvec\Gamma}$$

$$(c)\quad \frac{dL}{dvec\Delta} = \frac{dvecD}{dvec\Delta}\cdot\left[\frac{dvecF_D}{dvecD}\cdot\frac{dvecH_1}{dvecF_D}\cdot\frac{dL}{dvecH_1} + \frac{dvecH_2}{dvecD}\cdot\frac{dL}{dvecH_2} + \frac{dvecF_1}{dvecD}\cdot\frac{dvecJ_1}{dvecF_1}\cdot\frac{dL}{dvecJ_1}\right] = R\delta\cdot\frac{dL}{dvec\Delta} \qquad (2.12)$$

$$(d)\quad \frac{dL}{dvecC} = \frac{dvecC^{-1}}{dvecC}\cdot\left[\frac{dvecA}{dvecC^{-1}}\cdot\frac{dL}{dvecA} + \frac{dvecB}{dvecC^{-1}}\cdot\frac{dL}{dvecB} + \frac{devecD}{dvecC^{-1}}\cdot\frac{dL}{dvecD} + \frac{dvec\bar{\Sigma}}{dvecC^{-1}}\cdot\frac{dvec\bar{\bar{\Sigma}}}{dvec\bar{\Sigma}}\cdot\frac{dvec\bar{\bar{\Sigma}}^{-1}}{dvec\bar{\bar{\Sigma}}}\cdot\frac{dL}{dvec\bar{\bar{\Sigma}}^{-1}}\right]$$

$$= Rc\cdot\frac{dL}{dvecC}$$

$$(e)\quad \frac{dL}{dvecS} = \frac{dvec\bar{\Sigma}}{dvecS}\cdot\frac{dvec\bar{\bar{\Sigma}}}{dvec\bar{\Sigma}}\cdot\frac{dvec\bar{\bar{\Sigma}}^{-1}}{dvec\bar{\bar{\Sigma}}}\cdot\frac{dL}{dvec\bar{\bar{\Sigma}}^{-1}} = Rs\cdot\frac{dL}{dvecS}$$

$$(f)\quad \frac{dL}{dvec\Lambda} = \frac{dvecH_1}{dvec\Lambda}\cdot\frac{dL}{dvecH_1} + \frac{dvec\bar{\bar{\Sigma}}}{dvec\Lambda}\cdot\frac{dvec\bar{\bar{\Sigma}}^{-1}}{dvec\bar{\bar{\Sigma}}}\cdot\frac{dL}{dvec\bar{\bar{\Sigma}}^{-1}} + \frac{dvecH_2}{dvec\Lambda}\cdot\frac{dL}{dvecH_2} + \frac{dvecJ_1}{dvec\Lambda}\cdot\frac{dL}{dvecJ_1} = 0$$

where $R\Omega$ is a $g^2 x g^2$ selector matrix of zeros with a unit entry on the diagonal in the jth row if the jth element of $vec\Omega$ is contrained to zero (excluded). Rs, Rc and $R\delta$ are the corresponding selector matrices for S, C and Δ respectively.

Of the derivatives in (2.12) only $dvecF_D/dvecD$, $dvecF_D/dvecA$ and $dvec\bar{\bar{\Sigma}}/dvec\Sigma$, $dvec\bar{\Sigma}/dvecA^{-1}$ and $dvec\bar{\Sigma}/dvec\Gamma$ are sufficiently obscure given the rules in (2.10) to deserve explicit analysis.

Recall that

$$F_D = A^{-1}F_D{}^2 + D$$

Now taking vecs of both sides and differentiating with respect to some vector say vecQ (Q is an m x n matrix) gives

$$\frac{dvecF_D}{dvecQ} = \frac{dvecA^{-1}}{dvecQ} \cdot (F_D{}^2 \otimes I) + \frac{dvecF_D^2}{dvecQ}(I \otimes A'^{-1}) + \frac{dvecD}{dvecQ}$$

Reapplying the same rule to $dvecF_D{}^2/dvecQ$ and solving for $dvecF_D/dvecQ$ gives

$$\frac{dvecF_D}{dvecQ} = \left[\frac{dvecA^{-1}}{dvecQ} \cdot (F_D{}^2 \otimes I) + \frac{dvecD}{dvecQ}\right]\left[(F_D \otimes A'^{-1}) + (I \otimes F_D{}'A'^{-1})\right]^{-1} \qquad (2.13)$$

Turning now to $dvec\bar{\bar{\Sigma}}/dvecQ$, this is the $mnx(gT)^2$ matrix

$$\frac{dvec\bar{\Sigma}}{dvecQ} = \left[Z_{11}\ Z_{21}\ Z_{31} \ldots Z_{g1}\ Z_{12}\ Z_{22}\ Z_{32} \ldots Z_{g2} \ldots Z_{1T}\ Z_{2T}\ Z_{3T} \ldots Z_{gT}\right]$$

where $Z_{i1} = \{v_i w_i [0 \ldots. 0]\}$

$Z_{i2} = \{\bar{w}_i v_i w_i [0 \ldots. 0]\}$

$Z_{i3} = \{[0]\bar{w}_i v_i w_i [0 \ldots. 0]\}$

.

.

$Z_{(T-2)} = \{[0 \ldots. 0]\bar{w}_i v_i w_i [0]\}$

$Z_{i(T-1)} = \{[0 \ldots. 0]\bar{w}_i v_i w_i\}$

$Z_{iT} = \{[0 \ldots. 0]\bar{w}_i v_i\}$

and where v_i is the mnxg vector which forms the ith set of g columns (from left to right) of

$$\frac{d\,\mathrm{vec}(\Sigma-(I-\Lambda)A^{-1}\Sigma A'^{-1}(I-\Lambda))}{d\,\mathrm{vec}Q},$$

w_i is the mnxg matrix which forms the ith set of g columns (from left to right) of

$$\frac{d\,\mathrm{vec}(-\Sigma A'^{-1}(I-\Lambda))}{d\,\mathrm{vec}Q},$$

and where $\bar{w}_i$ is the mnxg matrix chosen as w_i but from

$$\frac{d\,\mathrm{vec}(-(I-\Lambda)A^{-1}\Sigma)}{d\,\mathrm{vec}Q}$$

The F.O.C. in (2.12) then define M.L.E.'s of the parameters of any dynamic simultaneous equation model with forward R.E. of time horizon one.

In general no neat explicit solution will be available for the maximands in (2.12). It is therefore suggested that (2.12)(a) to (2.12)(f) be solved block by block iteratively using a NAG routine such as E04FAF (residual sum of squares minimisation routine) to find values of the maximands that satisfy the F.O.C. at each iteration.

In the next section we describe a programme CLARE (Computationally efficient Likelihood estimation Algorithm for Rational Expectations) which solves (2.12)(a) to (2.12)(f) with D = 0 (no predetermined variables) and Δ = I (forward solution). The programming of the general F.O.C. in (2.12) in full however is left as the subject of future research.

We now move to tackle the two very important generalisations to the above namely that of an expectations horizon greater than one and the case where Ω is singular. We deal with them in this order.

To accomodate an expectation horizon of not greater than n it is sufficient to consider the model

$$Cy_t = \Omega y^e_{t+n} + \Gamma x_t + \Delta y_{t-1} + u_t$$

with quasi reduced form

$$y_t = Ay^e_{t+n} + Bx_t + Dy_{t-1} + v_t \tag{2.14}$$

where notation is now obvious.

The general solution to (2.14) may now be written in an analagous way to (2.4) as

$$\begin{aligned} y_t = {} & \Lambda F_D y_{t-1} - (I-\Lambda)A^{-1}Dy_{t-n-1} + Bx_t - (I-\Lambda)A^{-1}Bx_{t-n} \\ & + [\Lambda F_1-(I-\Lambda)BP]x_{t-1} + v_t - (I-\Lambda)A^{-1}v_{t-n-1} \end{aligned} \tag{2.15}$$

with (2.7)(b) and (c) replaced with

$$\begin{aligned} &(a)\ F_1 = A\sum_{i=0}^{n} F_D{}^i F_1 P^{n-i} + A\sum_{i=0}^{n} F_D{}^i BP^{n-i+1} \\ &(b)\ F_D = AF_D{}^{n+1} + D \end{aligned} \tag{2.16}$$

Note that we may still obtain derivatives such as $dvecF_1/dvecQ$ by successive application of the vec differentiation rule (2.10)(f). In particular

$$\begin{aligned} &(a)\ \frac{dvecF_1}{dvecQ} = \left[\sum_{i=0}^{n} \frac{dvecAF_D{}^i}{dvecQ} \{[F_1P^{n-i}\otimes I] + [BP^{n-i+1}\otimes I]\}\right] \\ &\qquad \left[I-\sum_{i=0}^{n}(P^{n-i}\otimes A'F_D{}^i)\right]^{-1} \\ &(b)\ \frac{dvecAF_D{}^i}{dvecQ} = \frac{dvecA}{dvecQ}(F_D{}^i\otimes I) + \frac{dvecF_D{}^i}{dvecQ}(I\otimes A') \\ &\text{and (c)}\ \frac{dvecF_D{}^i}{dvecQ} = \frac{dvecF_D}{dvecQ}(F_D{}^{i-1}\otimes I) + \frac{dvecF_D{}^{i-1}}{dvecQ}(I\otimes F_D{}') \end{aligned} \tag{2.17}$$

(2.17)(a) to (2.17)(c) may be used to construct derivatives analogous to those in (2.12). Writing a programme for the case of an n-period expectations horizon is thus a similar although considerably more complex task compared with that based on (2.12).

Finally we consider the model in (2.3) but with A singular so that A^{-1} does not exist.

To overcome this problem we seek a transformation of the model such that the coefficient matrix on y^e_{t+1} is invertible. Rao (1965) shows that for any nxm matrix of rank r (say A) there exists nonsingular square matrices N(nxn) and M(mxm) such that

$$A = N \begin{bmatrix} I_r & 0 \\ 0 & 0 \end{bmatrix} M$$

where N is orthogonal so that

$$N'N = I$$

In our case n=m=g and so we may rewrite (2.3) as

$$\begin{bmatrix} N_1' \\ N_2' \end{bmatrix} y_t = \begin{bmatrix} M_{11} & M_{12} \\ & 0 \end{bmatrix} y^e_{t+1} + \begin{bmatrix} (N'B)_1 \\ (N'B)_2 \end{bmatrix} x_t + \begin{bmatrix} (N'D)_1 \\ (N'D)_2 \end{bmatrix} y_{t-1} + N'v_t \qquad (2.18)$$

where N_1', N_2' are the rxg,(g-r)xg submatrices of N' and are thus of full rank. $(N'B)_1$, $(N'B)_2$ and $(N'D)_1$, $(N'D)_2$ are the respective rxk,(g-r)xk submatrices of N'B and N'D. Taking the second (lower) block of equations, advancing by one period and taking expectations conditional on Ω_{t-1} gives

$$N_2'y^e_{t+1} = (N'B)_2P^2x_{t-1} + (N'D)_2y_t^e$$

Now substituting for y_t^e in terms of y_t and the error in the current expectation and rearranging gives

$$(N'D)_2y_t = N_2'y^e_{t+1} + (N'D)_2Bx_t - [(N'B)_2P^2 + (N'D)_2BP]x_{t-1} + (N'D)_2v_t$$

Stacking this with the upper set of equations in (2.18) gives the new system

$$\begin{bmatrix} N_1' \\ (N'D)_2 \end{bmatrix} y_t = \begin{bmatrix} M_1 \\ N_2' \end{bmatrix} y^e_{t+1} + \begin{bmatrix} N_1 \\ (N'D)_2B \end{bmatrix} \begin{bmatrix} 0 \\ -(N'B)_2P^2+(N'D)_2BP \end{bmatrix} \begin{bmatrix} x_t \\ x_{t-1} \end{bmatrix}$$
$$+ \begin{bmatrix} (N'D)_1 \\ 0 \end{bmatrix} y_{t-1} + \begin{bmatrix} N_1' \\ (N'D)_2 \end{bmatrix} v_t \qquad (2.19)$$

or

$$\bar{C}y_t = \bar{\Omega}y^e_{t+1} + \bar{\Gamma}x_t + \bar{\Delta}y_{t-1} + \bar{w}v_t \qquad (2.20)$$

Note that now P(in (2.2)) = $\bar{P}$ where

$$\bar{P} = \begin{bmatrix} P & 0 \\ 0 & I \end{bmatrix}$$

$\bar{\Omega}$ in (2.20) is now non singular.

We consider our decomposition to be superior to the process described in Pagan (1982) called the 'shuffle algorithm' as our method converges in one step enabling the form (2.19) to be subject readily to analysis and in particular to differentiation. The final form is now written as in (2.4) with A replaced by $\bar{A}$, B with $\bar{B}$ etc, and the F.O.C.'s are as in (2.12) except the maximands are now $\bar{C}$, $\bar{\Omega}$, $\bar{\Gamma}$, $\bar{\Delta}$, Λ and $\bar{w}$. These may be transformed to C, Ω, Γ, Δ Λ by premultiplication of the relevant F.O.C. by matrix terms such as

$$\frac{d\,vec\bar{C}}{d\,vecC}, \quad \frac{d\,vec\bar{\Omega}}{d\,vecC}, \quad \frac{d\,vec\bar{\Gamma}}{d\,vecC}, \quad \text{etc.}$$

In practice most applied economists would be satisfied with the forward solution (Λ=I) and so the fact that Ω is not of full rank is irrelevant. We believe that unless there is compelling a priori evidence to the contrary, the forward solution should be imposed since purely backward looking solutions rule out the possibility of news about future exogenous events affecting current endogenous variables. In reality the merest casual empiricism will reveal that such news has a profound influence on current economic events in sectors of the economy where forward expectations are relevant to decision making.

As suggested in chapter 4 evidence in favour of (but not against) imposing the forward solution could take the form of examining the eigen values of a first stage estimate of A and if they all lie inside the unit circle then this restriction could be imposed on A itself and the forward solution imposed with more confidence than otherwise.

These reservations aside, the analysis in this section when implemented in the form of a programme can provide a relatively efficient means of deriving FIML estimates of both the structural parameters and the solution parameters (the λ's in Λ). This programme would be capable of handling most simultaneous econometric models.

The next section describes an implementation of the analysis in this paper for the case of D=0 and Λ=I, the implementation taking the form of the estimation program CLARE.

6.3 CLARE (Computationally efficient likelihood algorithm for estimating R.E.)

In this section we describe a programme CLARE which delivers FIML estimates of the parameters of the model

(a) $\quad Cy_t = \Omega y^e_{t+1} + \Gamma x_t + u_t$

with quasi reduced form

(b) $\quad y_t = Ay^e_{t+1} + Bx_t + v_t \qquad (3.1)$

and x-processes

(c) $\quad x_t = Px_{t-1} + \varepsilon_t$

under the assumption that all the eigen values of A lie inside the unit circle leaving the forward solution unique amongst the set of stationery solutions.

The reduced form (solution) to (3.1) is simply

$$y_t = Fx_{t-1} + Bx_t + v_t \qquad (3.2)$$

where

$$F = AFP + ABP^2 \qquad (3.3)$$

We may write the log likelihood function as

$$L = T/2 \log|\Sigma|^{-1} - 1/2 \, tr\Sigma^{-1}(Y-X_1F'-XB')'(Y-X_1F'-XB') \qquad (3.4)$$

where all matrices are as in the previous section. Using the rules in (2.10) it is simple but tedious to show that the F.O.C. for maximization of (3.4) are

$$\text{(a)} \quad \frac{dL}{dvec\Omega} \equiv -(I\otimes C'^{-1})[(FP+BP^2)\otimes I][I-P\otimes\Omega']^{-1}.vec(\Sigma^{-1}R'X_1) = R_\Omega \frac{dL}{dvec\alpha}$$

$$\text{(b)} \quad \frac{dL}{dvec\Gamma} \equiv -(I\otimes C'^{-1})\left[[P^2\otimes A'] \; [I-P\otimes A']^{-1}vec(\Sigma^{-1}R'X_1)+vec(\Sigma^{-1}R'X)\right]$$

$$= R_\Gamma \frac{dL}{dvec\Gamma} \qquad (3.5)$$

$$(c)\ \frac{dL}{dvecC} = (C^{-1} \otimes C'^{-1}) \Big[(\Gamma \otimes I)\{(P^2 \otimes A')(I-P\otimes A')^{-1} vec(\Sigma^{-1} R'X_1)\}$$

$$+ vec(\Sigma^{-1}R'X)\} + (\Gamma \otimes I)\{[(FP+BP^2)\otimes I](I-P\otimes A')^{-1} vec(\Sigma^{-1}R'X_1)\}$$

$$+ (I^P(C'\Sigma^{-1}\otimes I) + (I\otimes C'\Sigma^{-1}))vec(T/2\Sigma-1/2\ R'R) \Big] = R_c \frac{dL}{dvecC}$$

where I^P is a permutation of the $g^2 x g^2$ identity matrix and is equal to $dvecC'^{-1}/dvecC^{-1}$ and where Σ is unrestricted so that $dL/dvec\Sigma^{-1} = 0$ solves uniquely for Σ and S and is written in the standard way as

$$(d)\ \frac{dL}{dvec\Sigma^{-1}} = 0 \Rightarrow \Sigma = \frac{R'R}{T}$$

Finally $R = Y-X_1F'-XB'$.

Note that it is only possible to derive from (3.5) explicit solutions for vecF, $vec\Sigma$ and $vec\hat{\Gamma}$ where $vec\hat{\Gamma}$ is subject to exclusion restrictions such that

$$(I-R_\Gamma)\ vec\Gamma = vec\hat{\Gamma}$$

The solution for $vec\hat{\Gamma}$ is tricky and achieved as follows. Solve the leftmost equality in (3.5)(b) for $vec\Gamma$ in terms of $dL/dvec\Gamma$ to get a form such as (for example)

$$vec\Gamma = Q^{-1}\ (g- \frac{dL}{dvec\Gamma}) \tag{3.6}$$

where Q is a gkxgk matrix of full rank and g is a gkx1 vector. Now replace $dL/dvec\Gamma$ in (3.6) with the value it assumes to satisfy the F.O.C. when evaluated at $vec\hat{\Gamma}$ so that (3.6) becomes

$$vec\Gamma = Q^{-1}(g-R_\Gamma[g-Q(I-R_\Gamma)vec\Gamma]) \tag{3.7}$$

Now solve for $vec\Gamma$ to get explicit form for $vec\hat{\Gamma}$ that satisfies the F.O.C. in the rightmost equality of (3.5)(b) as

$$vec\hat{\Gamma} = [I-Q^{-1}R_\Gamma Q(I-R_\Gamma)]^{-1}Q^{-1}(I-R_\Gamma)g \tag{3.8}$$

More explicitly for our likelihood function $vec\hat{\Gamma}$ is

$$vec\hat{\Gamma} = [(I\otimes C'^{-1})E(I\otimes C^{-1})-R_\Gamma(I\otimes C'^{-1})E(I\otimes C^{-1})(I-R_\Gamma)]^{-1} \tag{3.9}$$

$$[(I\otimes C'^{-1})-R_{\Gamma}(I\otimes C'^{-1})]D$$

where $E = (p^2\otimes A')(I-p\otimes A')^{-1}[(X_1'X\otimes\Sigma^{-1})+(X'X\otimes\Sigma^{-1})] + [X_1'X_1\otimes\Sigma^{-1}+X'X_1\otimes\Sigma^{-1}](I-p'\otimes A)^{-1}(p^{2\prime}\otimes A)$

and $D = [(p^2\otimes A')(I-p\otimes A')^{-1}\mathrm{vec}\Sigma^{-1}(Y'X_1)+\mathrm{vec}\Sigma^{-1}(Y'X)]$

The five blocks of equations for $R_c dL/d\mathrm{vec}\hat{C}$, $R_{\Omega}dL/d\mathrm{vec}\hat{\Omega}$, vecF, $\mathrm{vec}\hat{\Gamma}$ and $\mathrm{vec}\Sigma$ were programmed in Algol 60 along with a subroutine to calculate the Hessian (to provide asymptotic standard errors), the final value of the likelihood function at the maximum and a few other diagnostic statistics (D.W. and R^2). The NAG library sum of squares minimisation routine E04FAA was used to solve the equations $R_c\ dL/d\mathrm{vec}\hat{C}$ and $R_{\Omega}dL/d\mathrm{vec}\hat{\Omega}$ for $\mathrm{vec}\hat{C}$ and $\mathrm{vec}\hat{\Omega}$ respectively at each iteration. One iteration of CLARE solves for $\mathrm{vec}\hat{\Gamma}$, $\mathrm{vec}\hat{C}$, $\mathrm{vec}\Sigma$, $\mathrm{vec}\hat{\Omega}$, and vecF in that order and on convergence the programme calls the subroutine HESS which provides asymptotic standard errors and some diagnostic statistics.

Data are input by means of a single datafile (input stream CLAREIP on the Warwick B67) and output is to a disc outputstream called CLAREOP. All data must be entered in free format in blocks as set out in table 6.1.

IG, IK, IT, IC, IA, IRG are integers describing the number of endogenous variables, exogenous variables, observations, and nonexcluded elements of vecC, $\mathrm{vec}\Omega$ and $\mathrm{vec}\Gamma$ respectively. MCAL specifies the maximum number of iterations allowed to solve the F.O.C. using NAG Routine E04FAA. Recommended value for this parameter if in doubt is 50. EPS sets the step length for numerical differentiation of the F.O.C's for calculation of the Hessian matrix. Its recommended value is 0.1xTTOL. The latter is a measure of convergence. When the sum of the squared updates on all the parameters is less than TTOL then the programme will enter HESS and terminate. For a model of around 10 parameters for example, TTOL=0.001 will give (at least) estimation accuracy in the second decimal place. STPMXA and STPMXC are real arrays which set the maximum step length used by E04FAA to numerically differentiate the F.O.C. for $\mathrm{vec}\hat{C}$ and $\mathrm{vec}\hat{\Omega}$ respectively. Each step should be set to about one tenth of the final value expected for the relevant estimate.

TABLE 6.1

	PROGRAMME NAME	TYPE
PARAMETERS	MCAL,EPS,TTOL	I,R,R
	STPMXA[IA]	A
	STPMXC[IC]	A
	UPDA,UPDC,UPDG	R,R,R
	IG,IK,IT,IC,IA,IRG	I,I,I,I,I,I
	XTOL,FTOL1	R,R
	VRC[IG2]	A
	VRA[IG2]	A
	VRG[IGK]	A
hATA and START VALUES	P[IK,IK]	A
	Σ[IG,IG]	A
	VCR[IC]	A
	VAR[IA]	A
	VGR[IRG]	A
	X[IT,IK]	A
	XI[IT,1K]	A
	Y[IT,IG]	A

KEY: I=INTEGER R=REAL A=REAL ARRAY [.,.]=ARRAY DIMENSIONS

UPDA, UPDC and UPDG are the gain factors for the iterative procedure. Values of 1.0, 1.5 and 1.0 were found to work well but if in doubt setting them all to unity will be robust. XTOL and FTOL1 set the accuracy to which E04FAA solves the F.O.C. Recommended values are 10^{-5} and 10^{-10} respectively. These values are not critical however since convergence seemed to be unaffected by errors in solving the F.O.C's of vec$\hat{C}$ and vec$\hat{\Omega}$. Obviously the smaller the values the longer the programme spends in E04FAA but the shorter the time to convergence of all the parameters. The arrays VRC, VRA and VRG should contain a 1 or 0 according to whether the corresponding element of vecC and vecΓ is included or not. Note that the diagonal elements of C are considered excluded.

Turning to the data, P is the kxk parameter matrix in (3.1)(c) and is entered row by row. Σ provides start values for the variance covariance matrix of v_t and again is entered row by row. VCR, VAR and VGR are start values for the included elements of vecC, vecΩ and vecΓ respectively. The elements should be entered in order of appearance in their respective vecs. Finally, the TxK, TxK and Txg data matrices of X, X1, and Y should be entered column by column i.e., in standard time series form.

On the B67 at Warwick processing time used to provide FIML estimates of the 12 parameters in the model in the next section varied from about 15 to 30 minutes where accuracy was to the fourth decimal place. This time is between 30 and 60 times that taken by the TSP job which provided limited information 2SLS estimates as start values for the parameters. Timing seems largely to depend on the fit of the model.

Finally, output is fairly brief and gives the estimates of C, Ω, Γ and Σ respectively row by row. Beneath these in the "diagnostic statistics" section is provided the diagonal elements of the Hessian function

$$-\left[\frac{d^2L}{\delta z \delta z'}\right] \qquad \text{where } z = \begin{bmatrix} \text{vec}\,\hat{C} \\ \text{vec}\,\Omega \\ \text{vec}\,\Gamma \end{bmatrix}$$

These are estimates of asymptotic standard errors. Finally after the D.W. and R^2 statistics comes the value of the L.R. test which compares the value of the likelihood function at the point of the R.E. and structural restrictions with that at the unrestricted point. High values of this statistic thus imply rejection of the R.E. restrictions by this criterion.

6.4 A small Monte Carlo Study

In this section we compare the performance of FIML against a limited information (IV) method due to McCallum (1977) by means of a small Monte Carlo Study of the following two equation model

$$\begin{bmatrix} 1.0 & C_{12} \\ C_{21} & 1.0 \end{bmatrix}\begin{bmatrix} y_{1t} \\ y_{2t} \end{bmatrix} = \begin{bmatrix} a_{11} & 0 \\ 0 & a_{22} \end{bmatrix}\begin{bmatrix} y^e_{1t+1} \\ y^e_{2t+1} \end{bmatrix} + \begin{bmatrix} B_{11} & B_{12} & 0 & 0 \\ B_{21} & B_{22} & 0 & 0 \end{bmatrix}\begin{bmatrix} x_{1t} \\ x_{2t} \\ x_{1t-1} \\ x_{2t-1} \end{bmatrix} + \begin{bmatrix} u_{1t} \\ u_{2t} \end{bmatrix} \tag{4.1}$$

or in terms of (3.1)(a)

$$Cy_t = \Omega y^e_{t+1} + \Gamma x_t + u_t \tag{4.2}$$

with AR(2) processes for the x_{it}

$$x_t = Px_{t-1} + \varepsilon_t \qquad (4.3)$$

where

$$P = \begin{bmatrix} 0.8 & 0.0 & -0.2 & 0.0 \\ 0.0 & 0.6 & 0.0 & -0.1 \\ 1.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 1.0 & 0.0 & 0.0 \end{bmatrix}, \quad \varepsilon_t = \begin{bmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ 0.0 \\ 0.0 \end{bmatrix} \quad \begin{matrix} (\ E(\varepsilon_{1t}^2) = \sigma_{\varepsilon_1}^2) \\ (\ E(\varepsilon_{2t}^2) = \sigma_{\varepsilon_2}^2) \end{matrix}$$

Recall that the reduced form in terms of observables is

$$y_t = Fx_{t-1} + Bx_t + v_t$$

where

$$\text{vec}F = [I-P'\otimes C^{-1}\bar{\Omega}]\text{vec}[C^{-1}\bar{\Omega}C^{-1}BP^2]$$

so that the restrictions on F depend upon the value assumed by P. Now the Monte Carlo exercise was conducted with the likelihood function being conditional on the x-processes and thus on the 'true' value of P. Limited information methods however do not impose restrictions implicitly and so could not make use of the 'true' value of P. As a result, $\sigma_{\varepsilon_1}^2$ and $\sigma_{\varepsilon_2}^2$ were chosen so as to give a good fit for the AR processes thus guaranteeing that the restrictions on F were more or less reflected by those obtained usng estimates of P ($\hat{P}$ using O.L.S on (4.3) say) from the generated data. In other words, care was taken to face the limited information method with data that corresponded to a genuine R.E. model.

The number of observations chosen for the exercise was 46 as this broadly corresponds with the typical length of post war time series (especially quarterly data) in applied economics.

The competitor for FIML chosen was McCallum's method. (McCallum (1977)). This was described in section 1 of this paper.

Accuracy of the FIML estimates was correct to four decimal places. Inspection of the Hessian for negative semi-definiteness was made for each set as a check on the validity of maximum. Some of the runs were checked for local maxima by restarting from various start values and no such problems arose.

Three sets of experiments were conducted. One with a very high R^2 (0.98) another with an R^2 more typically experienced in empirical macroeconomics (0.78) and a third set with a very poor fit of around 0.45. The first two sets contain four experiments and the third, two of independent data sets. Convergence time increased enormously as the fit decreased (hence only two experiments in the last set) it being double its value for high R^2 than for low R^2.

Finally, the variance covariance matrix of the structural errors were unrestricted and uninteresting and so were concentrated out of the likelihood function. This concentration was useful in that it reduced the dimension of the Hessian and so decreased the scope for numerical error in its calculation. The results of our 3 sets of experiments are reported in tables 5.2 to 5.4.

A glance at these tables show that despite the very good fit in the data sets in table 5.2 both vec$\hat{c}$ and vec$\hat{\Omega}$ are very poorly estimated indeed. This may be because R.E. terms contribute less than 10% to the 'explained variance' of the y_t and so are not likely to be well defined by any estimation procedure. This view is reinforced by the dual fact that the coefficients on the variables that contribute most to the explained variance of the y_t (Γ_{11} and Γ_{22} in particular) are very precisely estimated in virtually all our data sets and also that a halving of the R^2 to 0.45 is not reflected fully by a likewise proportional increase in bias.

Table 5.5 presents estimates of the means and variances of the 8 coefficients under both methods to provide summary measures of bias and efficiency. In columns 1 and 2 and 4 and 5 are standard mean and variance quantities calculated using the 4 estimation points provided by sets 1 and 2 and the 2 points of set 3 (variances were obviously not calculated for the latter). Explicitly

$$\text{Mean}(\theta) = \frac{\sum_{i=1}^{n} \theta_i}{n}, \quad \text{Variance}(\theta) = \left(\sum_{i=1}^{n} \theta_i^2 - n.[\text{Mean}(\theta)]^2\right)/n$$

where θ_i denotes an estimate of the parameter θ.

COLUMN	1	2	3		4	5
	FIML		SET 1 (R^2=0.98)		McCALLUM	
	Mean	Variance	Actual	Asymptotic Variance	Mean	Variance
c_{12}	-0.4371	0.00561	-0.5	0.00935	0.0467	0.17311
c_{21}	0.4113	0.01013	0.5	0.00641	-0.2834	0.04870
a_{11}	0.9919	0.00066	1.0	0.01844	0.4643	0.09442
a_{22}	1.0164	0.00301	1.0	0.00979	-0.3982	0.04987
Γ_{11}	0.4050	0.00003	0.4	0.000312	0.4364	0.00067
Γ_{12}	-0.0662	0.00200	-0.1	0.00344	0.1805	0.05990
Γ_{21}	0.1680	0.00350	0.2	0.00216	-0.1652	0.00606
Γ_{22}	0.5714	0.00063	0.6	0.00033	0.6342	0.00898
	(n=4)				(n=4)	
	FIML		SET 2 (R^2=0.78)		McCALLUM	
	Mean	Variance	Actual	Asymptotic Variance	Mean	Variance
c_{12}	-0.5754	0.07590	-0.5	0.28922	-0.4914	0.09576
c_{21}	0.3540	0.08366	0.5	0.15684	-0.3265	0.16910
a_{11}	0.9350	0.03678	1.0	0.07719	0.7611	0.23020
a_{22}	0.9211	0.01487	1.0	0.22356	0.4642	0.23935
Γ_{11}	0.4186	0.00078	0.4	0.00495	0.4720	0.00162
Γ_{12}	-0.1646	0.02572	-0.1	0.05652	-0.1206	0.02655
Γ_{21}	0.1586	0.03471	0.02	0.05416	-0.2545	0.03577
Γ_{22}	0.5311	0.00732	0.6	0.00601	0.4273	0.00758
	(n=4)				(n=4)	
	FIML		SET 3 (R^2=0.45)		McCALLUM	
	Mean	Variance	Actual		Mean	Variance
c_{12}	-0.4555	-	-0.5		0.0822	-
c_{21}	0.5974	-	0.5		-0.6768	-
a_{11}	0.8973	-	1.0		3.7211	-
a_{22}	0.6046	-	1.0		1.0664	-
Γ_{11}	0.4491	-	0.4		0.3733	-
Γ_{12}	0.1657	-	-0.1		-0.1241	-
Γ_{21}	0.1051	-	0.2		-0.4289	-
Γ_{22}	0.5775	-	0.6		0.4953	-
	(n=2)				(n=2)	

Finally, asymptotic standard errors for set 1 are provided. These are just averages for each set of the negative of the diagonal elements of the inverse of the Hessian matrix of the likelihood function. This should give some idea as to the accuracy of the Hessian as a measure of variance/covariance of the residuals.

Referring to the table it is clear that on both counts of mean and variance FIML is uniformly superior in virtually all the cases. Whilst this result is to be expected of FIML vis a vis limited information methods in correctly specified models, the high degree of bias and low efficiency of the limited information estimates in our study is both surprising and disturbing bearing in mind the method's widespread use. In poorly fitting models this problem seems highly acute with biases as high as 620% and many estimates wrongly signed.

7. A TEST OF THE R.E.H. IN THE CONTEXT OF A SIMPLE FINANCIAL MODEL.

7.1 Introduction

The purpose of this chapter is twofold. Firstly, using the results and apparatus in the previous chapter we wish to assess empirical evidence for the hypothesis that rational expectations (henceforth R.E.) play a crucial role in determining asset demands in financal markets. The second objective derives from the first. Investigators in this area had tried to explain the empirical importance of lagged asset stocks in their demand functions in terms of costly stock adjustment (see for example the work on discount house portfolios by Parkin (1970)). Our model leads us to a different interpretation of those empirical lags. Loosely speaking, our model has asset demand functions determining prices that, given an exogenous asset supply, clear the markets. Under the R.E. hypothesis, expectations of future asset prices, a crucial determinant of asset demands, are formed using lagged values of these stocks (and all other exogenous variables in the model). The reduced form of our model then includes lagged asset stocks that are associated not with costly adjustments but with expectation formation.

In the following section we describe the asset model with an exogenous variable sub model. Section 3 presents parameter estimates and Section 4 describes a test of the hypothesis and in particular focusses on the nature of the alternative hypothesis of the test. Section 5 provides a summary and conclusion.

7.2 The Model

The model is designed to focus on asset demands and so it treats supplies as exogenous. Aggregation over assets (four assets) and sectors (1 sector) is very coarse. Two asset prices are determined from two demand equations (gilts and equities). The third asset (money) is separable from (exogenous to) the rest of the model and because it is not of interest is not estimated. The fourth asset is a residual asset given by a wealth identity. There are seven exogenous variables, two asset stock quantity indices (gilts and equities), a short rate of interest, a price level, nominal income, nominal wealth and nominal dividends and these are modelled as joint AR processes. These were chosen because we believe they are a minimal set of key exogenous variables for the model. The joint AR model that links them, a sort of macroeconomic sub model is the simplest basis from which to form R.E. of future asset prices that is reasonably defensible. (Henceforth we refer to this AR model as the sub model and to that of the endogenous variables as the main model).

The four assets identified by the main model are: money used for transactions purposes (M2); gilt edged stocks with over two years to maturity; company equities and a residual asset which largely consists of highly liquid assets (building society deposits, treasury bills, local authority deposits and gilts with less than two years to maturity). There is no disaggregation over agents; the collection of a data base for such an exercise would in itself be a large undertaking. This absence of disaggregation severely limits the quantitative usefulness of the model (see Weale (1984) for a discussion of disaggregation in this context). We believe however that we still have a reasonable structure for testing the R.E. hyposthesis.

The structure of the model rests on a particular view of the behaviour of companies and of the authorities. The key policy intrument is the short rate of interest on liquid assets (our residual categuory) which we proxy with the treasury bill rate (r_t).

The authorities are faced with the following budget constraint

$$PSBR_t \equiv \Delta VGB_t + \Delta HPM_t - REV_t \qquad (2.1)$$

(2.1) simply states that a fiscal deficit (PSBR) must be financed by either gilts sales (equal to the change in the value of gilt holdings (ΔVGB_t) minus revaluations of outstanding debt (REV_t)) or by issuing high powered money. This excludes short period assets such as treasury bills and so represents a longer term view of the constraint. Given that the authorities have a target short rate of interest and given that there is some fixed relationship between high powered money and some definition of liquidity (M3 say) then the term ΔHPM_t is constrained to be that quantity that leads to a market clearing money stock. Any other value will cause upward or downward pressure on short rates. This leaves the current supply of new gilts (in value terms) driven by the PSBR and the short rate of interest. Now our model is in terms of stocks and for the supply of gilts it is useful to write the identity

$$GB_t \equiv GB_{t-1} + \Delta GB_t$$

where GB_t is the 'quantity' of bonds of an 'average' type (average maturity and price) outstanding .

Given the structure of policy described above then the number of new bonds supplied is driven by the PSBR and the targetted short rate of interest and will be independent of their price. This horizontal supply curve argument is however only approximately true. The number of bonds required to finance a given deficit will vary inversely with their sale price. The argument is constructed implicitly around an equilibrium price. In any event in the medium and long term the PSBR will be far and away the most important influence on the new supply of (and, therefore the outstanding stock of) gilts. There is still the argument that the government will be tempted to issue more (less) gilts when the price is high (low) accepting the short run consequences of this funding action for the short rate of interest. There is however, strong anecdotal evidence to suggest that such responses were, at the very most spasmodic up until the start of 1972.

> "No attempt was made to peg long rates or to offset market trends in the gilt market entirely but the bank usually 'leant into the wind' to reduce the rate of change of market prices in the interests of maintaining a broad orderly market for debt". (Goodhart (1984) p.92)

It goes without saying that funding activities which alter the structure of interest rates within our average (price) measure are not accounted for in our model. We assume that such activities do not affect the price of our 'average' bond. Taking the view that the PSBR is exogenous to our model it does not seen too unreasonable to model the (quantity) supply of gilts as an exogenous variable.

The supply of equities is treated in a similar fashion. We assume that investment is interest inelastic so that as with gilts we have a horizontal supply curve. Of course this view is not new. The Keynsian liquidity trap rests on interest inelastic investment. However, equity capital although the most important is not the only source of investment funds. Debentures and other loans may be called upon when the price of equities is too low. Further, the variance of equity prices far exceeds that of gilts so that a given level of investment (and it is the value of investment that we argue is exogenous) will require varying quantities of equity to finance it over our sample period. Sadly we are forced to abstract from such realities and maintain our heroic assumption that the supply of equity capital is independent of its price. One saving grace on this score is that we have GDP in our exogenous variable sub model and this is undoubtedly the most important variable influencing investment decisions.

Financial wealth (in our model, the sum of our four assets) is forecast as an exogenous variable. Broadly speaking the stock of wealth is the integral of saving over the past so that lagged wealth and GDP terms should be a sufficient basis for predicting current wealth. Once again this is not wholly true because wealth rises and falls with asset prices through revaluations. As with bonds and equities we argue that the influence of prices is minor relative to the determinants we have included.

Nominal income (GDP) is modelled primarily as an autoregressive process. This is not to say that asset prices (the endogenous variables in our model) do not affect GDP, in reality they probably do. For simplicity's sake their influence is ignored (although we still have a short rate of interest in the sub model). Our treatment of GDP allows the possibility of a trade cycle and so is a reasonable basis for expectations, at least in the context of our simple macroeconomic sub model.

Aggregate dividends, related to aggregate profits should be adequately explained by their own past and by (lagged) GDP. Dividends play a key role in the model. Unlike the coupons on gilts, nominal dividends are index linked. Broadly speaking (for a given income distribution) they will rise and fall with nominal GDP and not with the price level. Inflation then only plays a role in the exogenous variable sub model (as a determinant of short rates in fact). It has no role in the main model because all the variables relevant to returns on assets and competing assets are already present (namely, capital gains, current prices and dividends). The implicit assumption then is that real assets (other than equity) whose nominal return is primarily the rate of inflation, are not an option considered by our wealth holders.

This then completes our description of the sub model. Although there is an economic structure behind it, no overidentifying structural restrictions were imposed on the joint AR reduced form. Any exclusion restrictions arising from recursivity were however imposed. The sub model with parameter estimates is presented in the next section.

Finally we turn to the main model. The asset demand functions are simple log linear forms explaining quantity indices (outstanding stocks divided by a price index) in terms of own and competing capital gains and own and competing current returns. Wealth is a scaling variable. These quantity indices are exogenous so that given an equilibrium condition the two asset prices must clear the markets. The normalisation given below in equation (2.2) is therefore a little misleading.

The current return on equities is measured as aggregate nominal dividends divided by a price index giving the identity

$$RC \equiv D_t - PC_t + \text{constant}$$

where RC and PC are the rate of return on and price of equities and D_t is aggregate nominal dividends. Here and henceforth all variables are natural logarithms unless other wise stated.

Using this identity to replace current returns in the asset demand functions we described above gives us our main model.

$$
\begin{aligned}
&\text{(a)} \quad GB_t = \alpha_{11}(PG^e_{t+1} - PG_t) + \alpha_{12}(PC^e_{t+1} - PC_t) + C_{11}PG_t \\
&\qquad + C_{21}PC_t + \gamma_{14}r_t + \gamma_{17}W_t + \gamma_{19}D_t + \gamma_{110} + u_{1t} \\
&\text{(b)} \quad CB_t = \alpha_{21}(PG^e_{t+1} - PG_t) + \alpha_{22}(PC^e_{t+1} - PC_t) + C_{22}PC_t \\
&\qquad + C_{21}PG_t + \gamma_{24}r_t + \gamma_{27}W_t + \gamma_{29}D_t + \gamma_{210} + u_{2t}
\end{aligned}
\tag{2.2}
$$

GB and CB are the outstanding quantities of gilts and equities and PG and PC their respective prices. W and D are aggregate nominal wealth and dividends respectively and u_1 and u_2 are error terms. Superscript 'e' denotes a rational expectation formed at time t-1.

In writing (2.2) we have ignored completely the redemption values of government bonds. Because our data excludes short bonds we argue that the present (discounted) value of redemption monies is small enough to be excluded. In any case we would only expect it to be relevant when the average maturity date of the bonds in our sample is highly variable and we have no reason to believe this is so. Aside from this (2.2) includes all the variables of interest to asset holders. Explicitly these are capital gains $((PG^e_{t+1}-PG_t)$ and $(PC^e_{t+1}-PC_t))$, current returns ((constant coupon - PG_t), (D_t-PC_t) and r_t) and of course wealth (W_t).

If investors were neutral to risk we would expect their response to expected capital gains to be equal to their responses to current returns. No such constraint is imposed in (2.2) so leaving investors' attitude to risk unspecified. In any event the capital gains terms are approximately percentages whilst the current returns terms are logs of percentages and it is not clear what constraint risk neutrality implies for the coefficients.

A priori we would expect α_{11}, α_{22}, C_{12}, C_{21}, γ_{17}, γ_{27} and γ_{29} to be positive and α_{21}, α_{12}, C_{11}, C_{22}, γ_{14}, γ_{24} and γ_{19} to be negative.

Note that there are two restrictions arising from the identity linking D and PC namely

$$C_{12} = - \gamma_{19}$$

and $C_{22} = -\gamma_{29}$

In the estimated model, (where the equations are normalised on PG_t and PC_t) these simple restrictions become nonlinear. Unfortunately our estimation routine CLARE has no facility for imposing such restrictions and so the coefficients were estimated freely. The estimates for the main model are presented in the next section.

7.3 Estimates of the model

(i) The Data

The data are quarterly with the basic sample running from 1967(1) to 1971(4). Additional observations were collected spanning 1972(1) to 1976(4) but, in the case of asset stocks these data are highly unreliable as we discuss below.

The stocks, provided by the C.S.O. but not published were measured at the end of each quarter. Gilts consist of all outstanding government stock with more than two years to mature and equities consist purely of ordinary shares. Both were measured from the liabilities side of the balance sheet provided by the C.S.O. The 3.5% war loan price index was chosen to proxy PG, primarily because this issue still represents about one third of total gilts holdings. For the equity price (PC), the FT ordinary share index of industrial shares was chosen although over the entire period there were no substantial differences between the available series. Both price indices were measured on the last Friday of each quarter.

Dividends were calculated as the product of the FT thirty share dividend yield and the outstanding stock of equities. Nominal GDP was measured at factor cost from the expenditure side and the price level was the retail prices index. Both of these series are readily available from the E.T.A.S. (1977). The end of quarter treasury bill rate series is published in Financial Statistics.

Because the asset prices are measured at the end of the quarter it is vital that the book value of the stocks are right up to date when measured. A priori we might expect our quantity indices to be of a poor quality especially when the stock prices are highly volatile within and across quarters. A glance at the data shows that relative to our basic sample (1967(1) to 1971(4)) equity prices during the later period to 1976(4) were highly volatile. In particular, between 1974(4) and 1975(1) the index rose by over 83% and between 1973(3) and 1973(4) fell by 25%. In sharp comparison, the largest movement in our basic sample is only 12%. Now, our quantity index (for equities) is meant to represent the physical capital stock and can only contract through time because of bankruptcies or redemptions. Both of these influences on the capital stock over the period to 1976 were probably limited to a few percent of the stock. Referring to table 7.1 we see, however that our quantity index records over a 10% fall between 1972(4) and 1973(1), a staggering fall of 43% between 1974(4) and 1975(1) and a 7% fall between 1975(4) and 1976(1). A check with the data compilers at the C.S.O. revealed that the three quarters 1972(1) to 1972(3) are grossly underestimated but, at the time of writing we have received no revised data. We infer therefore that whilst the data of our basic sample has no discontinuities and anomalies, there are sharp discontinuities in the later sample as the graph in table 7.1 shows.

In marked contrast, the gilts quantity index graphed in table 7.1 is well behaved over the entire period to 1976(4). This is probably because our measure of gilts (or any measure for that matter) is far less volatile than the price of equities. A further explanation lies in our belief that the Bank of England keeps a much more up to date and accurate book record of outstanding government debt (gilts) than company secretary's do of company debt (equity).

These inaccuracies were unfortunate in that we were forced to constrain our basic sample to a mere 20 observations. There was, however another reason for this choice of sample, namely that there was a marked structural break in the behaviour of our exogenous variables around the end of 1971. This is hardly surprising considering the events taking place at the time. Competition and credit control restructured the banking sector and the clearing banks cartels were broken up. The

TABLE 7.1

QUANTITY INDICES FOR THE LATER SAMPLE

exchange rate was floated and provided a new target for the short rate of interest. Finally the miners' strike in 1974 was a peculiar factor influencing GDP in the later sample.

Bearing all these factors in mind, the focus of interest falls on our basic sample. We do use our later sample to casually assess parameter stability, etc. but these are only of passing interest and we hesitate to draw any results or inferences from this period.

(ii) <u>The exogenous variable sub model</u>

Because we have only 20 observations in our basic sample we had to restrict ourselves to lags of two in our AR model. Third order lagged dependent variables were tried but none were significant. As we noted above no attempt was made to identify structural parameters of the submodel by means of imposing overidentifying restrictions so the reduced forms presented were used directly to generate the R.E.

The seven equation are set out in (3.1)(a) to (3.1)(g) (standard errors are in brackets).

$$\text{(a)}\quad GB_t = \underset{(0.29)}{0.387} + \underset{(0.05)}{0.943}\, GB_{t-1}$$

$$s.e. = 0.033 \quad R^2 = 0.960 \quad \bar{R}^2 = 0.957 \quad h = -0.899$$

$$\text{(b)}\quad CB_t = \underset{(1.53)}{-3.139} + \underset{(0.19)}{0.502}\, CB_{t-1} + \underset{(0.25)}{0.600}\, Y_{t-2}$$

$$s.e. = 0.054 \quad R^2 = 0.871 \quad \bar{R}^2 = 0.855 \quad h = 1.84$$

$$\text{(c)}\quad P_t = \underset{(0.19)}{-1.160} + \underset{(0.05)}{0.772}\, P_{t-1} + \underset{(0.04)}{0.225}\, Y_{t-1}$$

$$s.e. = 0.055 \quad R^2 = 0.997 \quad \bar{R}^2 = 0.996 \quad h = 0.716 \qquad (3.1)$$

$$\text{(d)}\quad r_t = \underset{(2.66)}{-0.931} - \underset{(0.87)}{2.103}\, P_{t-1} + \underset{(0.14)}{0.876}\, r_{t-1} + \underset{(0.62)}{1.024}\, Y_{t-1}$$

$$s.e. = 0.085 \quad R^2 = 0.800 \quad \bar{R}^2 = 0.756 \quad h = -0.378$$

$$\text{(e)}\quad Y_t = \underset{(0.89)}{-0.65} + \underset{(0.22)}{0.550}\, Y_{t-1} + \underset{(0.24)}{0.542}\, Y_{t-2} - \underset{(0.06)}{0.081}\, r_{t-1}$$

$$s.e. = 0.028 \quad R^2 = 0.908 \quad \bar{R}^2 = 0.890 \quad h = -1.682$$

$$\text{(f)}\quad D_t = \underset{(1.59)}{2.446} + \underset{(0.20)}{0.545}\, D_{t-1} + \underset{(0.16)}{0.271}\, Y_{t-2}$$

$$s.e. = 0.055 \quad R^2 = 0.659 \quad \bar{R}^2 = 0.616 \quad h = 1.684$$

$$\text{(g)}\quad W_t = \underset{(0.50)}{0.403} + \underset{(0.18)}{1.531}\, W_{t-1} - \underset{(0.16)}{0.716}\, W_{t-2} + \underset{(0.07)}{0.189}\, Y_{t-1}$$

$$s.e. = 0.019 \quad R^2 = 0.976 \quad \bar{R}^2 = 0.971 \quad h = -1.359$$

Y_t and P_t are nominal GDP and a price level respectively. h is Durbin's h statistic.

Because the equations in (3.1) represent a reduced form it is hard to interpret the coefficients. However casual inspection of the system reveals that the model has the sort of dynamic responses that accord with our a priori view of the world. Lagged GDP plays a significant and sensible role in generating wealth, the supply of equities and dividends. This is also true in the equation for the price level where it picks up the pervasive influence of excess demand.

Taking the coefficients on Y_{t-1} and P_{t-1} in (3.1)(d) we see that, as would expect lagged real GDP ($Y_{t-1}-P_{t-1}$) has a downward effect on short ra Unfortunately we are still left with a separate negative influence of the price level on the rate which is contrary to our intuition. Rather we would expect higher prices to lead to an increase not a decrease in the rate. GDP and the quantity of gilts are explained well by their past although r_{t-1} has a small (and not very significant) negative influence on the former. We argue than that on the whole the sub model in (3.1) is sensible and a reasonable basis for the generation of R.E. of the exogenous variables as required by the main model.

(iii) The model of asset demands (the main model)

Because of the exogeneity structure of the asset demand functions, they have been normalised to leave the endogenous variables (asset prices) on the left hand side. Conditional on the parameters of the sub model the system in (2.2) was estimated using the programme CLARE. CLARE imposes the familiar 'forward' solution on the model. This is unique when the eigen values of the matrix premultiplying the R.E. vector in the quasi reduced form lie inside the unit circle. The estimates below satisfy this condition so that the imposition of this solution is ex post justified. Parameter estimates for the main model are laid out in equations (3.2)(a) and (3.2)(b).

$$\text{(a)}\quad PG_t = \underset{(1.05)}{5.833} - \underset{(0.01)}{0.266}\, PC_t + \underset{(0.01)}{0.028}\, PG^e_{t+1} - \underset{(0.03)}{0.141}\, PC^e_{t+1}$$

$$- \underset{(0.02)}{1.141}\, GB_t - \underset{(0.001)}{0.102}\, r_t + \underset{(0.03)}{0.950}\, W_t - \underset{(0.01)}{0.308}\, D_t$$

$$R^2 = 0.999 \quad D.W. = 1.64 \qquad (3.2)$$

$$\text{(b)}\quad PC_t = \underset{(5.12)}{-11.966} + \underset{(0.04)}{0.099}\, PG_t + \underset{(0.01)}{0.293}\, PC^e_{t+1} - \underset{(0.02)}{0.223}\, PG^e_{t+1}$$

$$- \underset{(0.02)}{1.615}\, CB_t + \underset{(0.001)}{0.008}\, r_t + \underset{(0.01)}{1.434}\, W_t + \underset{(0.04)}{0.719}\, D_t$$

$$R^2 = 0.999 \quad D.W. = 1.75 \qquad \text{Log of Likelihood Function} = 72.81$$

To interpret these estimates more easily we refer to table 7.2 where the parameters in the asset demand equatons (2.2)(a) and (2.2)(b) implied by (3.4)(a) and (3.4)(b) respectively are displayed.

Out of the sixteen structural parameters only three have perverse signs and two of these are of a negligible order of magnitude. The cross

Table 7.2

Parameter estimates of the asset demand model in equation (2.2)

EQUATION 1			EQUATION 2		
Parameter	Basic sample	Later sample	Parameter	Basic sample	Later sample
α_{11}	0.025	-0.071	α_{21}	-0.138	-0.556
α_{12}	-0.124	-0.155	α_{22}	0.181	0.731
c_{11}	-0.852	-1.425	c_{21}	-0.077	0.298
c_{12}	-0.357	-0.191	c_{22}	-0.438	-0.206
γ_{14}	-0.089	-0.151	γ_{24}	0.005	0.044
γ_{17}	0.833	1.228	γ_{27}	0.888	0.586
γ_{19}	-0.270	-0.004	γ_{29}	0.445	0.037
γ_{110}	5.112	-2.403	γ_{210}	7.409	0.650
D.W.	1.64	2.02	D.W.	1.75	1.78

price effects in both equations and the coefficient on r_t in the second equation are the parameters in question. The unimposed restriction that the coefficient on D_t be equal and opposite to that on PC_t is satisfied almost exactly in the second equation but not by any criterion is it satisfied in the first where the coefficient on PC_t is perverse. With these minor exceptions the coefficients are on the whole, sensible. We note that the own price effects are much larger than the effects of capital gains (the cross price effects are of course perverse). It is quite possible that the variance of expected capital gains are responsible for their discounting. The closeness to unity of the wealth terms ensure that stock holdings as a share of wealth are broadly stable as wealth grows (although the share contracts slightly with growing wealth).

All coefficients are significant by the asymptotic t-ratio criterion. The t-ratios are in fact, alarmingly large in certain cases. The wealth term in the second equation has the largest t-ratio at around 200 and two others are over 100. Whilst the information matrix was estimated numerically and not calculated exactly from analytical second derivatives we do not believe these numerical estimates to be poor. The estimated matrix was relatively insensitive to choice of steplength, dispelling worries in this quarter. The estimated variance covariance matrix of the quasi reduced form had elements of order of magnitude 10^{-4}. Taking expectations as data (as the quasi reduced form does) it seems that the order of magnitude of OLS t-ratios from this quasi reduced form would be the same as reported. In short then, it seems that these small standard errors simply reflect well defined parameter estimates.

Finally the Durbin-Watson (D.W.) statistics although in the inconclusive region are close enough to the upper bounds to be considered satisfactory. We must bear in mind that D.W. statistics have been found to be powerful against alternative hypotheses (forms of misspecification) besides that of first order error autocorrelation. All in all then, we believe that our estimates are satisfactory and plausible.

Before we move to a test of the R.E. hypothesis we thought it useful to reestimate our model for the data we have from 1972(1) to 1976(4). Because of the poor quality of the data in this period we treat these estimates with only passing interest. The estimates are tabulated in column B of table 7.2 and comparing these with column A we see that although all but three of the parameters are of the same sign and all but 5 the same order of magnitude, we would hesitate to describe the parameters as stable. The two constants and the coefficient on the own capital gain have shifted substantially. Further, the dividends terms in both equations are insignificant and do not satisfy the unimposed constraint vis a vis the coefficients on PC. The perverse coefficient on r_t in the second equation is also insignificant. Just how much of the parameter instability is due to our coarse model, how much is due to the events noted above in the sample and how much is a consequence of poor quality data (the gilts quantity index) is not clear. Because the focus of this paper is a test of the R.E. hypothesis we do not dwell on the issue here.

Finally, we performed a simple simulation exercise with the model using estimates from the basic sample into the later sample. For this purpose, one period ahead asset prices were regressed on the variables in the information set and predictions from these equations were used as proxies for the R.E. terms. This does not correspond to a consistent expectations forecast. However, bearing in mind that this exercise is of limited interest, the method is a quick and easy way to generate within sample unbiased forecasts.

TABLE 7.3

FORECAST RESULTS FOR BOND PRICES

TABLE 7.4

FORECAST RESULTS FOR EQUITY PRICES

The results are graphed in tables 7.3 and 7.4. These tables show clearly that whilst the predicted series have large residuals by the end of 1976 and whilst these residuals are all positive (PG) or all negative (PC), the major changes in the series are picked up fairly well and the dynamics seem well described. Undoubtedly having coefficients whose signs are not perverse contributes to the latter. Taken literally, however the forecasts are bad both series being in error by 100% at the end of the forecast period. Again we do not know where to lay the blame for this poor forecast.

To sum up then we are able to say that we have successfully estimated a coarse, simple but a priori sensible model of asset demands. The parameter estimates we have obtained whilst reasonable from an a priori point of view are not stable over the period 1972(1) to 1976(4). The forecast reinforces this view of instability whilst providing some comfort in the fact some of the out of sample dynamics are forecast by the model.

We now turn to the focus of this chapter; a test of the R.E. hypothesis.

7.4 A Test of the R.E. Hypothesis

(i) The null hypothesis

It is obvious from the nature of R.E. that no test of the hypothesis is independent of the model used to construct the test. In particular the simple hypothesis that expectations in asset markets are formed rationally is untestable as such. In this paper, however we have tried to develop a model that is as general as possible. The assumption of imperfect substitutibility between assets implicit in our model takes in the case where there is perfect substitutibility. Higher degrees of substitution will be reflected in larger own and across price responses and in the limit the total rate of return (current return plus capital gain) on all assets is forced to be equal. In this case our equations reduce to the familiar term structure equation always associated with efficient markets in the literature. Tests of the R.E. hypothesis have in the past been undertaken predominantly against such a term structure equation and, it is in this respect that the work in this paper is fundamentally different.

(ii) <u>The alternative hypothesis</u>

Wallis (1980) and Hoffman and Schmidt (1981) put forward likelihood based tests of the within and cross equation restrictions that R.E. impose on a reduced form and it is such a test that we adopt here. The problem with this procedure is that the interpretation or economic significance of the alternative hypothesese encompassed in the unrestricted reduced form is often rather vague. It may be the case that a reduced form containing the variables in the information set plus the current exogenous variables encompasses most or many alternative (expectations) hypothesese of interest but this is unlikely, especially in a small and simple static model such as the one we are considering. A particular alternative of interest to us here is that of adaptive expectations. In our model, this assumption in place of R.E. would imply a reduced form for (2.2) of

$$
\begin{aligned}
\text{(a)}\quad GB_t &= \theta(L)[\Pi_{11}+\Pi_{12}PG_t+\Pi_{13}PC_t+\Pi_{14}r_t+\Pi_{15}W_t+\Pi_{16}D_t] \\
&\quad +(1-\theta_2 L)^2. \\
&\quad (1-\theta_1)^2 PG_{t-1}+(1-\theta_1 L)^2(1-\theta_2)^2 PC_{t-1} \\
&\quad +(1-\theta(L))GB_t+\theta(L)u_t \\
\text{(b)}\quad CB_t &= \theta(L)[\Pi_{21}+\Pi_{22}PG_t+\Pi_{23}PC_t+\Pi_{24}r_t+\Pi_{25}W_t+\Pi_{26}D_t] \\
&\quad +(1-\theta_2 L)^2(1-\theta_1)^2 PG_{t-1}+(1-\theta_1 L)^2(1-\theta_2)^2 PC_{t-1} \\
&\quad +(1-\theta(L))GB_t+\theta(L)u_{2t}
\end{aligned}
\qquad (4.1)
$$

A glance at our sub model in (3.1) reveals that even ignoring the M.A. error process our unrestricted reduced form will include no lagged price terms so that we cannot expect our test to have much power against this key alternative hypothesis. To test against adaptive expectations would require a non-nested test and it is not clear how if at all such a test could be computed.

One hypothesis of interest in our set of alternatives is centred on the idea originally due to Fischer that (in the long run) the real rate of interest was equal to a constant. This has led to a view that expectations of future bond and equity prices were formed from the current view of long term inflation (see for example, Frankel (1982)). This in turn has led to specifications which have lagged price level terms to proxy for such expectations. Because the price level enters our submodel its lagged values enter the unrestricted reduced from so that we expect our test to have good power against this particular alternative.

Returning to the introduction of this paper we suggested that the proper role for lagged stocks in an asset model is as part of a prediction process. Because lagged adjustment is rife in the asset markets literature it would be most interesting if this view was rejected in favour of R.E. Typically, expectations are subsumed in stock adjustment models so that our unrestricted reduced form contains all the variables relevant to this alternative model.

Before we proceed with the test we note another problem of the proposed test, namely that it does not allow the overidentifying restrictions on the quasi reduced form (the traditional reduced form that contains only predetermined, exogenous and expectations terms) to be tested separately. Fortunately our model is just identified by the standard criteria, that is, treating expectations terms as exogenous variables, the structure is just identified by the standard conditions and so the problem does not arise.

Moving to the test itself, we chose a likelihood ratio x^2 test as this was the most convenient likelihood based test available. The number of terms in the unrestriced reduced form (excluding elements of the error variance covariance matrix as this was unrestricted) was 30 and the number of parameters under the null was 16 leaving 14 restrictions and a x^2 statistic with 14 degrees of freedom. More explicitly we have

H_0:- The model 4(a) and 4(b)

H_1:- $y_t = \Pi_0 x_t + \Pi_1 x_{t-1} + \Pi_2 \bar{x}_{t-2}$

where $y_t = \begin{bmatrix} PG_t \\ PC_t \end{bmatrix}$, $x_t = \begin{bmatrix} GB_t \\ CB_t \\ P_t \\ r_t \\ Y_t \\ u_t \\ D_t \end{bmatrix}$ $\bar{x}_t = \begin{bmatrix} Y_t \\ W_t \end{bmatrix}$ constant

and Π_0, Π_1 and Π_2 are unrestriced.

The value of the x^2_{14} statistic for the test was 15.36. When compared with a critical value of 23.69 (5% tail) the restrictions escape rejection by a very comfortable margin. Indeed the probability of getting a x^2 value as low as ours is about 0.5.

In sum then we have tested our R.E. model against a general alternative and the results as far as the R.E. hypothesis is concerned are very encouraging.

7.5 Summary and Conclusion

In this chapter we have described estimated and tested a simple static R.E. model of aggregate asset demands. Whilst the model does not provide good forecasts of asset price levels outside its sample, it does produce a forecast that mimics the dynamic behaviour of the data quite well. A test failed to reject the restrictions implied by the hypothesis with a large margin of comfort. Whilst this test could be expected to have good power against certain alternatives of interest it is unlikely to have any power against adaptive expectations and so our success must be qualified and put in perspective.

We end with a note on the information set. Because it is an equilibrium model where prices clear markets, the latter must be observed by agents at the time of the market transactions. More properly then, our information set should have been all the exogenous variables dated at time t-1 plus the current asset prices. In the annex we derive the minimum mean square error predictor in this case. It is clear from this that if the extra information (current prices) is to be of use then the variance covariance matrix of both structural errors and of exogenous variable process errors must be known. The assumption that we have implicitly made in choosing the R.E. predictor is that the latter matrices are unknown (or at least are subjectively highly uncertain). Whilst this assumption may by unpalatable (especially since other unobservables, namely the parameters of the model are assumed known) estimation incorporating this predictor is a considerably more complex exercise.

APPENDIX

Consider the model

$$y_t = A(y^e_{t+1}|\Omega_t^*) + Bx_t + u_t; \tag{A1}$$

$$x_t = Px_{t-1} + E_t$$

where u_t ()$N(0,\Sigma_u)$ and ε_t ()$N(0,\Sigma_\varepsilon)$, Ω_t^* is an information set which includes x_{t-1-i}, y_{t-i} $\forall i \geq 0$. y_t and x_t are gx1 and kx1 vectors and A, B and P are all known conformable parameter matrices.

Solving (A1) forwards gives

$$y_t = \sum_{i=1}^{\infty} A^i(x^e_{t+i}|\Omega_t^*) + Bx_t + u_t \tag{A2}$$

Denoting

$$\hat{x}_t = x_t^e|\Omega_t^*$$

we can using the commutative law of expectations write

$$x^e_{t+i}|\Omega^* = E\{(x^e_{t+1}|\hat{\Omega}_t)|\Omega_t^*\} \ \forall i \geq 0 \tag{A3}$$

where $\hat{\Omega}_t$ is an information set containing $\hat{x}_t, x_{t-1-i}$ $\forall i \geq 0$

(A3) states quite simply that y_t is only of help in predicting x_{t+1} because it helps to predict x_t. Using (A3), (A2) becomes

$$y_t = \hat{x}_t + Bx_t + u_t$$

where

$$\mathrm{vec}F = \mathrm{vec}\sum_{i=1}^{\infty} A^iBP^i = (I-P'\otimes A)^{-1}\mathrm{vec}ABP^2$$

To derive a form for $\hat{x}_t$ we combine the reduced form of (A4) with the exogenous variable processes to make a state space model.

Because the model is linear the minimum mean square error estimator (MMSEE) of $\hat{x}_t$ will be linear in y_t and x_{t-1}. The vector x_t itself will appear in the observable reduced form because it enters the structure (A1) independently of any expectations mechanism. The observable reduced form can therefore be written as

$$y_t = \Phi x_{t-1} + \theta x_t + u_t \tag{A5}$$

where $u_t = Lu_t$ and $E(u_tu_t')$ is denoted as Σu. Combining (A5) with the exogenous variable processes gives us the familiar state space measurement and transient equations

$$\begin{bmatrix} x_t^e|x_{t-1} \\ \bar{y}_t \end{bmatrix} = \begin{bmatrix} I \\ \theta \end{bmatrix} x_t + \begin{bmatrix} \varepsilon_t \\ u_t \end{bmatrix}$$

where $\bar{y}_t = \bar{y}_t - \Phi x_{t-1}$
and

$$x_t = Px_{t-1} + E_t \quad (A7)$$

respectively.

(A6) states our objective; to measure the state variable x_t using the information set x_{t-1} and y_t. (A7) simply describes the motion of the state variable through time.

The MMSEE of x_t is just the Generalized Least Squares estimate of x_t (see Harvey (1981) pp.108-109). This is

$$\hat{x}_t = Px_{t-1} + K(\bar{y}_t - \theta x_t^e | x_{t-1}) = Px_{t-1} + K(\bar{y}_t - \theta Px_{t-1}) \quad (A8)$$

where K is analogous to the Kalman Filter Gain matrix and is written as

$$K = \Sigma_\varepsilon \theta'(\theta \Sigma_\varepsilon \theta' + \Sigma_u)^{-1} = \Sigma_\varepsilon \theta' \Sigma_{RE}^{-1}$$

Σ_{RE}, of course is the variance covariance matrix of the R.E. error in predicting y_t from x_{t-1-i} ($i \geq 0$).

Combining (A8) with (A4) gives

$$y_t = (I-FK)^{-1}F(I-K(\Phi+\theta P))x_{t-1} + (I-FK)^{-1}Bx_t + (I-FK)^{-1}u_t \quad (A9)$$

Equating coefficients with those of (A5) gives a solution for the parameter matrices Φ, θ and L. The general form of these solutions is uninformative but it is clear that the nonlinear within and cross equation restricitons on the reduced form are different from those of the R.E. case. The two predictors coincide in the limit as the largest element of Σ_u^{-1} tends to zero i.e. when the 'noise' contaminating the signal becomes large.

8. CONCLUSION

8.1 A summary of the main results and their significance

Undoubtedly the main contribution of this thesis is the simplification (and potential cost reduction) of FIML estimation of R.E. models. The approach described in chapter six is we believe superior (for linear models) to others in terms of cost and computational efficiency. More important perhaps is the degree of certainty with which our method delivers true FIML estimates. An approach that relies on numerical differentialtion may fail if the likelihood function is not sufficiently 'well-behaved' in the parameter space. If for example there are discontinuities or 'flats' in this space then there is no guarantee that such routines will converge to FIML estimates if they converge at all.

We have demonstrated that our approach is feasible by writing and using a programme (CLARE) designed to estimate the structural parameters of a particular class of linear R.E. models. The programme was then successfully applied to the data to estimate the parameters in a simple model of financial asset demands.

More important than these estimates was the test of the R.E. hypothesis in a financial assets model that the programme allowed us to carry out. There is a voluminous literature concerning R.E. in financial markets but virtually all the focus of attention has fallen on testing the joint hypothesis of R.E. and efficient markets. The popularity of such models we believe lies in the ease with which they allow estimation and testing to be carried out. Our empirical model, however, incorporates the efficient markets assumption as a special case and so our approach is more general than that found in the literature. Furthermore we took care to ensure that our test has reasonable power against alternative hypothesese of interest in contrast to the literature where this issue is largely neglected. The test, a likelihood ratio test, was passed with a large margin of comfort providing support for the R.E. hypothesis that is far more convincing than the support that has been offered in the literature in the past.

The results concerning simulation of R.E. models whilst of less practical importance are nonetheless quite informative. The universal adoption of terminal conditions by forecasters as a means of making the solution of an R.E. model determinate is undoubtedly due to the simplicity and low cost of the procedure. In showing that there is no theoretical justification for using terminal conditions we have sounded a clear warning to model users that a model solution chosen in this way will be arbitrary.

By contrast the work on the Fair-Anderson simulation method is more positive. Any model user wishing to simulate standard models under the assumption of R.E. will be able to assess the degree of bias he can expect in his policy multiplier estimates by looking at the form(s) of the O.E.P.(s) of the expectations variables. If they are close in M.S.E. to the extrapolative proxies used for estimating the parameters of the standard model then these biases are unlikely to be serious. Forecasters are also warned, however, that multipliers to the unanticipated components of policy are likely to be seriously biased in all cases.

The small amount of original work in chapter 2 is of passing interest only. Here we have shown that the simulation of theoretical models in the manner proposed by Fair (1974) is an arbitrary way to proceed. This is not to say that theoretical models should not be explored using numerical methods. Rather that the assumptions of the experiments and in particular the manner in which the model's coefficients were chosen should be made clear and should be supported by reasoned argument.

Finally, the results in the same chapter on Lipsey and Parkin's incomes policy study is of some historical interest. These results suggest that simultaneity bias in the intercept and slope dummies, which were included to measure the impact of the policy may have led the authors to underestimate the influence of a prices and incomes policy on the wage price relationship.

8.2 Unfinished work

It was never our objective to address ourselves to all of the problems that the existence of R.E. poses for policy analysis and although we believe we have approached the most important ones, there are still problems that this thesis has not fully resolved. In particular the procedure advanced in preference to the use of terminal conditions in chapter four is expensive, time consuming and not applicable to nonlinear models. With regard to the latter it is not clear whether or not there is a parameterisation of the complete set of solutions to a nonlinear model analagous to that of a linear model. One answer to the problem may be to pass the solution through a point considered 'reasonable' by the modeller (it is not unknown for modellers to use their judgement in similar circumstances). There is no compulsion, however, to make these terminal points equilibrium or steady state points. Whenever used a sensitivity analysis should be conducted to see if the choice of terminal values (or terminal data) is of any significance. We believe that in many cases it will not be important (such cases in linear models correspond to those where the forward solution is uniquely stationary).

One development of the thesis which begs for further work is the implementation of the results on estimation in chapter six. The programme CLARE was written under the burden of research time and computer resource constraints. It is therefore applicable only to a particular class of models and is not as computationally efficient as it could be. Software that efficiently solves the first order conditions of the likelihood function for the general case would be of great benefit to empirical economists. However, further research into the best (in terms of computational efficiency) means of solving these nonlinear equations would be required. For example, we believe that the Gauss-Seidel technique adopted for CLARE was relatively inefficient and that an investigation of an alternative, say Newtonian, method would be worthwhile.

As we suggested in the introduction to the thesis the development and use of powerful tests of the R.E. hypothesis is still a relatively untouched area. It does seem alarming that the hypothesis is so widely accepted and imposed in all sectors of econometric models with so little empirical support to back up its use. Again the possibility of cheap FIML estimation opened up by this thesis (conditional on the software being written) should encourage more empirical work in this area.

We recognise that the estimation and testing procedure we have advanced is not applicable to nonlinear models. In these cases we would recommend that as a prelude to the incorporation of R.E. into a particular sector of an econometric model that a small and simple linear model of the sector be built, estimated and tested as we have done for the financial sector in chapter six. Many large modelling schools maintain a small linear condensed form of their model so that this procedure should not raise many difficulties. Armed with this empirical support for the hypothesis the modeller may then proceed with a clearer conscience to incorporate the assumption into his model using standard limited information estimation techniques.

References

Anderson, L.C. and Jordan, J.L. (1968) "Monetary and Fiscal Actions: A Test of Their Relative Importance in Economic Stabilisation" The Federal Reserve Bank of St. Louis Review, Vol 50.

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

Barro, R.J. (1977) "Unanticipated money growth and unemployment in the United States", American Economic Review, vol. 67 (March), pp. 101-15.

Blanchard, O. (1980) "Backward and Forward Solutions for Economies with Rational Expectations", American Economic Review.

Bodkin, R.G. (1966) "The Wage, Price, Productivity Nexus", Philadelphia.

Brainard, W.C. (1967) "Uncertaintly and the Effectiveness of Policy", American Economic Review, Volume 62 (2).

Buiter, W.H. (1979) "Feedback and the use of current information: The use of general linear policy rules in rational expectations models", Bristol Universtiy Working Paper.

Burmeister, E. Flood, R. and Turnovsky, S. (1978), "Rational Expectations and Stability in a Stochastic Monetary Model of Inflation", mimeo, University of Virginia.

Calvo, G. (1977) "On the Indeterminacy of Interest Rates and Wages with Perfect Foresight - Some Examples", mimeo, Columbia University.

Chow, G.C. (1976) "The Analysis and Control of Dynamic Economic Systems", Wiley.

Chow, G.C. (1981) "Maximum Likelihood Estimation of Rational Expectations Models", Princeton University, Econometric Research Program Memorandum.

Desai, M. (1975) "The Phillips Curve: A Revisionist's Interpretation". Economica.

Driffill, E.J. (1980) "Time inconsistency and 'rules vs. discretion' in macroeconomic models with rational expectations", Universtiy of Southampton Working Paper.

Doob, J. (1953) "Stochastic Processes", Wiley.

Fair, R.C. (1976) "A Model of Macroeconomic Activity. Volume II: The Empirical Model", Ballinger.

Fair, R.C. (1979) "An analysis of a macroeconometric model with rational expectations in the bond and stock markets", American Economic Review, 69, 539-52.

Fair, R.C. and Taylor, J.B. (1983) "Solution and Maximum Likelihood Estimation of Dynamic Nonlinear Rational Expectations Models", Econometrica, Volume 51.

Frankel, J.A. (1982) "A technique for extracting a measure of Expected Inflation from the Interest Rate Term Structure", The Review of Economic Studies.

Goldfeld, S.M. and Blinder, A.S. (1972) "Some Implications of Endogenous Stabilisation Policy", Brookings Papers on Economic Activity.

Goodhart, C.A.E. (1984) "Monetary theory and practice: The U.K. experience", Macmillan.

Gourieroux, C., Laffont, J.J. and Monfort A. (1982), "Rational Expectations Models - Analysis of the Solutions", Econometrica.

Granger, C.W.J. and Morris, M.J. (1976) "Time series modelling and interpretation", Journal of the Royal Statistical Society A, 139, 246-257.

Hall, R. (1978) "The Macroeconomic Impact of Changes in Income Taxes in the Short and Medium Runs", Journal of Political Economy.

Hoffman, D.L. and Schmidt P. (1981) "Testing the restrictions implied by the rational expectations hypothesis", Journal of Econometrics.

Holly, S. and Beenstock M. (1980) "The Implications of Rational Expectations for the Forcasting and the Simulation of Econometric Models", Economic Forecasting Unit Discussion Paper.

Kydland, F.E. and Prescott, E.C. (1977) "Rules rather than discretion: the inconsistency of optimal plans", Journal of Political Economy, vol. 85 (June), pp. 473-91.

Lipsey, R.G. and Parkin, J.M. (1970) "Incomes Policy: A Re-appraisal", Economica.

Lucas, R.E., Jr. (1972) "Expectations and the neutrality of money", Journal of Economic Theory, 4, 103-24.

Lucas, R.E., Jr. (1973) "Some International evidence on output-inflation trade offs", American Economic Review, 63, 326-34.

Lucas, R.E., Jr. (1976) "Econometric policy evaluation: A critique", in The Phippips Curve and Labor Markets, ed. by K. Brunner and A.H. Meltzer, Amsterdam: North-Holland, 19-46.

Lucas, R.E. (1975) "An equilibrium model of the business cycle", Journal of Political Economy, vol. 83 (December), pp. 1113-44.

McCallum, B.T. (1976) "Rational Expectations and the Estimation of Econometric Models: An Alternative Procedure", International Economic Review, Volume 17.

McCallum, B.T. (1979) "The Current State of the Policy Ineffectiveness Debate", American Economic Review.

Minford, A.P. and Matthews, S.K. (1978) "A note on Terminal Conditions and the Analytic Solutions of Rational Expectations Models", L.I.T.P. Working Paper No. 7805.

Mirrlees, J. (1968) "Multi-Sector Growth Models", Lecture Notes for the M. Phil Degree at Oxford.

Moroney, J.R. and Mason, M. (1971) "The Dynamic Impacts of Autonomous Expenditures and the Monetary Base on Aggregate Income", Journal of Money, Credit and Banking.

Muth, R. (1960) "Optimal Properties of Expontially Weighted Forecasts", Journal of the American Statistical Association.

Nelson, C.R. (1975) "Rational expectations and the estimation of econometric models", International Economic Review, 16, 555-61.

Pagan, A.R. (1982) "Rational Expectations and the Theory of Policy" in "The Theory of Economic Policy: Statics and Dynamics", by Preston, A.J. and Pagan, A.R.

Parkin, M. (1970) "Discount house portfolio and debt selection", Review of Economic Studies.

Pesaran, M.H. (1982) "Identification of Rational Expectations Models", Journal of Econometrics, Vol.16.

Rao, R.C. (1965) "Linear Statistical Inference", Wiley.

Revankar, N.S. (1980) "Testing of the Rational Expectations Hypothesis", Econometrica, Volume 48.

Prothero, D.L. and Wallis, K.F. (1976) "Modelling macro-economic time series (with discussion)", Journal of the Royal Statistical Society, A, 139, 468-500.

Sargent, T.J. and Wallace, N. (1973) "Rational Expectations and the Dynamics of Hyperinflation", International Economic Review, Volume 4.

Sargent, T.J. and Wallace, N. (1975) "Rational xpectations, the optimal monetary instrument and the optimal money supply rule", Journal of Political Economy, vol. 83 (April), pp. 241-54.

Schmidt, P. and Hoffman, D.L. (1981) "Testing the Restrictions Implied by the Rational Expectations Hypothesis", Journal of Econometrics, Volume 15.

Taylor, J. (1977) "Conditions for Unique Solutions in Stochastic Macro-Economic Models with Rational Expectations", Econometrica, Volume 45.

Theil, H. and Bogaard, P.J.M. Van den. (1968) "An application of strategy and certainty equivalence concepts to the economy of the U.S., 1933-1936", in Macrodynamic Policymaking, Zellner A. (ed.)

Tinbergen, J. (1952) "On the Theory of Economic Policy", North-Holland.

Wallis, K.F. (1980) "Econometric implications of the rational expectations hypothesis", Econometrica.

Wallis, K.F. (1971) "Wages, Prices and Incomes Policies: Some Comments", Economica.

Weale, M. (1984) "A simple disaggregate model of the U.K. financial sector", Cambridge University, GPP no.540.

Wickens, M. (1982) "The Efficient Estimation of Econometric Models with Rational Expectations", The Review of Economic Studies.