Report on modelling and forecasting at the Bank of England

Report to the Court of Directors of the Bank of England on the modelling and forecasting systems within the Bank, prepared by Adrian Pagan.[1]

Executive summary

Models and forecasts are important inputs into any decision-making process whether it relates to business or to monetary policy. In the UK context the model used in the monetary policy process needs to incorporate the views of the MPC about the way the economy functions, ie to be theoretically coherent, and also be able to replicate historical data on the UK economy, ie to be empirically coherent. It is hard to achieve both of these simultaneously and some trade-off needs to be made when selecting the model.

The report documents a range of models that have been or are being used in central banks and that resolve the trade-off identified above in different ways. Which of these models is selected will ultimately depend on the preferences of the MPC, but it is important that, for any degree of theoretical coherence, the degree of empirical coherence should be maximised. In the report this is characterised as the desirability of being on the frontier that shows the best possible combinations of theoretical and empirical coherence that are attainable, rather than being inside it.

At this point in time two core (or key) models exist in the Bank—what I have termed the macro model (MM) and the new macro model (NMM). The NMM is currently still under development. In my judgment the MM is not at the frontier. Its structure does not fully accord with the MPC's beliefs about the functioning of the UK economy and it has some well-documented difficulties in matching historical outcomes for a set of variables such as inflation and GDP growth. The NMM is likely to score more highly with regard to theoretical coherence but evidence on its empirical coherence was not available at the time of the writing of this report. Because the MM

is off the frontier there is a strong argument for replacing it with some alternative model.

Many attempts have been made to improve the empirical coherence of the MM. Bank staff have been very active in modifying the equations of the MM for this purpose. Some solutions have emerged but the report suggests that these have not completely resolved the issue of empirical coherence, particularly with regard to the inflation process. Problems in predicting inflation have been a worldwide problem in the mid to late 1990s and it seems that quite new perspectives may be required in order to produce good predictions of it from a model.

The level of technical proficiency displayed by the modelling and forecasting teams in the Bank is already very high and I have made only a few minor suggestions about how it might be augmented.

The decision to proceed with the development of the NMM may have had the (possibly unintended) consequence of diverting resources from the maintenance of the MM and, more importantly, meant that the response to failures of the equations of the MM was to make relatively small changes to them rather than to explore alternative paradigms.

The forecasting process is distinct from the modelling process. It involves a series of technical adjustments to correct difficulties experienced when using the equations of the core model for forecasting. These relate to past performance but also aim to anticipate difficulties that might occur in the forecast period. This system now seems to be working very well in the Bank and has had some noticeable success in correcting some of the difficulties encountered when forecasting with the MM model.

Some variables, particularly exchange rates, are very difficult to forecast, and a range of alternative methods has been experimented with. I feel that the methods used by the Bank in this context are 'state of the art'.

Although the Bank has a core model that is used for most policy analyses and forecasting, it recognises that such a model cannot handle all the situations that arise in actual decision-making, eg the impact on costs of 11 September 2001. Special models need to be developed to deal with such events. It also may be desirable to look at the evidence from a range of models rather than a single one when assessing policy options. This leads to the desire to have a diverse set of models, or what the Bank refers to as a 'suite of models'. In my opinion there is a good diversity of models for policy analysis within the Bank but I express some concern over whether this is true when it comes to the task of forecasting. A large number of auxiliary models are used in the latter activity, but they have tended to derive from a prior use in investigating policy issues rather than being selected for their suitability for informing the forecast. Consequently, I recommend that more attention be paid to the selection of the suite of forecasting models.

There has been much criticism of the fact that forecast errors in two year ahead forecasts have been consistently of the same sign and that there has been a 'bias' in these forecasts. In the report I analyse these outcomes and show that one should expect runs of the same sign in forecast errors and that the 'bias' is probably as small as one could reasonably expect.

The periods of time in which the forecast errors were worst coincide with the times in which there was an unusual pattern to MPC forecasts, namely one in which expected inflation and GDP growth were negatively rather than positively related. This suggests that greater attention should be paid to joint outcomes of inflation and GDP growth. At present the fan charts give this information separately on each variable and I would recommend that it should be an objective to produce fan charts showing their expected joint outcomes.

The Bank has been quite sensitive to the need to perform ex-post forecast evaluation. Analysis that has been presented to the MPC has ranged from summarising the outcomes to attempting to ascertain the reasons for the errors, eg by decomposing the forecast errors for inflation in terms of the forecast errors for the influences on inflation of earnings, the exchange rate, etc. I feel that the work in this area has been of high quality and certainly of adequate quantity.

Many problems have had to be faced and solved in setting up a modelling and forecasting system that would adequately serve the unique structure of the UK monetary policy decision process. It was inevitable that some difficulties would arise with the initial solutions and that modifications would need to be made. The current system seems to be working very well.

Terms of reference

In October 2001, I was asked by the Court of Directors of the Bank of England to prepare a report for them on the modelling and forecasting systems within the Bank. The terms of reference given to me were:

'The reviewer shall provide for the Court of the Bank a report on the statistical and economic modelling and forecasting work carried out by the staff of the Bank for the MPC and evaluate whether that work is 'state of the art'.' The review should in particular:

- focus on the technical aspects of the modelling and forecasting process, rather than the procedural and presentational issues addressed by the Kohn Report, and judged against the purposes set out for the monetary policy regime;

- cover the full range of modelling and forecasting approaches presently employed by the Bank and note where these methods lag behind best practice or are capable of improvement;

- identify any additional techniques or approaches that could usefully be employed; and

- evaluate the procedures for ex-post forecast evaluation.

During the period from October 2001 until June 2002 I visited the Bank on a number of occasions, interviewing many of the people involved in the forecasting and modelling process and attending meetings concerned with those activities. I also spoke to all members of the MPC about their impressions of the process. I would like to thank all those who responded to my requests for information and for going over the procedures in some detail.

Structure of review

To structure my review I found it useful to conduct the discussion in terms of six themes, rather than organising it specifically into the categories of the terms of reference. These are:

1 Representing the economy.
2 Modelling the UK economy.
3 Projecting the UK economy.
4 Diversification of representations and projections.
5 Assessing the quality of projections.
6 Communicating the models and projections.

Interpretation of terms of reference

To begin the review, it is necessary to interpret some of the terms of reference. Implicit in them is a recognition that it may be important to divorce the questions of how one constructs a representation of the economy from the way in which projections are built up. An important argument in favour of the divorce is that monetary policy makers often need to use models to enhance their understanding of an economy, and what the historical data say about interactions within it, rather than just using them to make a projection. Deciding on a useful representation is generally referred to as the modelling process, while the way in which a projection is done is generally termed the forecasting process. Though the term 'forecasting' is technically inaccurate, as it implies an unconditional statement about future outcomes, and so makes no reference to any assumptions that underlie the statements, eg concerning exchange rate behaviour, we will tend to use the terms projection and forecasting interchangeably, since it is such common usage.

A further element in the terms of reference that needs some clarification is the mention of 'state of the art'. Inevitably this demands a reference point. Three possibilities suggest themselves—relative to what is being done in academia, relative to what is being done in other central banks, and relative to what might be done given the constraints that are placed upon the Bank of England by the way in which monetary policy decisions and projections are made in the United Kingdom. The latter process seems to be unique in the world. Policy decisions are made by a committee that is composed of both executive and non-executive members of the bank (the 'internals' and 'externals' respectively). Moreover, under the legislation setting up the MPC, the projections that are recorded in the *Inflation Report* are those approved by the MPC; in practice this requirement has been met by having the projections as being those of the MPC rather than the staff of the Bank of England. The fact that this institutional structure is very different from others means that procedures have often evolved to deal with this fact. 'State of the art' might therefore mean that the processes should be such that the client—the MPC—is satisfied that the processes enable it to perform its tasks effectively. I suspect that one needs to look at the processes from all of these angles, and will do so in the report, although the third angle mentioned above seems to be the appropriate standard for final evaluations.

Models and decisions

To make decisions about the direction of a business requires information. Today the process of constructing that information often starts with a representation of the activities of the business in a spreadsheet that summarises its financial accounts. This spreadsheet is often termed a model. Often there is a 'core' spreadsheet which is linked to other spreadsheets, with the latter providing a greater degree of disaggregation of a given item, such as sales into different regions or branches etc. The 'core' spreadsheet would rarely stay the same over long periods of time, since the business will take on new activities, and new competitors will arise, so that the 'model' of the business to be found in that spreadsheet will need to adapt to the new environment.

The spreadsheet provides a large amount of information to the decision-makers concerning the current state of the business. However, when making decisions about the future direction of the business, it will be necessary to project the items in the core spreadsheet. This will generally be done through a number of other spreadsheets, each of which describes rules for the projection of each of the items. Such rules need to recognise that there are interrelationships between many of the columns of the spreadsheet, eg if the business has many products that are substitutes, it may be that a rise in the sales of one product line will mean a decline in the others. Moreover, in making any projections, account has to be taken of competitors' reactions, and allowances must be made for things not under the control of the company, such as developments in the macroeconomy, exchange rate movements and tax rates that might be levied by governments.

The spreadsheet numbers that come from this process are of course just a body of information into the decisions that are to be made. There is no automatic rule that maps the numbers into a decision, although it is quite likely that there would be some predictable reactions by the decision-makers to certain outcomes, eg a projected drastic decline in sales of a product could lead to a cancellation of that line.

The monetary policy decision process corresponds quite closely to the above description. To make decisions, the monetary policy makers need a representation of the entity that they are attempting to direct. Since this is the macroeconomy, the primary representation comes from the national accounts, and so it focuses upon the major aggregates such as consumption, investment, exports etc. But many other series may be linked into these elements, eg consumption may be disaggregated in many ways. Moreover, in the same way that the spreadsheet model of a business may vary over time in response to changes in a business and its environment, so too may the models employed by a central bank.

Monetary policy decisions also require a projection. This projection needs to recognise the interrelationship between many of the series and also has to take account of the reactions of the private sector to actions of the monetary authority, particularly through a change in its expectations. Finally, just as in any business decision, there are many things that are outside the direct control of the monetary authority; examples being the exchange rate and the international economic situation, although monetary actions may influence both to some extent. Consequently, separate projections of these influences must be made.

The similarity in process and structure means that there are also many similarities in the way that decisions are made in business and by monetary policy makers. In particular, in both cases the models and projections are meant to inform the decision-maker rather than being automatic determinants of the decision. But there are also some significant differences.

First, most businesses have good information on the current state of revenues and costs, so that the core spreadsheet will contain up-to-date information on the activities of the business. For a monetary policy maker this is rarely true. The quarterly national accounts are published with a substantial lag. This means that items such as current-quarter GDP may not be known at the time of a monetary policy decision. Indeed, the outcome for a number of past quarters leading up to the point in time at which a decision on monetary policy is to be made may not be known. Instead there will be a large amount of partial information upon such items, and an important task for central bank staff will be to 'fill in' the gaps. It is also the case that the quarterly national accounts are prone to revision and sometimes a model can be a useful device for assessing whether a particular revision needs to be treated with a good deal of caution.

Second, the degree of disclosure of both the projections and the reasons for decisions is very different. Company directors certainly face the discipline of the market, but they only formally appear before shareholders once a year. In contrast, a central bank continuously publishes

a large amount of information about its projections and its thinking about what is happening to the macroeconomy. Moreover, in the UK case, the Governor and other MPC members are generally subject to questions by committees of the Houses of Parliament a number of times a year.

Finally, the risks that surround a projection in the eyes of the directors and chief executive officer of a company are rarely disclosed to either shareholders or the market in any precise way. In contrast, at least for the Bank of England, projections are presented in each of the quarterly *Inflation Reports*, minutes of the decision-making meetings are published and risks to the projections are quantified in fan charts for inflation and GDP growth.

The brief description of decision-making given above highlights the need to build models and to make projections. Hence it is appropriate that the terms of reference I was given focused upon these aspects and their effectiveness for the decisions on monetary policy to be made by the MPC.

Representing the economy

Representing the economy requires setting up some modelling system. There are basically three components to such a system:

1 Deciding on a model of the economy.
2 Setting up processes for reviewing and modifying the components of the model.
3 Creating processes for deciding whether a new model should be entertained and formulated.

A 'state of the art' modelling system needs to address each of these items and we will consider each in turn. Before doing so, a brief history of modelling the economy is needed in order to classify the core models in use at the Bank of England.

Modelling strategies

The design of a model is like the design of a journey. One could simply meander around the countryside with the vague idea of arriving at a given destination, making choices about routes as one goes along. Alternatively, one could have an idea of a region that one might like to visit, say Provence, have a map in one's head of how one should get there, and then use sign posts as one goes along to achieve that objective. Greater degrees of precision can be applied in the choice of destination, eg

Aix-en-Provence rather than Provence, and more forward planning of the route to be followed can be done. In the ultimate scenario, one could imagine using an on-board computer to provide one with an optimal route that would minimise travel time given the information fed into it. These two elements, the specification of a destination and the selection of a method to get there, are the key factors that appear over and over again in the history of model design.

How should a model be constructed that is intended to be a good representation of those features that are of primary interest to policy-makers, such as the MPC? Under the MPC's remit the variable of primary interest is the inflation rate, but the level of economic activity clearly plays an important role in its thinking. Consequently, one needs to ask how one is to build a model of these two variables. When the first macroeconometric models emerged this seemed a rather straightforward question. With regard to economic activity, the national accounts provided a set of identities linking items such as GDP, consumption, investment and the trade balance. On the price side, following a long tradition in applied microeconomics, the price of domestic goods could be regarded as a mark-up over variable costs, while the price of imported goods reflected international prices and the exchange rate. Together these two prices combined to produce an aggregate price level. Thus it was envisaged that one would simply write down the equations underlying those series that were the building blocks of GDP and the price level, and then introduce extra variables as needed, eg the mark-up might vary with the state of the economy, making it necessary to measure the latter variable. This strategy meant that even more equations needed to be set out, so as to explain the variables included at an earlier stage. Therefore, starting from the original two variables that were to be explained—economic activity and inflation—such a modelling strategy meant that many more variables needed to be modelled. By the late 1960s the approach had led to models with hundreds and, sometimes, thousands of equations. In terms of a journey this would correspond to the meandering strategy.

Despite the misgivings of some, these models seemed to work rather well, and it was not until the early 1970s that they began to produce poor forecasts. A number of reasons were then suggested for this outcome. One blamed the oil price shocks of 1973, since, though the models had an extensive set of equations for describing demand, the other half of the economist's 'scissors',

supply, was either not present or was present in a very rudimentary way. Proponents of this view of the cause of forecast failure therefore responded by expanding the already large models to incorporate a supply side. This took many forms, of which the simplest was just to recognise that there were constraints placed on output by the quantity of factors in the economy, and so an expansion of demand did not lead to a rise in output unless extra factors could be found and mobilised. If this was not possible the demand dissipated, either into higher prices or into a deficit on the current account. Today such constraints are present in all macro models used for policy. Indeed, there is probably no single concept that is as much debated and measured in a myriad of ways by central banks as the 'output gap', a quantity designed to measure the state of excess demand.

A radically different interpretation of the forecast failure was that the breakdown of the models was caused by shifts in the parameters of the model as a result of the private sector changing the way it formed expectations about future developments; models that ignored such reactions would then seem to be unstable and, presumably, predict poorly. More specifically, it was those elements of econometric models that represented the dynamic responses to events which would be expected to change. This interpretation became known as the 'Lucas critique'. Its prescription was that microeconomic theoretical foundations needed to be invoked to explain expenditure choices and pricing behaviour in the macroeconomy. In practice, this was interpreted as meaning that models needed to be derived that emphasised optimal choices by economic agents. The dynamics of economies were then seen as being dependent on a fundamental set of 'deep parameters' that were embedded in the functions being optimised to derive decision rules. Only these might be regarded as constant, whereas the intermediate parameters of the older style model specifications would depend upon the nature of the regimes that agents believed they were in when making decisions. It seems fair to say that academics were extremely impressed by this argument and they promptly lost interest in the type of models that were then in use for policy analysis. This lack of interest still seems true today, although there is now an increasing trend to question whether the Lucas critique is of much importance when it comes to assessing whether the dynamics are particularly sensitive to the type of 'regime changes' that occur in reality (Rudebusch (2002)).

After this juncture, academics increasingly built empirical macroeconomic models that closely followed what was being done in the theoretical arena, in that decisions on expenditures were seen as optimal decisions by economic agents, based on the constraints they faced when looking into the future. The search also began for models that could describe observed price-setting behaviour as an optimal response. Intertemporal decision-making was central to these endeavours and that fact required a careful specification of how expectations about the future were formed. Mostly, expectations were taken to be fully rational; a better description might have been 'model consistent', as, in practice, they were taken to coincide with the predictions of the model that was being developed, rather than incorporating all available information, as implied by rationality. It was then a short step to recognise that uncertainty about the future had to be allowed for and this was handled by visualising the economy as being subject to various stochastic shocks which were imperfectly predictable. These models became known as dynamic stochastic general equilibrium (DSGE) models and they remain the dominant mode of macroeconomic modelling in academia today. In terms of our journey analogy, both the destination and routes were fully prescribed as the byproduct of some optimisation exercise.

Modellers involved in the policy process had a somewhat different reaction to the Lucas critique. While acknowledging the theoretical soundness of the case, they tended to ignore it in practice, perhaps due to a suspicion that any observed parameter instability came from more mundane factors, such as poorly understood specifications. Nevertheless, they were receptive to some of the main themes coming out of academia, if not the way in which these had been implemented. For some time they too had been concerned that the 'bottom-up' strategy of building models on an equation-by-equation basis often led to poor performance of the complete model. To discover the properties of the system, modellers had increasingly resorted to simulation methods and 'stress tests' to highlight weaknesses in the completed models. But, even though weaknesses might be revealed in this way, it was often harder to know how to respond to such information. In turn, this generated the feeling that there was a case for a more 'top-down' approach to modelling, and it was the latter that was the important contribution of the academics' research agenda. Still there was concern about going completely to a 'top-down' approach.

Unlike academics, the policy analysts had to work with models that were quite large. A major reason for this was that it would be unlikely that a small-scale model, which just incorporated (say) output, inflation, an interest rate and an exchange rate, would suffice for actual policy analysis, even though these small models might be useful for thinking about policy issues. The need to explain policy actions to the public inevitably meant that a greater degree of disaggregation was necessary.

It was also felt that simply imposing a top-down perspective on a larger model was unlikely to produce a good match to the actual dynamic behaviour of the variables. Moreover, it was a major computational task to implement such models; one that would have sorely challenged the capacity of a supercomputer of the 1980s. What one needed were models that were sizable and yet whose properties were relatively easy to understand, ie a central organising principle had to be found that could be used on the scale necessary for policy modelling. There was an example of such a successful development—the computable general equilibrium (CGE) models developed to analyse tax and tariff issues; these were large, but the answers they provided were relatively easy to understand. The reason why the outcomes of such a large dimensional system were simple to understand was that they could be conceptualised as coming from multiple markets, whose supply and demand curves had different elasticities which depended upon the parameterisation of the model. It was not quite as easy to find the same simplicity in macroeconomic modelling, since the key problems related to dynamic responses, and there was little theory regarding these that had widespread acceptance. Thus it was inevitable that there would be compromise; a downsizing of the models to obtain clarity was accompanied by the imposition of a top-down way of thinking, as well as the employment of a variety of different strategies when matching the models to the data.

These principles led to the development of what might be called hybrid models. Their core organising principle was the segmentation of the representation task into two stages: cast in terms of a journey the two stages corresponded to first selecting a destination and then choosing the route for getting there. In the first stage it was assumed that there was an equilibrium path along which the economy was seen to be evolving. This path might be either implicit or explicit. The second stage

involved stipulating the nature of the adjustment to the path, ie the route to be followed. This division enabled one to focus upon different ways of dealing with each segment rather than trying to deal with both at the same time; the latter being a strategy that was characteristic of the 'bottom-up' methods. Moreover, it enabled one to retain and emphasise in a precise way a concept that underlay a lot of the early work in macromodelling—the idea of a 'gap'—since a 'gap' could now be viewed as a deviation from an equilibrium path. We will distinguish between Type I and Type II hybrid models, based on whether the long-run equilibrium part—the destination—is treated implicitly or explicitly in the model. Models in which it is implicit will be termed Type I hybrid models; those in which it is explicit will be designated as Type II.

Within the dual approach characterising Type I hybrid models, the long-run relationships governing equilibrium growth paths were constructed from fairly loose economic reasoning. Thus it might be assumed that certain sets of variables were in a constant long-run relationship, for example consumption might be taken to be a proportion *(k)* of either income or wealth. Then, if one formed the ratio of consumption to wealth at a particular point in time, the extent to which it departed from *k* would be an index of the disequilibrium in consumption, and adjustments would be expected to take place to restore the balance. Accordingly, such long-run relationship were often used to produce measures of 'gaps' in the goods, labour and money markets. Some of the earliest work in this vein placed much emphasis on measuring a particular gap, namely that between the demand and supply of money, but interest in this particular gap has declined since its zenith in the mid-1970s. What was particularly attractive in this two-stage approach was that the idea dovetailed very neatly with an emerging econometric literature on 'co-integration', wherein variables might exhibit trending behaviour but the gaps between them might be trendless.

Finally, Type I modellers had to face up to the issue of how to describe the adjustment to an equilibrium position. Again, there was some assistance from econometric developments in the form of equilibrium correction mechanisms (EqCMs), ie equations that showed the speed at which the disequilibrium gap would be closed. Davidson *et al* (1978) made the case for these representations in a forceful way.

Data entered the construction of Type I hybrid models through the fact that there were a number of unknown parameters that characterised the equilibrium positions and the rates of adjustment. Thus the level of retail prices would be a combination of the level of domestic and imported goods prices or, as mentioned above, consumption might be a fraction of household wealth. These weights might be estimated from the data or imposed. Furthermore, following the arguments of Davidson *et al* (1978), the parameters of the EqCM model describing the adjustment mechanism were left to be determined by the data. The divergence between actual and equilibrium values of a variable could then be taken as a measure of the extent of disequilibrium, ie a gap that was to be closed. Based on these principles Type I hybrid models were developed in a number of central banks; current examples would be the area-wide model at the European Central Bank (see Fagan *et al* (2001)), the RIMINI model at the Bank of Norway, and the Economic Group Model at the Reserve Bank of Australia (see Beechey *et al* (2000)).

It is important to note that no long-run equilibrium path for any of the variables was computed in Type I models. Rather it was just the relationship that must hold between the variables that was described. Type II hybrid models took the further step of working with an explicit description of the equilibrium paths of variables. In this they were closer to the nature of DSGE models and, in fact, they utilised the same optimising framework as employed in DSGE models. But they shared with Type I models the characteristic that, once the equilibrium path was tied down by some theoretical specification, the data were largely responsible for the determination of the dynamic adjustment process to that path, ie the EqCM form was used and estimated. A major difference from Type I models though was that some decisions were influenced by expectations about the future, ie the models incorporated 'forward-looking behaviour', and it was this fact that required the existence and calculation of steady-state growth paths for their variables. The latter were needed since, in computing expectations of future variables, it was necessary to know the point to which they would eventually converge, ie the steady-state path. The derivation of equilibrium paths was therefore a key element in their *modus operandi*.

Because of the strong use of economic theory in Type II hybrid models, the outcome of experiments performed with these tended to be relatively easy to understand, in the same way as the solutions from CGE models were.

They also provided a completely consistent treatment of stocks and flows, eg if the stock of debt was accumulating it would ultimately affect (say) expenditures, since the debt-income ratio needed to be restored to realistic levels. Some of the earliest econometric models did not have such constraints, and so 'free lunches' abounded. Even in Type I hybrid models stock-flow constraints were never treated in an entirely consistent way. This treatment did, however, have a cost, in that certain assumptions needed to be made about key ratios such as the level of foreign debt to GDP, and these could become key determinants of the outcomes of the model. A good description of the structure of these models is in Powell and Murphy (1997), and variants have been used by the Singapore Monetary Authority in its Monetary Model of Singapore and the New Zealand Treasury (see Szeto (2002)).

Perhaps one of the most interesting features of Type II hybrid models was that they reflected academic models in their determination of the equilibrium or steady-state path, but retained a core feature of the older tradition in having separate equations for many of the variables of interest to the policy-maker. They were not uniform in design though and differed in the emphasis laid on a number of key points; in particular the degree to which decisions were made to depend upon expectations about future events. Future expectations were sometimes limited to financial markets, with decisions by consumers and investors being only partially forward looking, since it was felt that putting too much emphasis upon views of the future was probably unrealistic, and certainly did not accord with much survey evidence on economic agents' behaviour.

Hybrid models were probably most popular in the late 1980s and early 1990s. At that point a number of institutions and individuals began to argue that the economic theory used to form the equilibrium paths should also be used to describe the adjustment path, rather than simply allowing the data to determine the latter, ie the dynamics should be intrinsic to the model rather than being extrinsic to it. The first central bank to adopt such a philosophy was the Bank of Canada in its QPM model—Coletti *et al* (1996)—and this was soon followed by the Reserve Bank of New Zealand and its FPS model—Black *et al* (1997). Just like in the DSGE models that academics worked with, it was the imposition of theoretical principles that gave the models desirable properties. Unlike academic work however, the new policy models had a number of concessions to

perceptions about actual economic responses. First, it was recognised that not all decisions could be usefully viewed as optimal responses to an uncertain future. As had become evident from much empirical research, the academic models that incorporated such an assumption failed to match the dynamics evident in the data. Hence the policy models always incorporated some inertia in decisions, through the idea that these were often made by rule-of-thumb rather than optimally, although, as Nickell (1985) observed, it was generally possible to set up an optimisation problem that would rationalise a wide range of rules-of-thumb. These policy models also took a more restrictive view of the nature of shocks. Indeed, this had initially been true of academic models. Shocks were viewed as being either permanent or transitory, unlike in DSGE models in which one fully specifies a process for the shocks, and so they can have both permanent and transitory elements. Because these models have many similarities to DSGE models, but the nature of the shocks is not fully specified, I will refer to them as incomplete dynamic stochastic general equilibrium models (IDSGE).

These models also witnessed a change in attitude towards the role of data in describing the adjustment paths to an equilibrium. Historical outcomes shape these models but in an imprecise way. Often the adjustment path to an equilibrium tends to be imposed by the beliefs of the modellers and the monetary policy decision-makers. This makes sense if the relevant data are rather limited, as was the case in New Zealand after the reforms of the 1970s and 1980s. But there was also a feeling that the data might not accurately measure the variables of interest, and this failure to match theoretical constructs and data might lead to frequent changes in the model. Thus the QPM builders had this to say about their motivation for being cautious about the role of data in quantifying their model:

'… the inability of relatively unstructured, estimated models to predict well for any length of time outside their estimation period seemed to indicate that small-sample econometric problems were perhaps more fundamental than had been appreciated and that too much attention had been paid to capturing the idiosyncrasies of particular samples. There had been a systematic tendency towards over-fitting equations and too little attention to capturing the underlying economics. It was concluded that the model should focus on capturing the fundamental economics necessary to describe how the macroeconomy functions

and, in particular, how policy works, and that it be calibrated to reflect staff judgment of appropriate properties rather than estimated by econometric techniques'. (Coletti *et al* (1996, page 14)).

Modelling conflicts

As recounted above, the history of economic modelling can be regarded as one of attempting to solve a conflict between the distinct desires that a model should be both theoretically and empirically coherent. By the first we mean that the model outcomes can be explained by reference to some agreed-upon conception of the way in which the economy is thought to function, while the second relates to the ability of the model to explain the history of that economy. For many reasons it has proven impossible to satisfy both desires simultaneously, and therefore a trade-off is perceived to exist. One might conceive of this trade-off as a curve like that in Figure 1. At one end of the curve are theoretical models that have never been exposed to an historical data set, while, at the other, there are models that fit every quirk in the data set but whose outcomes are impossible to interpret. Being at either of these points is not particularly attractive to a policy-maker and so models used in the policy process have always been located along the interior points on the curve.

Figure 1
Trade-off between theoretical and empirical coherence for models
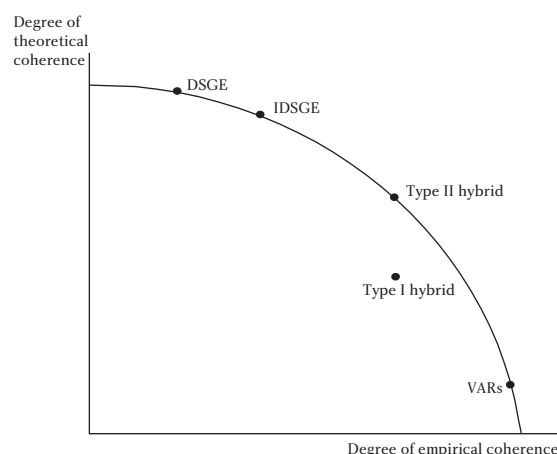
Of the categories of models listed previously, DSGE models tend to be closer to the left-hand end of the curve, while the early macro models were close to the right-hand end. Over time the curve has shifted outward and it has been possible to attain the same degree of empirical coherence with stronger theoretical constructs. Often this has simply been a reflection of the development of computer power: some theoretical

models that today appear to provide a reasonable match to the data could not have been solved 20 years ago. At any point in time, there will be a frontier of 'best-practice' models that shows the combinations of empirical and theoretical coherence that are attainable. There is no precise way of determining this frontier but sometimes opinions form about what is on and off the frontier. Thus, Type I hybrid models would now seem to be below the frontier, as one can achieve the same degree of empirical coherence with a clearer theoretical structure by using their Type II cousins.

However, just as with all trade-offs in economics, where a model is located on the best-practice curve is a function of the constraints that come from the institutional structure in which the models are to operate. For academics, best practice, or what is 'state of the art' for them, tends to be taken as being towards the left-hand end of the curve, although in recent times they have shown a greater interest in attaining empirical coherence. For policy modellers it has always been towards the right-hand end, although, as we have documented, the adoption of the hybrid class of models has moved the standard much closer to the centre of the curve, while the IDSGE models of some central banks lie even further up towards the left-hand end.

Modelling the UK economy

Since the formation of the MPC, the core model in use at the Bank of England has been the model variously called the medium-term macro model (MTMM) or the macro model (MM). It was first documented in Bank of England (1999) and updated in Bank of England (2000). We will refer to it with the acronym MM. In the past few years a new model has been under development as a potential core model. We will examine issues regarding the Bank of England's current core model and its prospective one under a number of headings that distinguish between the degree to which the overall framework is satisfactory and whether particular components of the model have any known inadequacies.

The overall design of the core models

The current core model (MM)

As we have detailed above, the current frontier for models in central banks would suggest that this is somewhere between Type II hybrid models and an IDSGE. I would classify the MM as a Type I hybrid

model since its equations can be thought of as having an EqCM structure, although this is not always obvious, and there is no explicit solution given for the steady-state path. One has to qualify the latter judgment by noting that there is mention in the 2000 model description of a version of the MM that does incorporate some forward-looking behaviour. However, published simulations of the model do not use such a device and it does not seem to have been used much in the monetary policy context. Thus, based on my classification, it would seem that the MM is not on the frontier and, in this sense, might be regarded as not entirely 'state of the art'. It seems highly likely that one could achieve the same empirical coherence with a stronger theoretical perspective that accords more closely with the conceptions of the MPC.

The question we might ask, however, is whether the MM is satisfactory when it comes to providing the information that the MPC needs for its decisions, ie perhaps it is off the frontier due to some constraints upon what is feasible. To determine an answer to this question it is natural to query the MPC on whether it is acceptable given the constraints or whether it might be regarded as being away from the frontier and thus capable of improvement. I received many opinions about this. Though there seemed to be a consensus that the model was not entirely satisfactory, there was much less consensus about what the perceived deficiencies were.

Some of the concerns stemmed from the fact that a number of the equations of the model had been known to possess deficiencies in fitting the history of the UK economy for some period of time—one can specifically mention those equations in the model describing price and wage movements and consumption. We will discuss this feature later in the report. It might be noted that the Bank of England is not alone in this regard. Many central banks have experienced difficulties in accounting for inflation and declining savings ratios in the 1990s, eg the Governor of the Bank of Canada, David Dodge (2002) says:

'With the low inflation target becoming increasingly credible, the whole nature of the inflation process seemed to change. The short-run response of inflation to measures of excess demand and supply appears to have fallen during this period. And the response of inflation to relative price shocks, such as changes in the exchange rate and energy prices, also seems to have declined.'

Though it is possible that these problems were worse at the Bank of England than in other central banks, it seems unproductive to engage in such comparative analysis when the relevant question should be whether adequate attempts were made to deal with any deficiencies when they became known. As this is a question pertaining to the methods for reviewing model performance, rather than to the question of overall design of models, I will examine it below.

We turn then to the question of overall model design, or theoretical coherence. Bank modellers and some MPC members felt that the MM was unsatisfactory, either because it failed to capture some linkages that were prominent in their thinking, eg an impact of profits upon investment expenditures, or because it was unlikely to produce trustworthy responses if one posed questions to it that involved the formation of expectations about the future. One can empathise with this latter viewpoint, since variables appear in it that are clearly forward looking, but which are modelled as depending simply on current outcomes. A good example of this would be equity prices, which are modelled as depending only on current nominal GDP (a proxy for dividends). Although we noted that there is mention of the potential for introducing forward-looking effects in some instances (although not for equity), the general feeling amongst modellers and some of the MPC members was that the way it had been performed was not very satisfactory, largely because it involved a grafting on to the model of a feature that did not fit well with its conceptual design. Since modellers in the United Kingdom have been world leaders in introducing such effects into large-scale macro models, it is perhaps a little surprising that the MM does not incorporate them when it is run in its standard mode. Perhaps this can be explained by an impression which I received that the original model was built under severe constraints at the time of the formation of the MPC. Although there was an existing model, it did not seem entirely suitable for the new monetary policy environment and so there was a need to modify it fairly quickly. At that time one could make the case that there had been more experience within the Bank with Type I hybrid models than with any other version.

As mentioned above, many reasons were given for regarding the theoretical structure of the MM as unsatisfactory. One of these, the preference for having a greater incidence of forward-looking elements in the core model, probably stems more from a desire for some

flexibility and the need to increase the level of confidence in the model being used than because it was felt that the absence of such mechanisms was responsible for any of the model's failures to fit the data. Indeed, although we do not have a lot of comparative research on the relative properties of models that do and do not incorporate forward-looking behaviour, recent research (Rudebusch (2002)), might suggest that the loss is not very great. Nevertheless, just as one can lose confidence in an old car's reliability in the face of the need for increasing levels of maintenance to fix its defects, one can lose confidence in a model like the MM when many adjustments have to be made to improve its performance. Potentially, this loss of confidence can have an impact on decisions, since experiments with human subjects have shown that the accuracy of predictions is positively related to the mood of a decision-maker. Consequently, it seems important to support mechanisms that will instil confidence in the systems used to inform policy analysis.

From this perspective, complaints about the absence of forward-looking expectations in the MM are a symptom of a deeper malaise. Even if forward-looking expectations had been incorporated in a satisfactory way into the MM, it seems highly likely that other reasons would have been found for dissatisfaction with it, and these negative feelings would have grown over time. One of these reasons is that the underlying framework was too vague and this often meant that it was hard to interpret some developments. Another comes from the knowledge that the degree of confidence in any decision support system by a decision-maker is very much dependent on the extent of that decision-maker's participation in its construction. Understandably, the MPC's involvement in the original development of the MM seems to have been minimal, and the short tenure of MPC members will always mean that, at any one time, there will be few members who have had a direct involvement with that process.

Finally, as the 'gap' terms that are a feature of Type I hybrid models come under intensive scrutiny, they are often found to be unreliable indicators. When they are defined quite precisely from some theoretical model, there is a chance that a new measure can be produced by varying the theory that has been employed, but, when they are just loosely thought of as indicators of the extent to which one is out of equilibrium, this is much harder to do. Within the MM there is a preponderance of the latter type of measures. If the equations being

driven by these gap terms perform credibly, then the fact that the relationship is a loose one probably does not lead to any serious doubts about the nature of the model. But, once the equations begin to fail, there can be a rapid loss of confidence in the model, unless it is possible to discuss alternative measurements of the gap terms by reference to the workings of a model. The MM is just not rich enough to do this in a satisfactory way.

An alternative core model (NMM)

Overall, I think it was well recognised among both Bank staff and the MPC that the MM was not really 'state of the art'. But while this may produce some pressures for its replacement, any action in this direction may not happen immediately. One can drive an old car for many years, repairing it when it fails, before one is forced to concede that there might be advantages in buying a new model. Replacement is made to look attractive, however, when we have a preference for going on journeys with a precise destination and we are concerned about whether our current car would be capable of getting there in all circumstances. In the same way, constant repair of a model eventually leads to a demand that it be replaced.

Apart from the direct financial costs of developing new models, it needs to be recognised that it takes time to decide on any new framework and to develop it. Thus there has been a project in the Bank for the past two years to develop a new model that would hopefully resolve some of the issues that had arisen with the MM. I will term this new model the NMM (new macro model). At the time of writing ,this model is still not fully operational, so that I do not feel that I can make any detailed comments upon it, but some observations need to be made.

In terms of my categories above, the NMM is an IDSGE model and so is at the frontier of central bank models. Indeed, from what I have seen of its structure, it significantly advances that frontier. Moreover, unlike existing IDSGE models which tend not to have been judged by their ability to replicate a wide range of data, it is intended that this one will be so judged, ie it aims to move the frontier relating theoretical and empirical coherence outwards. In this respect it is again novel. Although the model is being developed to replace the MM, it is the case that the MPC will be the final arbiters of whether it becomes the core model rather than being a supplementary one.

Replacing the MM with the NMM is likely to solve one of the sources of disquiet about the MM, namely that the destination (long-run equilibrium) will be much clearer in the NMM than in the MM, and some of the adjustment paths will be specified as part of the model structure. Because it has a better articulated theoretical structure it also has the potential to measure gaps in a different way than the MM does, see Neiss and Nelson (2001) for an example. But these are all potential gains. Whether they become actual ones remains to be seen. One problem that may arise in adopting the NMM stems from the nature of the MPC. Given its diversity, and the short tenure of its members, it may be difficult to get assent to any model that imposes strong theoretical specifications. Leaving things rather loose, as in Type I hybrid models, enables the model to be flexible and to accommodate different views. One will need to balance the benefits and costs of each approach. In this context, it is noticeable that, for the two central banks that have adopted IDSGE models, the policy decision is invested in the Governor and the discussion of options is performed by an internal committee of bank officials, ie the decisions are taken by internals alone. Perhaps some thought should have been given to moving to a Type II hybrid framework, since this has many similarities to the IDSGE constructs but retains the familiar structure of the MM. It should be said, however, that at this point, it is unclear how the NMM is to operate in the forecasting process and it may well end up as being a very sophisticated version of a Type II hybrid model.

The equations of the core model

Since it is the EqCM approach that is the basis of Type I hybrid models like the MM, it is natural to focus upon its components when asking about the adequacy of the equations contained within the MM. Thus one might query the set of variables that is to appear in an equilibrium relationship, the weights to be given to them, and the influences upon the adjustment paths. The first two are about how to measure the 'gaps' in a model like the MM in a satisfactory way and the last is about how to describe and capture effects on the speed of adjustment back to equilibrium.

There are many equations in the MM and it is outside the scope of this report to comment on them all. At various times there have been difficulties experienced with the equations describing consumption, investment and earnings. However, owing to the primacy of the inflation forecast in monetary policy decisions, it is useful to look at the equations pertaining to that

variable as illustrative of the types of problems that have been encountered in the use of the MM. Although the variable of ultimate interest is RPIX inflation, inflation in published versions of the MM is built up in a number of stages from the GDP deflator, import prices and various taxes.[1] If any of these forecast badly, then it is likely that forecasts of RPIX inflation would be inaccurate. Around the beginning of 1998 the GDP deflator equation started to exhibit constant overprediction, thereby initiating a widespread debate within the Bank and the MPC about the causes of this. Since the GDP deflator equation in Bank of England (2000) has inflation in that price index being driven by capacity utilisation (CAPU), unit labour costs (ulc) and various dynamic adjustment terms, it is natural that the debate should focus around the nature of these variables.

In particular, the debate centred upon the measurement of demand pressure and there were very public disagreements over the appropriate way to do this—see Wadhwani (2001)—with a number of members of the MPC clearly feeling that the measure that appeared in the model was a very poor one. However, even if one concurred with this position, exactly what should be done about it is less obvious. Suggesting an alternative specification of demand pressure is not enough, since the new specification has to be put together as part of a complete system. Certainly some of the proposals would have necessitated the adding of extra equations to the model in order to explain the new variables that were to replace CAPU. My impression is that the Bank modellers resisted this tendency towards an expansion of model size, and so were reluctant to depart from a measure that was generated within the model. Given that a proliferation of variables was the catalyst for early models becoming very large and unwieldy, one has to sympathise with this resistance.

Did any of the suggestions made about measuring CAPU in a different way lead to important changes in the GDP deflator equation? It seems not. Apart from the difficulties that would accompany an expansion in model size, it emerged during the period in which I was performing the review that new statistics released on the capital stock significantly modified the existing series on CAPU. Indeed, to such an extent that the fit of the equation with the new series was a considerable improvement on the old one. However, there is now some doubt about the nature of these data revisions and they seem to have been withdrawn. This episode illustrates a fundamental quandary for all users of models—sometimes it is not the model that is at fault, but the fact that available data do not accurately measure the concepts used in the model that is the explanation of poor fit. It should be said that there are a number of other equations in the MM where the fit has improved as a result of data revisions.

Moving away from the debate over CAPU, the External MPC Unit[2] in particular was very interested in the role of world prices in determining the GDP deflator, and argued very strongly for such an influence. It is now the case that there will be separate equations in the MM describing changes in the RPIX and GDP deflator indices and that world prices will have a role in influencing inflation.

Is it likely that these modifications will solve the problem with the GDP deflator equation? I have some doubts over this. One can see the fundamental difficulties facing the modellers in a simple way. In many countries inflation tends to be a fairly persistent process and that fact often aids predictability. To look at how persistent it is in the United Kingdom, we fit a third-order autoregressive process to the inflation series and sum the three coefficients. A value for the sum of unity would mean an extreme form of persistence, whereas a zero value means a lack of it. To see how persistence varies over time, we employ a rolling regression of 30 quarters. Thus the first estimate of the sum uses the 30 quarters starting in 1980 Q4, the next the 30 observations arising from 1981 Q1 onwards, and so on. Figure 2 plots these for a number of series— inflation in the GDP deflator (PGDP), unit labour cost growth rates (ULC), world price growth (adjusted for the exchange rate) (WPXADJ), and two measures of capacity utilisation, the one currently employed (CAPU) and that used in Bank of England (2000), labelled CAPU2.[3] It is

---

(1) A question that arises concerns what is the most useful degree of disaggregration. Opinion was divided on this. Some saw the existing degree as unnecessary. Another opinion was that a greater degree of disaggregation might be desirable. This was based on an interesting analogy with the situation in HM Treasury's model. Since the fiscal stance is the variable of most interest to the Treasury, this had led to a disaggregated tax system. It was argued that, because the MPC was entrusted with the task of achieving an inflation target, it might be appropriate for there to be a more disaggregated model of inflation. My feeling is that, unless the individual components had very similar equations describing their behaviour, the resulting model would be a little too complex for useful policy work.

(2) The External MPC Unit comprises Bank staff who have been seconded to assist the external members of the MPC in their research.

(3) The sample in Figure 2 runs from 1988 Q2 until 2000 Q1. The upper limit comes from the fact that the old capacity utilisation series was not available after the latter date.

Figure 2
Rolling regression estimates of sum of AR(3) coefficients for various series

Figure 3
**Rolling regression estimates of sum of AR(3) coefficients for RPIX inflation**
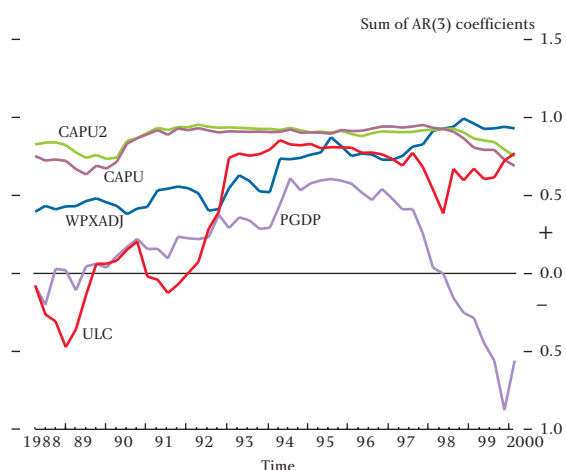
clear that, while there has been a very striking change in the degree of persistence of GDP deflator inflation from 1998, this is less true of any of the variables offered as explanators of it.[1] Without some convincing explanation for the decline in the persistence of the GDP deflator inflation, it seems likely that the equation will continue to cause problems. One possible explanation for the MM equation difficulties is that expectations of inflation do not have a direct effect on inflation outcomes, although there are good theoretical arguments for there to be such an effect. Given that expected future inflation may have declined since the late 1990s this could point to the need for such an extra variable to be incorporated into the current MM equations.

Is this decline in persistence also true of RPIX inflation? Wadhwani (2002) argued that it was and that there had been a large change after the MPC was formed. Figure 3 confirms this for seasonally adjusted RPIX inflation, although the changes occur a little later when one measures persistence as I have done rather than the way he did, which just involved the first-order autoregressive coefficient. Thus the comments quoted earlier by Dodge regarding changes in the inflation-generation process in Canada are certainly true in the United Kingdom. Such non-uniform structural changes place stresses upon the capabilities of a model like the MM to produce reasonable projections.

What should be done about an equation like that for the GDP deflator when it is clearly inadequate? In the first instance, new explanatory variables should be sought,
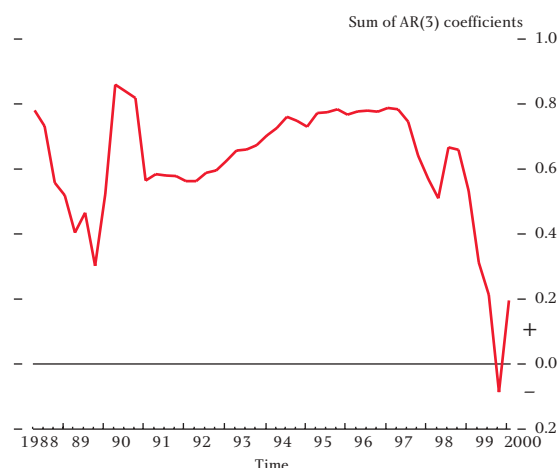
and there has certainly been a great deal of work done in this vein. A second option is to reconsider the nature of the specification, ie to build up a new theory of price determination. The changes in persistence suggest that some forces are at work that are not adequately captured by a modelling approach like that in the MM and point to the need to develop a new structural model. Finally, one might try to isolate the equation, so that its poor fit does not affect the projections of other variables in the model. In the current context, the link between the GDP deflator and RPIX would be severed and the latter would be modelled directly. Such a strategy seems a sensible one, and has recently been followed, although it may just be passing the problems on to the RPIX equation.

The modelling technology

Leaving aside the issue of the nature of specifications of the equations in the MM, one might ask whether the technology employed in designing and testing these specifications was the best possible. Since the dynamic relationships between the variables were left to the data to decide, it may be worth experimenting with the use of recently developed automatic model selection methods to achieve this, eg Pc-Gets (see Hendry and Krolzig (2001)). It was in fact noticeable that the dynamic specification of several of the equations had changed through time and some limited experimentation I had done suggested that the best dynamic structure might not have been chosen.

A second technical issue that was raised with me concerned the ability of researchers outside of the forecasting team to experiment with alternative

---

(1) It is evident though that the new capacity utilisation series does show a greater decline in persistence than the old one.

specifications of equations that were in need of repair. There was a perception that this was not as easy as it might be. One clearly wants to place some limits on this activity. Too frequent changes to the model can be disturbing for both making and communicating policy decisions. Only after a considerable degree of experimentation and a number of demonstrations that the changes do 'make a difference' should the model be amended. Moreover, any claims to superior performance will need to be verified by the forecasting team, and that could easily absorb a significant amount of their time. But currently the barriers to performing such work by outsiders do seem to me to be inordinately high. Most of these may be unintended as they derive from the separation of the tasks of data management, model estimation and model simulation. It would seem that, if the new model is adopted, then there will be greater integration of these tasks, although whether that will make it easier for outsiders to perform experiments on the model is unclear.

The review component of the modelling system

Although the broad structure of the model has remained fairly constant, there is a continuous process of experimentation with the specification of its equations. At least since mid-2000 there have been a number of procedures in place for reviewing the performance of the model equations in the interval between *Inflation Reports*. Bean and Jenkinson (2001) give a general description of this process. After each *Inflation Report* there is a meeting to review issues for the next forecasting round and some of these will involve discussion of the parts of the model that might need attention. The results of any work commissioned as a result of that meeting are then considered at a model review meeting before the next forecast round. The final decision whether to introduce any new specifications is made by the MPC. Modelling issues are also dealt with as part of the process for setting the future research programme of the Monetary Analysis department. The department suggests topics to pursue and the MPC is able to accept, reject, modify and prioritise these. This review of model performance is an important part of any modelling process. As we pointed out in the introduction, models are not fixed in stone, so that regular reviews are an important part of their use.

The fact that there has been public controversy over the nature of at least one of the equations suggests that the review process may not have worked as effectively in the early years of the MPC. Although I feel that this should

be less of an issue with the current structure, there may still be room for some improvement. The review process is most effective if there is considerable and ongoing experimentation with either entirely new specifications of equations or even the introduction of new paradigms, eg in the modelling of the price-setting mechanisms. This is a costly activity. It rarely leads to major changes in the model but, nevertheless, needs to be done. I saw a number of examples of this experimentation process while undertaking the review, particularly in relation to the inflation equations and, as has been observed, these experiments have finally led to a new set of inflation equations within the MM.

In general, while there was a commendable willingness to experiment, there seemed to be an attitude that it should be done in a restricted way, viz by tinkering with individual equations in the MM rather than by looking at quite radical changes to them. My overall impression was that too few resources are being allocated to thinking about quite different ways of modelling sectors, such as the price/wage sector, and that most attention has been devoted to focusing on extra variables to add to an equation or different ways of measuring the gaps, ie 'tinkering' and 'patching'. This may well have been the outcome of a belief that 'new paradigms' came under the NMM research programme. Now, there is always an innate conservatism amongst modellers and users of models when it comes to contemplating large changes in them. In this instance it seems to have been accentuated by a conscious decision to shift resources from the maintenance of the MM to the development of the NMM. For such a policy to have minimal impact on the viability of the core model, it was necessary to have a realistic timetable for the introduction of the new model into the monetary policy process. For a number of reasons there did appear to be overoptimism on this score. First, because any new model team is likely to be composed of very talented and experienced individuals, there is always the temptation to divert them from the task of constructing the new model into meeting the current demands of the monetary policy process. Although there may have been some diversions of this nature, they do appear to have been minimal and have actually been useful in learning about the capabilities of the NMM. Second, as already alluded to, the NMM is very much at the frontier of the class of IDSGE models, and so many more novel problems needed to be solved than if one had just used one of the existing IDSGE models, such as QPM or FPS. Some of these problems come from the need to adapt these models to the needs of what is a unique monetary policy process, but not all.

Much new thinking about models often comes from those who are a little removed from the day-to-day operation of models. Within the Bank structure this has often (but not exclusively) meant people working for the External MPC Unit. I think the latter have been important in generating new ideas about individual equations, eg in the inclusion of world prices into the price relationships, and in suggesting new ways of measuring the output gap. This fact illustrates the importance of the team that is directly engaged in model development and maintenance being open to proposals and criticisms from researchers who are outside the team, whether that be the External MPC Unit or others in Monetary Analysis.

The process of new model development

No model used in the policy process ever remains static. New paradigms emerge elsewhere and should be constantly reviewed to ascertain whether they might be incorporated into Bank procedures. Sometimes these adjustments can be handled as part of the model maintenance program, eg it might be decided to change the way in which wage behaviour is modelled. Others, however, constitute more fundamental shifts, eg it is possible that future models used in banks may be DSGE models, as the growth of computational power has meant that these models are starting to approach a size and complexity wherein they might actually be used in policy analysis. Thus the ECB has produced an experimental version along these lines—see Smets and Wouters (2002). Moreover, one can imagine that one of the major restrictions of existing DSGE models, viz that decisions are taken by a single representative agent, will be removed and replaced by multiple agents. Since early versions of these latter models seem to have the ability to explain many of the characteristics of financial time series, they may be very useful in central bank work for understanding movements in financial asset prices, and in explaining 'puzzles' such as the 'forward premium bias'. One may need to gather new data in order to calibrate these models, but it may also be possible to use many existing data on options and forward market prices for that task. Because one of the crucial aspects in forecasting is what should be done about equity prices and exchange rates, such models may ultimately prove very useful for central banks.

This raises the issue of how one encourages thinking about such new models. Most central banks keep an eye on such developments by ensuring that their staff attend conferences and seminars and produce summaries of these events that are distributed to other researchers. The quality of the summaries produced by Monetary Analysis staff is very high indeed. But I think something more will need to be done. Even if the NMM is not adopted as the new core model, I think the process of developing it was beneficial and has had, and will have, substantial spin-offs, in that it has suggested new directions for reconstructing the MM. If the NMM is adopted, I think it would be unfortunate if the process of thinking about new models was not done on an ongoing basis. One might envisage a small unit that performed this task. It would be best if it was constituted from researchers with a range of experience in forecasting and policy work with the core model, just as was true of the team building the NMM. My experience has been that the people most committed to developing a new model, and who are capable of going beyond simply modifying equations, are people who have worked on the existing core model and have become disenchanted with it. It seems important to channel these negative feelings into some positive directions.

The forecasting process

Although core models are the basis of any forecasting system, there are many issues to be considered concerning how one uses such a model in the process, and whether it is the only model to be used. As the latter aspect relates to diversification strategies, it will be left until the next section. Here we will consider the role of the core model in forecasting.

Forecasting with the core model within the Bank of England

Core models are never complete. There are always variables in them that are not modelled, but whose evolution will be an important determinant of outcomes. We generally refer to these variables as 'exogenous' to the model. International variables are an important example, but there are many others such as tax rates and some elements of government and private spending.[1] Even when forecasts of these exogenous variables have been assembled it may still not be possible to generate forecasts of the variables of interest from the core model. Because macroeconomic models are dynamic it is necessary to know the values of many variables

---

(1) In the case of international variables, the forecasts are generally made with the assistance of models built outside the Bank, such as NIGEM. I will not comment on this part of the modelling and forecasting systems since few people identified the international forecasts as major contributors to any forecast failures.

appearing in them up to the point in time that the forecast is being made. Thus, if one was forecasting inflation and growth in the fourth quarter of 2001, it is necessary to know (at least) GDP for the third quarter of that year. But the lag in the national accounts means that this outcome will generally be unknown, and so a forecast has to be made of it. This is done by using information from a variety of sources, eg survey data of expected output changes by firms. Within the Bank these forecasts are referred to as 'constraints'.

After deciding on a future path for the exogenous variables and the constraints, ideally one would produce projections with the core model automatically. But, in practice, these forecasts are generally modified in a number of ways. The modifications involve augmenting the base model in some way, so as to increase its effectiveness in the projection process. We can distinguish four types of adjustments that are commonly made:

1 Switching adjustments.
2 Past performance adjustments.
3 Future belief adjustments.
4 Profile and alignment adjustments.

Switching adjustments refer to the decision to switch the status of a variable that is endogenous in the base model to one that is exogenous over the forecast horizon. Thus, interest rates are often set at the current value or are constrained to follow a path that is determined by market expectations. Exchange rates may be set at the current value rather than evolving in line with the gap between short-term interest rates—the so-called uncovered interest parity assumption. Expected inflation rates may come directly from survey data or be extracted from the evidence in (say) index-linked bonds.

Past performance adjustments are adjustments made to compensate for some deficiency in the equation that has become evident from the history of the model forecasts or fit.

Future belief adjustments are made to enforce a variety of beliefs concerning developments over the forecast horizon, eg that the equilibrium point towards which the base model is directing the forecasts should be modified or perhaps that the speed of adjustment to the equilibrium point should be varied. This may be necessary since the equations of the base model implicitly embody assumptions about the long-run equilibrium growth paths of the variables, and it may be

felt that these have changed from their values in the sample period. A good example of the need for such an adjustment in recent years was the behaviour of consumption. The equations of the MM would imply a constant ratio of consumption to income in equilibrium, but often this was felt to be an unrealistic reference point in the forecast period. Sometimes the prior information justifying the adjustment related to institutional events, such as building society deregulation, which were hard to capture in a core model, and off-model experiments were used to produce some estimates of the transitory rise in consumption, which were then added on to the base model forecast. Another example was the desire to make some allowance for the effects upon productivity and costs of the events of 11 September 2001.

Finally, profile and alignment adjustments need to be performed in order to either smooth out the adjustments or to make them sum to some fixed amount.

Other classifications of adjustments have been given. Wallis and Whitley (1991) distinguish between automatic and discretionary adjustments. This distinction seems to be less descriptive of forecasting today (certainly for the Bank of England) than it was a decade ago. Then it was more common to make some automatic adjustments, in particular those that ensured that a perfect fit was obtained to an average of the outcomes over a short period immediately prior to the beginning of the forecast period. These automated adjustments were contrasted by Wallis and Whitley with those often made by 'judgmental' forecasters. Adjustments to reflect past forecast failure today are much less automatic and involve a good deal of judgment; hence our distinction between what one is responding to when making the judgments rather than the method of adjustment.

When making any adjustment that has an important influence on the forecast, it should be a requirement that a strong economic argument be advanced to justify it. It is this constraint that should keep the number and type of adjustments under some degree of control; without it the model can easily become submerged by the adjustments. The point is made very forcibly by Siviero and Terlizzese (2001, page 10) who say:

'...forecasters are tightly constrained, when making arbitrary adjustments to their forecasts, by the need to be explicit about the economic reasoning used to support their results'.

The technology for performing the adjustments in a model like the MM relies upon the fact that one can adjust the predictions of a variable by adding on to each equation the requisite amount that is desired for the forecast to change from that of the base model. It can be shown that, for models like the MM, the estimated 'intercept' terms entering any of its equations that have been cast into an EqCM form are an amalgam of a number of factors. These include the growth rates in all variables that describe the long-run equilibrium relationships, as well as the magnitude of various other parameters summarising the long-run and dynamic responses. Thus, if changes had occurred in any of these quantities over the sample period, such changes would tend to show up as a poor fit when the equation is estimated.

Ideally, one wants to make some explicit allowance for that fact. If one knew where the changes took place, one would utilise a smaller sample of observations. If one knew that the changes were smooth, one might try techniques that allow for evolving coefficients. But often these methods still cannot compensate for the shifts that have occurred, and it may be most convenient to simply make some allowance for these effects by adjusting the intercept in this equation. The same argument applies if it is thought that the same quantities might change over the forecast horizon.

Broadly speaking, past puzzle and profile/alignment adjustments are now done by the Bank staff in preparing the benchmark (or 'central tendency') forecast, and the MPC will make future prior adjustments before the forecasts appear in the *Inflation Report*. Of course there are always exceptions to this—adjustments performed for the likely future impact of 11 September were naturally first computed by the staff, and then reported to the MPC for a final decision on whether to apply them in an unmodified form. It is also the case that many 'past puzzle' adjustments have been arrived at after an extensive discussion between the staff and the MPC.

These adjustments are not trivial. Indeed, they can mean major changes in the forecasts from what the core model itself would produce if no adjustments were performed at all. For example, in Wadhwani (2001), the compensating and prior adjustments made to forecasts for the GDP deflator equation were removed, resulting in the forecasts made for the August 2001 round increasing from 2.5% to around 5%. Thus it becomes very difficult to know whether a good (or bad)

forecasting record is the consequence of the core model chosen or the adjustment mechanisms.

It is clear that one would like the forecasts coming from the core model to be the basis of discussion when the monetary policy decision is made, as these have a very clear interpretation and, as experiments have shown, the ability to provide a clear explanation of a forecast leads to better decisions. But the reality is that some adjustments are made by all central banks, and so there are really two issues to be decided: whether there are better ways of adjusting the forecasts and whether there has been adequate documentation and explanation of those that have been made. From what I saw of the process of preparing exogenous variable projections and the derivations of the constraints, I would have no hesitation in saying that the latter of the two issues is done very well. Regarding the former, though there are always suggestions that one might make about adjustments for specific equations, there is nothing that I saw which made me feel that the current methods were deficient. I was certainly satisfied with the professionalism shown by the Bank staff in carrying out these tasks. The same thing can be said about the method by which 'past puzzle' adjustments were prepared. I found that the explanations given for why the adjustment had to be made were very clear, and there was good disclosure of which of the equations of the core model were being modified and how much the forecast was modified by these changes.

Future adjustments are very difficult to assess. One might prescribe some formal ways of doing this if it was the prior of a single individual that had to be incorporated, but, in the UK context, the opinions of all the members of the MPC need to be melded into a set of adjustments. This is clearly a task that is more of an art than a science. It is also the case that it is not always possible to make an adjustment that will agree with all priors. This tension was resolved for some time by the inclusion of Table 6.B in some *Inflation Reports*; this table effectively registered any dissent by MPC members from the published central tendency forecast. This does not seem to be an area where one can say much about 'state of the art' methods, as it is unlikely that one could even get the members of the MPC to write down precisely what their priors are. About all one can ask is whether the staff at the Bank do respond in an adequate way to these prior beliefs when implementing the adjustments. My own opinion is that they have done a good job. I do note in passing that the prior adjustments that have been performed do seem to have

resulted in an improvement in the forecasts of a number of the key variables, such as inflation and consumption, and this alone shows the importance of them.

It is often the case in other central banks that more use is made of switching adjustments than in the Bank of England, eg replacing the investment equation forecasts with data collected from a survey of anticipated expenditures. It may be that such predictors are not as reliable in the United Kingdom as elsewhere. The main use here has been in handling items such as the exchange rate and equity prices. In the core model the exchange rate would change according to the short-run interest differentials that emerge between the United Kingdom and foreign countries—the uncovered interest parity (UIP) condition. An alternative might be to keep the exchange rate at its current level (actually an average of the 15 days prior to the forecast is used). Wadhwani (1999) notes that the difference in inflation forecasts between these two assumptions is 0.4 percentage points for the two year ahead inflation rate of RPIX. Consequently, there has been some controversy over what is the appropriate way of handling the exchange rate. Keeping it fixed at its current value is not very appealing if one knows that the short-term interest rate in the United Kingdom will rise. But allowing it to change to the full extent of the differential runs counter to a huge body of literature which, if anything, suggests the opposite response from that predicted by UIP. Research has suggested that the reasons for this unexpected response is that market participants form biased expectations of future exchange rate movements and that policy responses may be a contributor to such an outcome—McCallum (1994). Currently, the MPC's response to this dilemma is to make the change in the exchange rate only a proportion of the differential, which is a cautious response to our limited knowledge of what drives exchange rates, and represents a compromise between assigning a zero weight (an unchanged exchange rate) and unity (UIP). It may be that, as we begin to work with richer models that more closely replicate actual market trading in financial assets—ie they are peopled by agents who engage in momentum trading and contrarian strategies as well as working on fundamentals—better ways may be found to allow for appropriate exchange rate responses to policy changes. At the moment, it would be hard to devise a core model that is reasonably simple and yet which accounts for the seemingly perverse responses, so my feeling is that the current method is as close to state of the art as one could reasonably ask for.

Diversification of representations and projections

Most central banks are interested in diversifying the set of models that they use for representing and projecting the economy. Sometimes this fact is referred to as the institution possessing a 'suite of models'. We will describe these models as auxiliary models. Bank of England (1999) documents a variety of auxiliary models that have been used at various times within the Bank. Some of the models described in that document, eg the Batini-Haldane model, seem to have only been used for special purposes, and the most common alternatives to the MM model now seem to be various types of vector autoregressive (VAR) models. Although an eight-equation VAR was set out in Dhar *et al* (2000), mostly they are much smaller than that. These models were constructed primarily with the intention of informing the policy discussions, eg a small supply-side VAR was recently used to address questions relating to the impact of supply-side shocks, but in recent times they have also been introduced into the forecasting process. In the latter vein, since 2001 a number of small models that had been developed in the Bank, and which could be used to forecast GDP growth and RPIX inflation, have been used in an automatic way to generate forecasts. The range of possible outcomes from such models might be used both as an indicator of the degree of uncertainty in the projections and also as a guide to possible risks for the benchmark forecast.

In analysing diversification strategies it is useful to draw a distinction between those models using much the same information as the core model and those that work with an expanded data set, ie we will find it advantageous to distinguish between:

1 Models that largely reprocess information that is in the core model.
2 Models that aim to expand significantly the information set from that of the core model.

Diversifying policy-analytic models

The core model cannot incorporate all available information without becoming impossibly large. The subset of information used in it therefore represents the modellers' best judgment about what is potentially most important for the central task of projecting inflation and economic activity. Even then, some information might need to be excluded, since it may be hard to embed in whatever equations are used in the core model and yet

leave it in a tractable form. Indeed, this is one argument for having a suite of models—it enables one to keep the core model reasonably compact. A good example would be the role of money. This rarely appears in core models today. In the MM it appears as a mechanism for describing a monetary policy rule, but this is only in the simulation mode. It does appear in household financial wealth, and the latter variable influences a number of expenditures, but it is not separated out. The question that arises is whether information on money (and, perhaps more broadly, credit) provides useful information for forecasting—see Hauser and Brigden (2002) and Nelson (2000a,b). To shed light on this question small models have been developed in the Bank that allow money and credit to influence expenditure and inflation directly. This seems to be an efficient way of incorporating information that is not explicitly in the core model into the forecasting process. It seems likely that this work will be expanded and I would agree that this is a desirable development.

When considering policy options there are situations where one wishes either to trace out the consequences of alternative scenarios or to study the likely equilibrium position of a variable such as the real exchange rate. A literature has developed around the latter under the title of estimating a fundamental equilibrium exchange rate (FEER). A model like the MM is not well suited to any of these tasks. It does not have the structure to generate a well-defined FEER and it is unable to handle convincingly questions that often come up in scenario analysis, such as the consequences of a rise in the risk premium on assets. It is clear that there has been a need for some time for a model that resembles the NMM in order to handle questions such as these. Consequently, even if it was decided not to proceed with the NMM in a forecasting environment, it would seem worth retaining it as a way of generating policy scenarios that are likely to hold over the medium run. In this frame of reference it is not so critical that there be a good match to quarterly data sets.

Diversifying forecasting models

As mentioned above, many auxiliary models are used for producing forecasts of output and inflation. Currently, there are 32 of these in use at the Bank—the range includes VARs, time-varying component models and factor models. Possibly the only type of model that does not seem to have been used in a routine way would be non-linear models such as threshold autoregressions. Although there is a little evidence that the latter may be

useful for forecasting over the longer time horizons that are of interest to monetary policy makers, they have not been widely used amongst central banks.

Where I think there can be room for improvement is in the way that the auxiliary models are being selected. To date those selected have largely been byproducts of previous enquiries into some phenomenon, and have not been specifically designed for forecasting. To give an example of this point, although there are quite a few VARs within the current set of auxiliary models, there is an underrepresentation of the type that has been found useful for forecasting in other central banks—viz Bayesian vector autoregressions (BVARs). These models try to solve a problem in forecasting with VARs stemming from the fact that, in their basic form, far too many parameters need to be estimated from limited amounts of data. The solution is to impose some restrictions (priors) upon the set of parameters in order to effectively reduce the size of that set to a smaller dimension. Although these restrictions are 'statistical' rather than 'economic', they have proven to be effective—see Robertson and Tallman (1999) for the Federal Reserve Bank of Atlanta's experience. During the period of my review some models were introduced within the BVAR class, but I feel that more work needs to be done on this.

More generally, if the suite of models philosophy is to attain its designated purpose, then some resources will need to be devoted to considering the question of what type of alternative models should be employed, rather than simply having them be a byproduct of some past usage. For example, based on the work of Clements and Hendry (1998), there is a presumption that models using an EqCM structure may be inferior to those that ignore the disequilibrium term whenever there are likely to be substantial changes in the equilibrium paths over time, and it seems likely that this has occurred over the past decade. Thus the type of models being chosen as the auxiliary set should properly be chosen for their potential forecasting prowess.

It also seemed to me that there is a strong case for focusing on fewer auxiliary models in the forecasting process. For those auxiliary models that are largely just reprocessing the information in the core model, in principle it is possible to determine exactly why there is a difference between the core model and auxiliary model projections, since the latter model is virtually nested within the former. Such an exercise is to be recommended. It will generally be desirable for auxiliary

models to be designed so as to exploit genuinely useful extra information to that in the core model. Currently, those models that either emphasise money as a determinant of activity and inflation, or that work with a very large number of time series that have been reduced to a small number of factors, are good examples of what I would regard as genuine auxiliary models.

There are other questions that also need to be settled. One is whether the best strategy for generating alternative forecasts is to do so automatically, that is with little human intervention, from a large range of auxiliary models. A competing proposal would be to produce forecasts from a single model that is much smaller than the MM (say five or six equations) but in which the type of adjustments used in the MM are applied. Such a model might have to treat a number of variables as exogenous and then use some forecasts of them. This sort of model is close to what has been used at the Reserve Bank of Australia for some years, and it now seems as if a variant will be developed for use as a supplement to the core model in use at the Reserve Bank of New Zealand. In this connection it may be pertinent to note that the latter is in the same class as the NMM, and so it may be particularly important to address this issue if the NMM becomes the core model.

If it is decided to retain the feature of automatically generating forecasts then I think one should investigate whether the automatic selection methods currently go far enough. For example, there seems no reason why specifications of these models should remain constant from period to period. Indeed, the best one at any point in time could be chosen based on some automatic criterion, eg see Swanson and White (1997) for some evidence that this may be useful for forecasting over longer periods. There are other questions that also need to be looked at rather carefully when doing automatic forecasts, eg what vintage of data (in terms of revisions) should be selected; whether parameters should be chosen to (say) minimise forecast error at an horizon of four to six quarters rather than the one quarter ahead criterion used by estimators such as regression; and even what the criteria for evaluating a good forecasting model should be. I would recommend that a small group be set up to look closely at what would be a useful range of alternative models. In doing so one should pay some attention to thinking about the use of 'economic' rather than 'statistical' priors upon the coefficients of the VARs, ie following Ingram and Whiteman (1994) and Del Negro and Schorfheide

(2002) one might utilise some priors from an existing economic model that embodies some of the beliefs of the Bank staff and the MPC. A good example would be to base the prior on the NMM, but it would be possible to think about using the MM for this purpose.

The literature on forecasting often emphasises one other point about the utility of producing a range of forecasts, viz that averaging of them may produce a better forecast than would be available from any individual one. But, as Hendry and Clements (2002) point out, this may actually produce inferior forecasts if the different models utilise the same set of information and some of these encompass others; whether it does or not depends a good deal on how much structural change has occurred over the sample. At the moment, with the exception of the factor models, one might expect that some of the auxiliary models used at the Bank will encompass others and so this might cast some doubt upon summarising the information from the existing auxiliary models as simple averages. At the moment, the auxiliary model forecasts are presented to the MPC both as averages and in raw form, although it seems more likely that most attention would be paid to the averages. Again, this is a reason for some careful planning concerning the nature of the auxiliary models.

The quality of projections

How well have the modelling and forecasting systems performed? Making a strong distinction between the 'pure' model forecasts that employ no adjustments of any sort, and those forecasts that emerge after the adjustment process, it seems highly likely that the 'pure' model forecasts of inflation and output change would have been rather poor. If this is true then the forecasting system has adjusted the 'pure' model projections in the right direction. Exactly how much of this adjustment can be attributed to the Bank staff, and how much to the MPC, is probably impossible to determine. Over most of the time since the formation of the MPC there was not a forecast prepared by the staff. Today, although the staff do prepare a 'benchmark' forecast, this forecast inevitably reflects past judgments by the MPC.

There has been a lot of comment upon the relationship of the published forecasts to the outcomes. This comment has dwelt on three items—the 'bias' in forecasts, the extent to which forecasts of inflation have been consistently higher than the outcomes, and, to a lesser degree, the patterns in the forecast errors.

It is worth examining each of these separately but it will prove efficient to analyse the issue of persistence first.

Persistence in forecast errors

To analyse forecast errors we need to state what type of forecast is being considered. There are a number of possibilities. Thus, one might concentrate upon the ability to forecast outcomes for the next quarter. More relevant to policy is the ability to forecast annual inflation either over the next year or perhaps in one or two years' time. It is these forecasts that have attracted most attention and their prediction errors have been summarised and briefly analysed in the August 2001 and 2002 issues of the *Inflation Report*. In the August 2002 edition it was observed that, over the period between February 1998 and May 2001, the two year ahead forecast of the annual inflation rate had resulted in an overprediction of 0.5 percentage points, while growth had been slightly overpredicted.

No analysis was provided of other characteristics of the errors, but in a number of speeches, an ex-member of the MPC, Sushil Wadhwani, has pointed to the fact that all the forecast errors in the two year ahead prediction of inflation have been overestimates. For the one year ahead forecasts, there have been nine overpredictions and five underpredictions. Thus it is clear that the forecast errors are very persistent. Technically, they are positively correlated, ie if there is a positive forecast error for one period it tends to persist for many periods.

Those drawing attention to this phenomenon seem to believe that such an event is highly unlikely. Is this true? In the August 2001 *Inflation Report* analysis of the MPC forecast errors it was mentioned (page 58) that the persistence of forecast errors was likely to be particularly evident 'when projections are presented as four-quarter rates of change'. I will support this claim, arguing that the probability of such runs is in fact quite high, even for a forecaster who actually knew the process that generated inflation. The reason arises from the interaction of the nature of the inflation data and the quantities that are being forecast.

What we know about the quarterly inflation rate is that it is often a very persistent process, ie the outcome in the current quarter is highly correlated with the outcome in the previous quarter.[1] A simple polar case that emphasises this persistence, and which is easy to

analyse, is to assume that the quarterly inflation rate is a random walk. Then we consider a forecaster who forecasts with this model, and so is actually making the best possible forecasts. In the appendix it is shown that the correlation between the forecast errors varies depending on how far ahead we are attempting to forecast. If one is predicting quarterly inflation one quarter ahead it is zero; the correlation rises to 0.66 if we are predicting the annual inflation rate over the next year, and to 0.9 if we are predicting the annual inflation rate two years into the future. For the latter case, in four out of every ten realisations there will be six or more occurrences of forecast error of the same sign; and two out of ten times one will see eight in a row.

As should be stressed, in the experiment mentioned above the forecaster knows the actual model that generated the data, and so the forecast is of the highest quality. Yet runs of overpredictions and underpredictions are probable. The analysis in the appendix also indicates that we should see little in the way of consistent overprediction or underprediction of quarterly inflation rates one quarter ahead, a greater incidence of them when predicting the annual rate one year ahead, and a still larger one for the annual rate two years ahead. For the 18 quarterly forecasts made from August 1997 until November 2001, 10 of the forecasts were overpredictions and 8 were underpredictions, while the relationship between the one year and two year ahead forecasts was exactly as predicted by our simple model. Thus this simple analysis points to the fact that an observed run of overpredictions of inflation does not tell us much about the quality of the forecasts.

What about predicting GDP growth? There is much smaller persistence in the quarterly growth rates of GDP, so we might analyse it by assuming that the process being forecast has zero correlation rather than being a random walk. Then the forecast error correlation for the annual GDP growth rate one year ahead is 0.75, and it remains at this value when looking two years out. For the two year out forecasts there were six underpredictions and four overpredictions, which seems consistent with the degree of correlation that would be in the forecast errors.

Bias in forecasts

What about bias? In the August 2002 *Inflation Report* it was reported that the mean forecast error in inflation

---

(1) As was seen in Figure 3, the correlation in RPIX inflation was around 0.75 until late 1997, but then began to decline.

two years out made over the period from February 1998 to May 2001 had been 0.5 percentage points, and that there had been a small average overprediction of GDP growth. The latter has varied a lot over time—in the August 2001 *Inflation Report* it was indicated that there had been an underprediction up to that point in time.

It is useful to adopt our simple model of persistent inflation to analyse such outcomes. Now it is assumed that the forecaster makes an average forecast error for the quarterly inflation rate of b, ie if quarterly inflation is actually 0.6% then the average prediction of it might be 0.62%, giving a bias of $b = 0.02$ percentage points. In the appendix it is shown that this bias would become $10b$ when predicting the annual rate one year ahead and $26b$ in predicting the annual rate two years ahead. Thus, even a very small bias in predicting the inflation rate on a quarterly basis means a large bias in the annual estimate two years out. If, for example, there is a bias of 0.5 percentage points per annum in the annual rate of inflation two years out, then this would be consistent with a bias of 0.02 percentage points in the quarterly rate. In fact, the latter is the average bias in the quarterly inflation projection over the period from 1997 until 2001. One might note that a bias of 0.02 percentage points in a quarterly inflation prediction seems extremely small. It might also be noted that, if inflation were not a persistent process, then the bias in any annual inflation rate would be $4b$. Hence, the decline in the persistence of RPIX inflation recorded earlier suggests that there might be smaller average errors after 1999, which is consistent with the analysis provided in the August 2002 *Inflation Report*.

The argument above should not be construed as saying that the models being used for forecasting are accurate or that they cannot be improved on. Our analysis was simplified by treating inflation as a random walk, although our conclusions would be qualitatively the same if we allowed for a more realistic level of dependence. Rather our aim was to illustrate in a simple way why it is difficult to use statistics on inflation predictions two years into the future to assess the quality of forecasting procedures. Indeed, it might be argued that an observed bias in the forecast of the annual inflation rate two years out of just 0.5 percentage points is a tribute to the abilities of the Bank staff and the MPC to offset known inadequacies in the models.
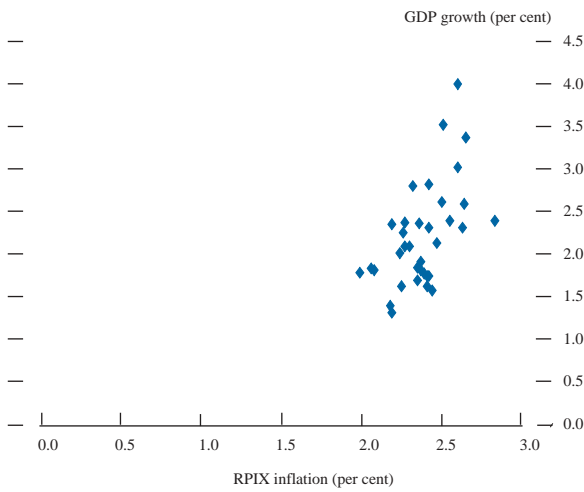
Patterns of forecast errors

There are also some joint patterns in forecast errors that have been commented upon. In particular, overprediction of inflation has been associated with underprediction of GDP growth in the forecasts prepared between August 1998 and May 1999. Explanations of this conjunction offered in the August 2002 *Inflation Report* have centred upon the strength of sterling and the fact that the pessimistic expectations generated by the Russian debt crisis, and the near-failure of the Long-Term Capital Management hedge fund, failed to materialise. These are certainly possible reasons for the observed patterns, but an analysis of the forecasts made during the period August 1998-May 1999 suggests that there may have been other factors at work. Figures 4 to 7 deal with this issue.
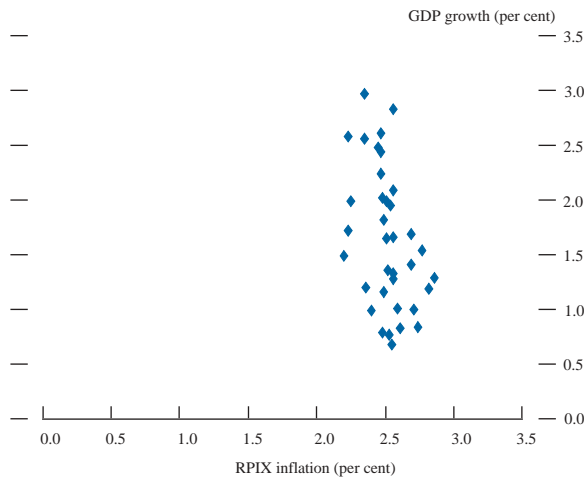
Each time a forecast was prepared, projections were made quarter-by-quarter for nine quarters into the future.[1] Hence, over the four forecast rounds between August 1997 and May 1998, the MPC made 36 projections of inflation and output growth. Figure 4 contains these 36 combinations of expected inflation and growth for the period between August 1997 and May 1998, in the form of a cross-plot of GDP growth against inflation. Figure 5 does the same for the period August 1998 to May 1999, and so on. Thus these graphs show the relationship that the MPC expected between inflation and output growth outcomes, and it does not relate to what actually occurred. Normally, we would expect a positive relationship between these two variables, ie higher expected growth would be associated with higher expected inflation. It is noticeable that this is true for all the years, except for the forecasts between August 1998 and May 1999. In this period of time the MPC expected a combination of high inflation and low growth. Now normally we would think that such a combination would be appropriate if the economy was being subject to supply-side shocks, eg a rise in oil prices, but that does not seem consistent with the story being told above for weak GDP growth. It seems more consistent with some comments that I received to the effect that, at the time of these forecasts, earnings growth seemed to be exceeding what would have been expected given the state of demand, and such an outcome might have been interpreted as a supply-side shock. I think this analysis shows that there are often insights available about errors in the forecasting process from looking at the inflation and growth forecasts

---

(1) One of these is for the current quarter in which the forecast is made as it is unknown at the time of forecast.
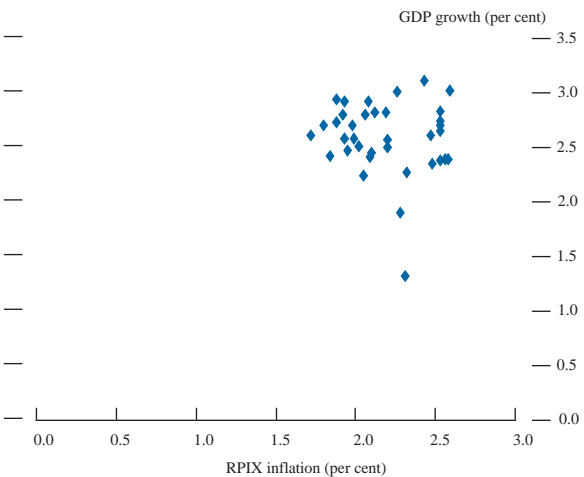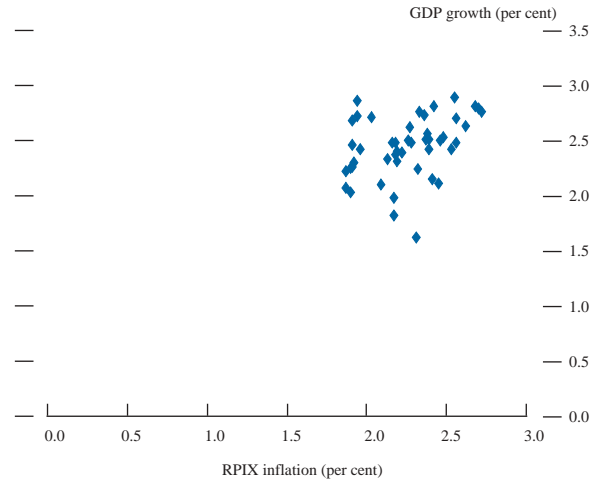
Figure 4
MPC forecasts, August 1997-May 1998

GDP growth (per cent)

RPIX inflation (per cent)

Figure 7
MPC forecasts, August 2000-May 2001

GDP growth (per cent)

RPIX inflation (per cent)

Figure 5
MPC forecasts, August 1998-May 1999

GDP growth (per cent)

RPIX inflation (per cent)

Figure 6
MPC forecasts, August 1999-May 2000

GDP growth (per cent)

RPIX inflation (per cent)

together rather than separately—the latter is what is provided by the fan charts. I would certainly recommend that more attention be paid to examining the joint outcomes when considering the forecasts being made at any point in time.

Communicating the models and projections

The Bank publishes a great deal of information on its projections, models and research. It is to be highly commended for this. Indeed, I think they are a model for other central banks. I certainly found that the material needed to write the report was largely available from published sources. Nevertheless, there are a few areas in which I think there can be some improvement.

The first item relates to the description of the model. As explained earlier, the type of model used at the Bank has a dual structure, involving an equilibrium position and a description of how the system responds to disequilibrium. It is logical then to present the equations of the model according to this principle. That was done for many of the equations in the MM description, but not for one of the most important, the earnings equation. It takes some effort to write it in the requisite form, so as to understand what the underlying structure of the relationship determining earnings is. I would recommend that, in order to consistently achieve clarity, all the equations of the model be written in EqCM form.

A second item relates to the presentation of the risks of the forecasts, ie the fan charts. There has been some criticism of these—see Wallis (2001)—one of which

involves the scaling factor used to determine the variability of the fan chart. In particular, it has been suggested that the value used is too high since it may have been constructed using data that are a mixture of different monetary policy regimes. The criticism seems a potentially correct one and this has been noted by the Bank in the August 2002 *Inflation Report*. Ultimately, this is an issue of whether the decline in the volatility of inflation is a temporary or permanent phenomenon. Around the world we have seen declines in the volatility of series such as GDP growth and inflation during the 1990s, but to date there has been no convincing explanation of these outcomes. Hence it remains unclear whether one should recalibrate the fan charts with a lower variability factor. Leaving this calibration issue aside, there is a good case for providing more information on the joint outcomes for inflation and growth. In the previous section I pointed out the benefits that can come from such a joint assessment. I am not sure that it is necessary to publish a bivariate fan chart to achieve this objective. It might be sufficient to provide the information needed to construct such an entity in the file that is now being published containing the quantitative parameters used in constructing the univariate fan charts.

Conclusion

When the MPC was formed in 1997 there were clearly going to be some stresses placed upon the existing modelling and forecasting systems. The novelty of having the monetary policy makers take responsibility for the forecast necessarily required some adjustments to previous operating procedures. Because it was a unique institutional arrangement, there was no real guidance to be had from other institutions over how to proceed. In the past few years a reasonably stable set of

activities connected with the monetary policy cycle seems to have emerged, and the system now seems to be working quite efficiently. There are clearly problems with the inputs into the decisions, as the core model has not performed in a satisfactory way, and that has required the MPC and Bank staff to make many adjustments to its forecasts. In this they have performed very creditably. As I have argued in the report, much of the criticism made of the actual forecasting record seems rather harsh and fails to take into account the nature of the series being forecast.

There is little doubt that the core model needs to be improved. It is possible that the new model that is being developed will be a more satisfactory vehicle for policy analysis and forecasting than the current one, but, until one sees its performance during a forecasting round, it is impossible to make a judgment about it. At this stage, all one can say is that it would appear to be 'best practice' and to correct many of the difficulties that have emerged when operating the current model. I also feel that greater attention needs to be paid to the suite of auxiliary models. In particular, the number used to make alternative forecasts could usefully be reduced and more attention paid to designing them as forecast vehicles, rather than just adapting models that were used for some other task.

Finally, the Bank has been quite sensitive to the need to perform ex-post forecast evaluation. Analysis that has been presented to the MPC has ranged from summarising the outcomes to attempting to ascertain the reasons for the errors, eg by decomposing the forecast errors for inflation in terms of the forecast errors for the influences on inflation of earnings, the exchange rate etc. I feel that the work in this area has been of high quality and certainly of adequate quantity.

Appendix

Let p_t be the quarterly variable to be forecast. A one quarter ahead projection involves forming the expected value of p_{t+1} given the information available at time t, $E_t(p_{t+1})$; an annual one year ahead forecast is $E_t(p_{t+1} + p_{t+2} + p_{t+3} + p_{t+4})$; and an annual inflation forecast two years ahead is $E_t(p_{t+5} + p_{t+6} + p_{t+7} + p_{t+8})$.

Now consider the case where p_t is a pure random walk, ie $p_t = p_{t-1} + e_t$, where e_t is an identically and independently distributed random variable with zero expectation and variance v. Then we would have as forecasts:

One quarter ahead
$E_t(p_{t+1}) = E_t(p_t + e_{t+1}) = p_t$

Annual inflation one year ahead
$E_t(p_{t+1} + p_{t+2} + p_{t+3} + p_{t+4}) = E_t(p_t + e_{t+1} + p_t + e_{t+2} + e_{t+1} + p_t + e_{t+3} + e_{t+2} + e_{t+1} + p_t + e_{t+4} + e_{t+3} + e_{t+2} + e_{t+1}) = E_t(4p_t + 4e_{t+1} + 3e_{t+2} + 2e_{t+3} + e_{t+4}) = 4p_t$

Annual inflation two years ahead
$E_t(p_{t+5} + p_{t+6} + p_{t+7} + p_{t+8}) = E_t(4p_t + 4e_{t+1} + 4e_{t+2} + 4e_{t+3} + 4e_{t+4} + 4e_{t+5} + 3e_{t+6} + 2e_{t+7} + e_{t+8}) = 4p_t$

Thus if the forecaster knew that the p_t followed a random walk the forecast errors would be:

One quarter ahead
$\phi_t = e_{t+1}$

Annual inflation one year ahead
$\phi_t = 4e_{t+1} + 3e_{t+2} + 2e_{t+3} + e_{t+4}$

Annual inflation two years ahead
$\phi_t = 4e_{t+1} + 4e_{t+2} + 4e_{t+3} + 4e_{t+4} + 4e_{t+5} + 3e_{t+6} + 2e_{t+7} + e_{t+8}$

The covariance between ϕ_t and ϕ_{t-1} will be:

One quarter ahead
Zero

Annual inflation one year ahead
$20v$

Annual inflation two years ahead
$84v$

while the variance of ϕ_t will be:

One quarter ahead
v

Annual inflation one year ahead
$30v$

Annual inflation two years ahead

94v

giving autocorrelations of zero, 0.66 and 0.89.

If the process for p_t had just been $p_t = e_t$ then the expected value of p_t for all forecast horizons would be zero and the forecast errors would be:

One quarter ahead

e_{t+1}

Annual inflation one year ahead

$e_{t+1} + e_{t+2} + e_{t+3} + e_{t+4}$

Annual inflation two years ahead

$e_{t+4} + e_{t+5} + e_{t+6} + e_{t+7}$

from which the autocorrelations are zero, 0.75 and 0.75.

Finally suppose that the forecaster makes a mistake and thinks that the process for p_t has the form:

$$p_t = b + p_{t-1} + e_t$$

This implies a bias of b in the quarterly forecast as it would forecast b rather than zero. Following the same derivations as above, we see that the forecast made of the annual inflation one year out will be $4p + b + 2b + 3b + 4b$ which produces a bias of $10b$. Continuing in the same way a bias of $26b$ is found for the annual inflation two years out. When the process is believed to be of the form $p_t = b + e_t$, whereas it is actually just e_t, the derivations above show that the bias is $4b$ for both of the annual forecasts.

References

Bank of England (1999), *Economic models at the Bank of England.*

Bank of England (2000), *Economic models at the Bank of England, September 2000 update.*

Bean, C and Jenkinson, N (2001), 'The formulation of monetary policy at the Bank of England', *Bank of England Quarterly Bulletin*, Winter, pages 434–41.

Beechey, M, Bharucha, N, Cagliarini, A, Gruen, D and Thompson, C (2000), 'A small model of the Australian macroeconomy', *Reserve Bank of Australia Discussion Paper*, 2000-05.

Black, R, Cassino, V, Drew, A, Hansen, E, Hunt, B, Rose, D and Scott, A (1997), 'The forecasting and policy system: the core model', *Reserve Bank of New Zealand Research Paper,* 43.

Clements, M P and Hendry, D F (1998), *Forecasting economic time series*, Cambridge University Press.

Coletti, D, Hunt, B, Rose, D and Tetlow, R (1996), 'The Bank of Canada's new quarterly Projection Model, Part 3: The dynamic model: QPM ', Bank of Canada, *Technical Report*, 75.

Del Negro, M and Schorfheide, F (2002), 'Priors from general equilibrium models for VARs', *Working paper 2002-14*, Federal Reserve Bank of Atlanta.

Davidson, J E H, Hendry, D F, Srba, F and Yeo, J S (1978), 'Econometric modelling of the aggregate time-series relationship between consumers' expenditure and income in the United Kingdom', *Economic Journal*, Vol. 88, pages 661–92.

Dhar, S, Pain, D and Thomas, R (2000), 'A small structural empirical model of the UK monetary transmission mechanism', *Bank of England Discussion Paper no. 113.*

Dodge, D (2002), 'Remarks prepared for the central bank governor's panel on inflation targeting', The American Economic Association and the North American Economics and Finance Association Atlanta, Georgia, January.

Fagan, G, Henry, J and Mestre, R (2001), 'An area-wide model (AWM) for the euro area', *Working paper No. 42*, European Central Bank.

Hauser, A and Brigden, A (2002), 'Money and credit in an inflation targeting regime', *Bank of England Quarterly Bulletin*, Autumn, pages 299–307.

Hendry, D F and Clements, M P (2002), 'Pooling of forecasts', *Econometrics Journal*, Vol. 5, pages 1–26.

Hendry, D F and Krolzig, H-M (2001), 'New developments in automatic general-to-specific modelling', in Stigum, B P (ed), *Econometrics and the Philosophy of Economics, Cambridge*, MA: MIT Press.

Ingram, B F and Whiteman, C H (1994), 'Towards a new Minnesota prior: forecasting macroeconomic series using real business cycle priors', *Journal of Monetary Economics*, Vol. 47, pages 497–510.

McCallum, B (1994), 'A re-consideration of the uncovered interest parity relationship', *Journal of Monetary Economics*, Vol. 33, pages 105–32.

Neiss, K S and Nelson, E (2001), 'The real interest rate gap as an inflation indicator', *Bank of England Working Paper no. 130*.

Nelson, E (2000a), 'Direct effects of base money on aggregate demand', *Journal of Monetary Economics*, Vol. 49, pages 687–708.

Nelson, E (2000b), 'The future of monetary aggregates in monetary policy analysis', Carnegie-Rochester Conference on Public Policy.

Nickell, S (1985), 'Error correction, partial adjustment and all that: an expository note', *Oxford Bulletin of Economics and Statistics*, Vol. 47, pages 119–30.

Powell, A A and Murphy C (1997), *Inside a modern macroeconometric model*, Berlin: Springer-Verlag.

Robertson, J C and Tallman, E W (1999), 'Vector autoregressions: forecasting and reality', *Federal Reserve Bank of Atlanta Economic Review*, Vol. 84 (First Quarter), pages 4–18.

Rudebusch, G D (2002), 'Assessing the Lucas critique in monetary policy models', *Working Paper 2002-02*, Federal Reserve Bank of San Francisco.

Siviero, S and Terlizzese, D (2001), 'Macroeconomic forecasting: debunking a few old wives' tales', *Discussion Paper 395*, Banca d́Italia.

Smets, F and Wouters, R (2002), 'An estimated stochastic dynamic general equilibrium model of the euro area', *Working Paper No. 171*, European Central Bank.

Swanson, N and White, H (1997), 'A model selection approach to real-time macroeconomic forecasting using linear models and artificial neural networks', *Review of Economics and Statistics*, Vol. 79, pages 540–50.

Szeto, K L (2002), 'A dynamic computable general equilibrium model of the New Zealand economy', *Working Paper 02/07*, New Zealand Treasury.

Wadhwani, S (1999), 'Currency puzzles', speech at London School of Economics.

Wadhwani, S (2001), 'Do we have a New Economy?', paper delivered at a CEPR/ESI Conference, Helsinki.

Wadhwani, S (2002), 'The MPC: some further challenges', speech given at National Institute of Economic and Social Research Governor's Seminar.

Wallis, K F (2001), 'Chi-squared tests of interval and density forecasts, and the Bank of England's fan charts', *International Journal of Forecasting*, (forthcoming).

Wallis, K F and Whitley, J D (1991), 'Sources of errors in forecasts and expectations: UK economic models', *Journal of Forecasting*, Vol. 10, pages 231–53.